palgrave macmillan

Advances in Risk Management

Edited by

Greg N. Gregoriou

ADVANCES IN RISK MANAGEMENT

Also edited by Greg N. Gregoriou

ASSET ALLOCATION AND INTERNATIONAL INVESTMENTS
DIVERSIFICATION AND PORTFOLIO MANAGEMENT OF MUTUAL FUNDS
PERFORMANCE OF MUTUAL FUNDS

Advances in Risk Management

Edited by

GREG N. GREGORIOU

palgrave macmillan

First published 2007 by
PALGRAVE MACMILLAN
Houndmills, Basingstoke, Hampshire RG21 6XS and
175 Fifth Avenue, New York, N.Y. 10010
Companies and representatives throughout the world

PALGRAVE MACMILLAN is the global academic imprint of the Palgrave Macmillan division of St. Martin's Press, LLC and of Palgrave Macmillan Ltd. Macmillan® is a registered trademark in the United States, United Kingdom and other countries. Palgrave is a registered trademark in the European Union and other countries.

ISBN-13: 978–0–230–01916–4
ISBN-10: 0–230–01916–1

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources.

A catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data
Advances in risk management / edited by Greg N. Gregoriou.
p. cm. — (Finance and capital markets series)
Includes bibliographical references and index.
ISBN 0–230–01916–1 (cloth)
1. Investment analysis. 2. Financial risk management. I. Gregoriou, Greg N., 1956–
II. Series: Finance and capital markets
HG4529.A36 2006
332.1068'1—dc22 2006045747

10 9 8 7 6 5 4 3 2 1
16 15 14 13 12 11 10 09 08 07

Printed and bound in Great Britain by
Antony Rowe Ltd, Chippenham and Eastbourne

Contents

Acknowledgements

I would like to thank Stephen Rutt, Publishing Director, and Alexandra Dawe, Assistant Editor, at Palgrave Macmillan for their suggestions, efficiency and helpful comments throughout the production process, as well as Keith Povey (with Paul Dennison and Nick Fox) for copyediting and editorial supervision of the highest order. In addition, I would like to thank the numerous anonymous referees in the US and Europe during the review and selection process of the articles proposed for this volume.

Notes on the Contributors

The Editor

Greg N. Gregoriou is Associate Professor of Finance and coordinator of faculty research in the School of Business and Economics at the State University of New York (Plattsburgh). He obtained his PhD (Finance) from the University of Quebec at Montreal and is the hedge fund editor for the peer-reviewed journal *Derivatives Use, Trading and Regulation*, published by Palgrave Macmillan, based in the UK. He has authored over 50 articles on hedge funds, and managed futures in various US and UK peer-reviewed publications, including the *Journal of Portfolio Management*, the Journal of Futures Markets, the European Journal of Finance, the Journal of Asset Management, the European Journal of Operational Research, and Annals of Operations Research. He has published four books with John Wiley and Sons Inc. and four with Elsevier.

The Contributors

Taras Beletski is a PhD student in financial mathematics and a member of the Graduate Research Training Programme, Mathematics and Practice, at the University of Kaiserslautern (Germany). After having worked for the financial mathematics group of the research centre CAESAR in Bonn (Germany), he joined the financial mathematics group of Prof. Dr Ralf Korn in Kaiserslautern. His main research areas and interests are continuous-time portfolio optimization, modeling of inflation and statistical methods related to the Kalman filter.

Olha Bodnar is a research assistant in the Department of Statistics, European University Viadrina, Frankfurt (Oder), Germany. She obtained her PhD in Statistics from the European University Viadrina, Frankfurt (Oder), Germany. Ms Bodnar has worked on the Scientific Project Statistik Interaktiv (Multimedia Teaching Program), European University Viadrina, Frankfurt (Oder), Germany, and is a specialist in the Risk Department, Western Ukrainian Commercial Bank, Lviv, Ukraine.

Emanuele Borgonovo holds a PhD in Risk and Probability from the Massachusetts Institute of Technology. He is currently an Assistant Professor in the Department of Quantitative Methods at Bocconi University where he teaches pure and applied mathematics courses. His research focuses on operations research and on sensitivity analysis of mathematical models.

Riccardo Bramante is Professor of Economic Statistics at Milan Catholic University. His research interests include technologies for financial markets, risk analysis and time series modeling. He works as a consultant to primary financial institutions. He is also a member of the Applied Statistics Laboratory of the Milan Catholic University.

Alain Charbonneau gained his MSc in Mathematics in 1989 and a PhD in Mathematics from the Université Laval (Québec) in 2003. He is a Professor of Mathematics and Numerical Methods in the Department of Computer Sciences at the Université du Québec en Outaouais. His current research is in the areas of numerical simulation in optical waveguides, finite elements methods, and numerical methods.

Helena Chuliá is a researcher at the University of Valencia (Spain). She obtained a degree in Management and Business Administration from the University of Valencia (Spain). She is currently working on her PhD thesis on quantitative finance. Her current areas of interest focus on applied financial econometrics, portfolio management and international finance.

Francisco J. Climent is Professor of Finance in the Department of Financial Economics at the University of Valencia (Spain). He received a PhD in financial economics at the University of Valencia, and has published articles in financial journals such as the *Journal of Financial Markets*, *International Review of Economics and Finance*, *Journal of Asset Management*, *Review of Financial Markets*, *Investigaciones Económicas*, *Revista Española de Financiación y Contabilidad* and *Revista Europea de Dirección y de Economía de la Empresa*. His current areas of interest are international finance, integration among financial markets, risk management, population economics and financial econometrics.

Yves Crama is Professor of Operations Research and Production Management, and Director General of the HEC Management School of the University of Liège in Belgium. He holds a PhD in operations research from Rutgers University, USA. He is interested in the theory and in the applications of optimization, mathematical modeling and algorithms, with an emphasis on applications arising in production/operations management and in finance. He has published about 60 papers on these topics in leading international journals, he is the coauthor of a monograph on production planning, and he is the associate editor of several scientific journals.

Giuseppe Di Graziano holds a PhD in applied mathematics from the University of Cambridge. Previously, he has been a member of the quantitative research team of Lehman Brothers in London. He is the author of several publications in credit derivatives, modeling and the Markov-chain approach to derivatives pricing.

Abdeljalil El-Moussadek has recently obtained his MSc in applied finance at l'École des Sciences de la Gestion, University of Quebec at Montreal.

Jean-David Fermanian is the Head of Risk at Cooperneff AM. Previously, he was Head of Risk Methodologies at Ixis CIB, after being Professor in statistics at ENSAE (Paris) and head of the statistics laboratory at the Center for Research in Economics and Statistics (CREST). His research interests include particularly survival analysis, credit portfolio modeling, simulated methods and copulas. He has published numerous articles in economics, statistics and financial econometrics. He graduated from the École Normale Supérieure and ENSAE, he holds a doctorate in statistics from the University Paris 6.

Giampaolo Gabbi is Professor of Banking and of Risk Management at the University of Siena, Italy, and a senior teacher at SDA Bocconi Milan, where he coordinates several executive courses on financial forecasting and risk management. He is head of the financial areas of the Masters in Economics of the University of Siena. Professor Gabbi holds a PhD in Banking and Corporate Management, and has published many books and articles in referred journals, including *Decision Technologies for Computational Management Science, the MTA Journal, Managerial Finance*, and the *European Journal of Finance*.

Stefano Galluccio is in charge of exotic interest-rate and hybrid derivatives trading at BNP Paribas in London. Prior to this, he was deputy head of interest rate derivatives research at BNP Paribas London. Galluccio holds a PhD in mathematical physics from the Ecole Polytechnique of Lausanne and held academic positions in Boston and Paris. He is the author of several publications on infinite dimensional term-structure modeling, market models for interest-rate derivatives, jump-diffusion modeling, hybrid derivatives and model misspecification theory.

Hayette Gatfaoui graduated in 2002 with a PhD on "Default Risk Valuation of Financial Assets" at the University of Paris 1. He has taught for five years at the University of Paris 1 (Pantheon-Sorbonne), and is now an Associate Professor at Rouen Graduate School of Management. He specialized in applied mathematics (holds a master's degree in stochastic modeling for finance and economics), and is currently advising financial firms about risk measurement and risk management topics for asset management as well as credit risk management purposes. Dr Gatfaoui is also a referee for the *International Journal of Theoretical and Applied Finance* (*IJTAF*). Briefly, his current research areas concern risk typology in financial markets, quantitative finance and risk(s) analysis.

Georges Hübner is the Deloitte Professor of Financial Management at HEC Management School, University of Liège. He is also Professor of Finance at Maastricht University and EDHEC (Lille-Nice) and Research Director at the Luxembourg School of Finance, University of Luxembourg. He has taught at the executive and postgraduate levels in several countries in Europe, America, Africa and Asia, and intervenes in the FRM and CAIA curricula. He holds a PhD in management from INSEAD, and has published several books and papers on operational, market and credit risk. In 2002 he received the Iddo Sarnat Award for the best paper published in the *Journal of Banking and Finance*. He is also the inventor of the Generalized Treynor Ratio for the measurement of portfolio performance, published in the *Review of Finance*.

Stephen Jewson works at Risk Management Solutions in London where he manages a group that produces commercial software and meteorological data for the weather derivatives industry. Prior to RMS he worked at the universities of Reading, Monash, Bologna and Oxford. He has a doctorate from Oxford in atmospheric science and a degree in mathematics from Cambridge. Dr Jewson has published a large number of articles in the fields of fundamental climate research, applied meteorology and weather derivatives pricing and is the lead author of the book *Weather Derivative Valuation*, published by Cambridge University Press. His main interest is in taking industrial problems related to meteorology and formulating and solving them using best-practice mathematical modelling methods.

Thadavillil Jithendranathan is an Associate Professor of Finance at the University of St Thomas, St Paul, Minnesota. He obtained his PhD in finance from the City University of New York. Thadavillil has published research papers in financial intermediation, international finance and investments.

His dissertation on currency futures options won the Oscar Larson award for the best dissertation in Business at the City University of New York. Prior to his teaching career, Thadavillil has worked in the areas of accounting and management in various countries around the world.

Ralf Korn studied mathematics at the University of Mainz (Germany) where he also obtained his PhD. He has been Professor of Financial Mathematics and Stochastic Control at the University of Kaiserslautern (Germany) since 1999. He has written three books in financial mathematics among them the well-known *Optimal portfolios* (1997) and *Option Pricing and Portfolio Optimization* (2001), and has published numerous articles on continuous-time finance. His main research interests are continuous-time portfolio optimization, transaction costs, value preserving strategies, crash modeling and worst-case approaches, inflation and numerical methods. He is a board member of the German Society for Insurance and Financial Mathematics, DGVFM, a member of the Scientific Advisory Board of the Fraunhofer Institute for Industrial Mathematics, ITWM (where he also heads the group on financial mathematics that has done numerous joint projects with the banking industry), speaker of the excellence cluster "Dependable adaptive systems and mathematical modeling" and associate editor of various journals (among them *Mathematical Finance*). He has organized many international conferences and summer schools in financial mathematics and is currently dean of the Department of Mathematics at the University of Kaiserslautern.

Annick Lambert obtained a Certificat supérieur d'études statistiques at the Institut de Statistique de l'Université de Paris in 1973 and a Diplôme d'études approfondies (DEA) in Applied Mathematics at the Université Claude Bernard (Paris) in 1974. She received a MSc in statistics in 1975 and a PhD in statistics in 1978 from the University of Wisconsin at Madison. She is a Professor of Statistics at the Université du Québec en Outaouais. Her current research interests cover asymptotic distribution theory, non parametric statistics, and linear statistical models.

François-Serge Lhabitant is responsible for investment research at Kedge Capital. He was previously the Head of Quantitative Risk Management at Union Bancaire Privée (Geneva), and a Director at UBS/Global Asset Management. On the academic side, Dr Lhabitant is a Professor of Finance at the University of Lausanne and at EDHEC Business School. His specialist skills are in the areas of alternative investment (hedge funds) and emerging markets. He is the author of five books on these two subjects and has published numerous research and scientific articles. He is a member of the Scientific Council of the *Autorité des Marches Financiers*, the French regulatory body.

Claudio Marsala is quantitative portfolio manager in Ras Asset Management and prior to this spent several years in the risk management department of Ras Asset Management, that he joined in 2001. His main focus is on the practical application of econometric models to portfolio management. He studied economics and econometrics in Pisa and holds a Masters degree in finance from CORIPE, University of Turin.

Manuel Moreno holds a PhD in economics from the Universidad Carlos III of Madrid and a BSc in mathematics from the Universidad Complutense of Madrid. He is currently Assistant Professor of Financial Economics and Accounting at the University of Castilla La-Mancha at Toledo (Spain), associate editor of *Revista de Economía Financiera and a* member of GARP (Global Association of Risk Professionals). He has previously held teaching and research positions at the Financial Option Research Centre (Warwick Business School), Universidad Carlos III de Madrid, IESE Business School and Instituto de Empresa Business School. In the past, he was the co-Director of the Masters in Finance at the University Pompeu Fabra in Barcelona (Spain). His research interests focus on finance in continuous time with special emphasis on derivatives markets, financial engineering applications, pricing of derivatives, empirical analysis of different pricing models, portfolio management and term structure models. His research has been published in a number of academic journals including *Review of Derivatives Research* and *Journal of Futures Markets* as well as in professional volumes. He has presented his work at different international conferences and has given invited talks in many academic and non-academic institutions.

Massimiliano Pallotta joined the risk management team of Ras Asset Management in 1997, working on the quantitative and technological aspects of the development of the company's internal model; in June 2004 he joined the quantitative portfolio management unit. He studied mathematics in Milan and has practical experience on the more advanced technological aspects of the risk management and portfolio management process.

Jean-Paul Paquin received a MSc in Business Administration (applied mathematics) in 1969 and a MSc in economics in 1971 from the Université de Montréal, and a PhD in economics (econometrics) in 1979 from the University of Ottawa. His work experience includes extensive field work in financial evaluation and economic planning (input–output analysis) for the World Bank, financial risk assessment for the Department of National Defence of Canada, and strategic planning for a major Canadian Credit Union. He currently teaches project financial evaluation at the Université du Québec in the Master's program in project management and is an Associate Professor and researcher at ISMANS (France). His most recent research, publications and conferences cover project financial evaluation, project risk management and quality management ("The Earned Quality Method", IEEE Transactions on Engineering Management, Feb. 2000), and strategic management ("Market Force Field" – ASEM 2004 and 2005 annual conferences).

Jean-Philippe Peters joined the Risk Management team at Deloitte Luxembourg in 2002. His expertise focuses on operational risk measurement and management for banking institutions within the Basel II framework. Beside practical assignments with clients, he has been involved in academic research activities that led to the publication of several papers and two books. Prior to Deloitte, he was a researcher at the University of Liège (Belgium) and studied volatility modeling with applications to risk management. This led to the release of GARCH, a software package to estimate volatility models. Jean-Philippe holds a MA in business management from the University of Liège and he is a certified Financial Risk Manager (FRM) by the Global Association of Risk Professionals. Since 2004, he has been a PhD student in finance at the HEC Management School of the University of Liège.

Marco Percoco, PhD, is a Research Fellow in the Department of Economics at Bocconi University (Milan, Italy) where he teaches several graduate courses in the field of applied economics. His main research interests are in public policy, infrastructure finance and economics and sensitivity analysis of economic models.

Amiyatosh Purnanandam is an Assistant Professor of Finance at the Ross School of Business at the University of Michigan, Ann Arbor. He obtained his PhD in finance from Cornell University. He holds an MBA and BTech (Electrical Eng.) from the Indian Institute of Management and the Indian Institute of Technology respectively. His primary research interests are banking, risk management and security issuance decisions of firms. His research has been published in finance journals such as the *Review of Financial Studies*, the *Journal of Financial Services Research* and *Finance Research Letters*.

Pierre Rostan, PhD, is Professor of Finance at Audencia Nantes (France), School of Management. He was previously a derivatives research analyst for the Montreal Stock Exchange. He is the author of a book on the computational aspects of financial derivatives and has published numerous articles on interest rate derivatives.

Mohammed Sbai followed engineering studies in the École Nationale des Ponts et Chaussées, specializing in applied mathematics. From 2004 to 2005, Mohammed worked at IXIS CIB for a one-year internship under the supervision of Jean-David Fermanian. Currently he is pursuing a Masters in applied mathematics for finance.

Pilar Soriano is a researcher at the University of Valencia (Spain). She obtained her BA in European Management from the University of Westminster (UK) and has a degree in management and business administration from the University of Valencia. She is currently working on her PhD thesis on quantitative finance. Her current areas of interest focus on financial econometrics, international finance and risk management.

Raymond Théoret, PhD, is Professor of Finance at the École des Sciences de la Gestion of the University of Quebec (Montréal, Canada). He has written many books and articles on banking, portfolio management and the technical aspects of financial engineering. He was previously senior economist in a major Canadian financial institution and was also Professor at HEC Montréal.

Hipòlit Torró is Professor of Finance in the department of financial economics at the University of Valencia (Spain). He obtained a degree in economics and business with honors and a PhD in financial economics at the University of Valencia and a MSc in financial mathematics at the Universities of Heriot-Watt and Edinburgh in Scotland. He has published articles in financial journals such as the Journal of Futures Markets, the Journal of Risk Finance, Moneda y Crédito, Revista Española de Financiación y Contabilidad. His current areas of interest are portfolio management (hedging and trading rules) and financial modeling of stock, interest rates, bonds and commodities (weather and electricity).

Mitch Warachka is an Assistant Professor in the Lee Kong Chian School of Business at the Singapore Management University. He obtained his PhD in finance from Cornell University and an MBA from the University of Chicago. His research interests include asset pricing, derivative securities and risk management. He research has appeared in publications such as the *Journal of Financial Economics*, *Review of Financial Studies*, *Journal of Banking and Finance* as well as the *Journal of Risk*.

Raffaele Zenti has been risk manager at Ras Asset Management from 1997 to June 2004, working on the development of the internal risk model and on the definition and application of risk policies to actively managed portfolios. Since 2004 he has been a quantitative portfolio manager: he supervises a department that manages several portfolios using non-subjective models. He studied economics and statistics in Turin and is Lecturer and the Master, course in finance of CORIPE, University of Turin.

William T. Ziemba is the Alumni Professor of Financial Modeling and Stochastic Optimization (Emeritus) at the University of British Columbia, Vancouver, Canada. In 2005 he was Visiting Professor of Finance at the finance department of MIT's Sloan School of Management, and the statistics department of the University of Washington. In 2006 he is a Visiting Professor at the ISMA Centre, University of Reading, EDHEC in Nice, France and Luiss University in Rome teaching courses in the great investors, asset-liability management and security market anomalies. Besides being active in his own publishing (articles and books) and lectures around the world, he is series editor for North-Holland's Handbooks in Finance.

Yonggan Zhao is an Assistant Professor of Banking and Finance at Nanyang Business School, Nanyang Technological University, Singapore, where he has been teaching since 2001. His PhD is from the University of British Columbia. His general research interests encompass the theoretical and numerical investigation of financial investment models, dynamic portfolio management and option valuation and replication. He has a particular interest in dynamic models of risk control and risk management in an incomplete market setting where practical constraints exist.

Introduction

Chapter 1 examines the estimation of operational risk exposure of financial institutions, and its dependence on the floor level at which operational losses are collected. The chapter shows that the choice of the collection threshold is not likely to influence the economic capital if extreme loss events are properly accounted for. Overall, the choice of the collection threshold should rather be guided by a simple profit/cost analysis than by regulatory arbitrage considerations.

Chapter 2 introduces a risk measure defined on portfolio holdings. In contrast to terminal portfolio values, this domain is conducive to having diversification reduce portfolio risk. The risk of a portfolio is determined by its distance from a set of acceptable portfolios. More importantly, this distance involves as many components as there are available assets, which includes but is not limited to risk-free capital. As a consequence, the role of derivative as well as insurance contracts in risk management is recognized.

Chapter 3 looks at the sensitivity analysis of volatility and return models that can be thought of as an essential ingredient in portfolio management. The Differential Importance Measure (DIM) is a generalization of local sensitivity analysis techniques and provides insights for the analysis of the impact of parameter changes. By considering a portfolio GARCH model, we make use of the DIM to identify the most important stocks in a given portfolio, i.e. those stocks whose change is meant to generate substantial changes in the portfolio return volatility. In order to provide some empirical application of the proposed technique, we consider a portfolio of 30 stocks, replicating the Dow Jones Index composition as at 2002.

Chapter 4 presents several applications of a two-factor continuous-time model of the term structure of interest rates, previously presented in Moreno (2003), for managing interest rate risk. New measures that generalize conventional duration and convexity are presented and applied in different situations to manage market and yield curve risks. After showing how to immunize a bond portfolio with bond options, the authors present and illustrate numerically how these new measures can solve the limitations of conventional duration.

Chapter 5 reviews the recent literature about stochastic volatility and builds on the works of Nelson which reconcile continuous and discrete volatility processes. The authors use the Extended Kalman Filter to deal with the issue of the unobserved volatility of the yield curve. The authors also introduce Bollinger bands as a brand-new variance reduction technique for improving the Monte Carlo performance; a technique never applied before to yield curve forecasting.

Chapter 6 examines the modern credit risk valuation which focuses on the soundness of the risk assessment process since Basel II directives. Any risk assessment requires comprehending the volatility of credit risky assets with accuracy. For this purpose, the authors state a flexible credit risk valuation framework while allowing such a volatility to evolve stochastically. Hence, the structural approach of credit risk along with the modern option pricing theory allows for an interesting and flexible stochastic credit risk valuation framework.

Chapter 7 investigates simple intensity models that induce dependence levels comparable to those induced by a Merton-style model using a simulation model. The authors compare the respective loss distributions obtained in each framework and provide some dependence indicators. Moreover, they specify two promising and original intensity-based models that emphasize their results: correlated frailty and alpha-stable distributions.

Chapter 8 discusses various mathematical techniques that can be used for the modelling of weather derivatives portfolios. In particular, the authors describe extensions to the most commonly used simulation algorithm. These extensions include methods that improve estimates of the correlation structure, deal with non-normality, incorporate hedging constraints, estimate sampling error, allow consistency between single contract pricing and portfolio modeling, and give quick estimates of VaR.

Chapter 9 links nominal interest payments (as in typical bond contracts) with the demand for real payments (as in pension contracts), and models for the inflation and for valuing inflation linked products. Here, the authors introduce a simple continuous-time framework that is economically justified and similar to the Garman–Kohlhagen model for foreign currencies. It allows for valuation of inflation-linked derivatives, optimal investment into such products and hedging of inflation risk. Explicit solutions for all these tasks are provided and permit an easy implementation and calibration in real world markets.

Chapter 10 examines the explosive growth in the use of financial models in recent years that has allowed for the creation of more diverse financial products and the development of new markets for such products. However, it also has some drawbacks, such as the creation of a new type of risk called model risk. The latter arises as a consequence of incorrect modelling, model identification or specification errors, inadequate estimation procedures, as well as mathematical and statistical properties of financial models applied in imperfect financial markets. Although models vary in their sophistication, they all need to be subjected to an effective validation process to minimize the risk of model errors.

Chapter 11 investgates the crucial question among risk managers and regulators; whether Value-at-Risk models are accurate enough. The authors propose a methodology based on a cross-section analysis of portfolios, aimed to assess the goodness of VaR using a simultaneous analysis of a multitude of simulated portfolios, created starting from a common investment universe. This enhances the exploitation of the information content of data, broadening the perspective of risk assessment.

Chapter 12 analyses the shocks in correlations that could significantly alter outcomes in portfolio optimization and risk management estimates. The chapter examines the relation between exponential correlation changes and volatility for the different movements of markets and studies the magnitude of errors among equity investments in the USA, the Euro area and Japanese markets.

Chapter 13 explores the historical values of the asset returns process, from which is derived the sequential control procedures for monitoring the changes in the covariance matrix of asset returns that could influence the selection of an optimal portfolio. In order to reduce the dimensionality of the control problem we focus essentially on the transformation of the optimal portfolio weights vector.

Chapter 14 reiterates the notion whereby one of the factors that contributes to the portfolio diversification benefit is the correlation between the asset returns. Correlations are time varying and the traditional method of using unconditional correlations in portfolio optimization models may not capture the time-varying nature of asset return correlations. In this chapter the authors compare the *ex post* performance of portfolios created using unconditional correlations against those created using Dynamic Conditional Correlation (DCC). The results using 20 stocks from the Dow Jones Industrial Average show that portfolios created using the DCC model outperformed those created using the unconditional correlations.

Chapter 15 deals with the evaluation of risky capital investment projects when total risk is relevant. The authors demonstrate mathematically that the NPV probability distribution does not conform strictly to the central limit theorem asymptotic properties, whereas first-order autoregressive stochastic stationary processes do. However, through simulation runs and statistical tests, the authors show under realistic conditions that the CLT does apply to the NPV probability distribution provided the discount rate does not exceed some threshold value.

Chapter 16 analyses the volatility transmission between the US and Spanish stock markets using a recent sample period including September 11. The analysis is based on a multivariate GARCH model which takes into account both the asymmetric volatility phenomenon and the non-synchronous trading problem. An examination of Asymmetric Volatility Impulse-Response Functions (AVIRF) confirms that volatility transmission patterns between both markets have changed as a result of the terrorist attacks.

Chapter 17 examines the volatility transmission between large and small firms in Europe using Germany, France and UK stockmarket data. The empirical results indicate that volatility spillovers take place between both kinds of firms and that the volatility feedback hypothesis can explain asymmetric volatility and covariance. Additionally, evidence is obtained showing that in order to avoid error specification in the beta coefficient, it is necessary to use a conditional model.

Chapter 18 analyzes the impact of model misspecification on the replication error associated with trading contingent claims in arbitrage free markets. A general formula is determined for the total hedging error in the light of stochastic volatility and numerical tests are performed on European options to estimate the replication error probability density function.

CHAPTER 1

Impact of the Collection Threshold on the Determination of the Capital Charge for Operational Risk

Yves Crama, Georges Hübner and Jean-Philippe Peters*

1.1 INTRODUCTION

In 2004, the Basel Committee on Banking Supervision (hereafter the Basel Committee) released the Revised Framework of the International Convergence of Capital Measurement and Capital Standards (hereafter Basel II). Together with new rules governing the calculation of regulatory capital charge for credit risk, Basel II introduces explicit recommendations with regard to *operational risk*, defined by the Basel Committee as the "risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. This definition includes legal risk, but excludes strategic and reputational risk" (BCBS, 2004).

* Georges Hübner gratefully acknowledges financial support from Deloitte Luxembourg and the Luxembourg School of Finance. This paper has won the Operational Risk & Compliance Achievement Award 2006, hosted by *Operational Risk Magazine*, in the best academic paper category.

Basel II leaves the choice between three approaches for quantifying the regulatory capital for operational risk. Both the Basic Indicator Approach (BIA) and the Standardized Approach (SA) define the operational risk capital of a business line as a fraction of its gross income, thus explicitly assuming that operational risk is related to size. Under the Advanced Measurement Approach (AMA), banks can develop their own model for assessing the regulatory capital that covers their operational risk exposure over a one-year period within a confidence interval of 99.9 percent (henceforth Operational Value at Risk, or OpVaR). They must apply this model for each of the eight Business Lines and for each of the seven Loss Event Types defined in the Revised Framework. By default, capital charges associated to all 56 combinations are added to compute the regulatory capital requirement for operational risk.[1]

Although operational risk has been the focus of much attention in the manufacturing industry for several decades, most financial institutions have had a tendency to neglect this heterogeneous family of risks which, except for fraud, are often perceived as diffuse and peripheral. For the same reasons, until recently, very few banks had set up systematic procedures for the collection of data relative to operational losses. As a consequence of Basel II, however, many banks are now in the process of setting up a sound and homogeneous loss data collection system for all types of risks.

A question that often arises when implementing a loss data collection process is the determination of the collection threshold. Recording *all* the operational loss events is indeed impossible, or at least wasteful, as the cost (in terms both of systems and time) of the process would be much too high in regard to its potential benefits. Therefore, banks are led to fixing a minimum collection threshold under which losses are not collected.

While the literature on operational risk modeling is booming (see for example, Frachot, Georges and Roncalli (2001), Cruz (2002), Alexander (2003), Fontnouvelle, Jordan and Rosengren (2003), Fontnouvelle, Rosengren and Jordan (2004), Moscadelli (2004), or Chapelle, Crama, Hübner and Peters (2005)), few studies have paid specific attention to the choice of the collection threshold for operational risk modeling and to its impact on the capital charge.

This chapter examines the tradeoff between the cost of collecting data from a very low money value and the loss of information induced by a higher threshold. It is organized as follows. In section 1.2, we introduce the LDA method to model operational risk losses. Next, we discuss the loss data collection process, the related choice of the collection threshold and its impact on estimated parameters. Section 1.4 uses real life data to examine the impact of the collection threshold on the value of the capital charge for operational risk. Section 1.5 contains some conclusions.

1.2 MEASURING OPERATIONAL RISK

1.2.1 Overview

Although the application of AMA is in principle open to any proprietary model, the most popular methodology is by far the Loss Distribution Approach (LDA), a parametric technique that consists in separately estimating a frequency distribution for the occurrence of operational losses and a severity distribution for the economic impact of the individual losses (see for example, Klugman, Panjer and Willmott, 1998; Frachot, Georges and Roncalli, 2001; or Cruz, 2002). Both distributions are then combined through an n-convolution of the severity distribution with itself, where n is a random variable that follows the frequency distribution (see Frachot, Georges and Roncalli, 2001, for details).

The output of the LDA methodology is a full characterization of the distribution of annual aggregate operational losses of the bank. This distribution contains all relevant information for the computation of the regulatory capital charge to cover operational risk, as this capital charge is obtained by subtracting the expected loss from the 99.9 percent quantile of the distribution.[2]

1.2.2 Loss distribution approach

In this section, we discuss the methodological treatment of a series of internal loss data for a single category of risk events, so as to construct a complete distribution of operational losses.

As mentioned before, the LDA separately estimates the frequency and severity distributions of losses. The aggregate distribution of losses is then obtained by an n-fold convolution of the severity distribution with itself, where n is the (random) number of observations obtained from the frequency distribution. As an analytical solution to this problem is extremely difficult to derive in practice, we compute this convolution by Monte Carlo simulations. A precise overall characterization of both distributions is required to achieve a satisfactory level of accuracy.

Maximum Likelihood Estimation (MLE) techniques can be used to estimate the parameters of both distributions. From a statistical point of view, the MLE approach is considered to be the most robust and it yields estimators with good statistical properties (consistent, unbiased, efficient, sufficient and unique[3]).

More precisely, let $f(x;\theta)$ be a selected parametric density function, where θ denotes the vector of parameters, and let $F(x;\ \theta)$ be the cumulative distribution function (or CDF) associated to $f(x;\theta)$. Then, the corresponding log-likelihood function is

$$\ell(x;\theta) = \sum_{i=1}^{N} \ln(f_i(x_i;\theta)) \tag{1.1}$$

where $(x_1, \ldots, x_N)$ is the sample of observed ordinary losses. The maximum likelihood estimates of the parameters θ_j are obtained by solving the system of equations

$$\frac{\delta \ell}{\delta \theta_j} = 0$$

Frequency distribution

The frequency distribution models the occurrence of operational loss events recorded by the bank. This distribution is by definition discrete. It is frequently modeled either as a homogeneous Poisson or as a (negative) binomial distribution. The choice between these distributions may appear important as the intensity parameter is deterministic in the first case and stochastic in the second (see Embrechts *et al.*, 2003).

The mass function of the Poisson distribution is

$$\Pr(N = x) = \frac{e^{-\lambda}\lambda^x}{x!} \tag{1.2}$$

where λ is a positive integer. It can easily be estimated as λ is equal to both the mean and the variance of the Poisson distribution. Note also the following nice property of the Poisson distribution: if $X_1, X_2, \ldots, X_m$ are m independent random variables and $X_i \sim \text{Poisson}(\lambda_i)$, then $X_1 + X_2 + \cdots + X_m \sim \text{Poisson}(\lambda_1 + \lambda_2 + \cdots + \lambda_m)$.

The binomial distribution is given by

$$\Pr(N = x) = \binom{m}{x} p^x (1-q)^{m-x} \tag{1.3}$$

where $\binom{m}{x}$ is the binomial coefficient defined as $\frac{m(m-1)\ldots(m-x+1)}{x!}$, $p \in (0,1)$ and m is a positive integer. Contrary to the Poisson case, the mean is not equal to the variance for this distribution, as mean $= mp$ and variance $= mp(1-p)$. It follows that the mean is larger than the variance for the binomial distribution.

Finally, the negative binomial distribution has the following mass function

$$\Pr(N = x) = \binom{x+r-1}{x} p^r (1-p)^x \tag{1.4}$$

where $p \in (0,1)$ and r is a positive integer. The relationship between mean and variance is the opposite of the binomial as mean $= \frac{r(1-p)}{p}$ and variance $= \frac{r(1-p)}{p^2}$. Thus, the mean is smaller than the variance for the negative binomial distribution.

A good starting point to determine the most adequate frequency distribution is therefore to check the relationship between mean and variance of the observed frequency. If the observed variance is much higher (resp. lower) than the observed mean, a negative binomial (resp. binomial) distribution could be well-suited to model frequency.

Other techniques to discriminate between these distributions include goodness-of-fit tests such as the χ^2 test. The idea of this test is to split the population into k adjacent "classes" of equal width, and then to compute the following statistic:

$$\chi^2 = \sum_{j=1}^{k} \frac{(n_j - E_j)^2}{E_j}$$

where n_j is the number of elements observed in class j and E_j is the theoretical expected number of observations in the class. This test should be interpreted as follows: the lower χ^2, the better the fit.

If H_0 is true (for example, the observed series follows the tested distribution), χ^2 converges to the distribution function that lies between the chi-square distributions with $k-1$ and $k-m-1$ degrees of freedom (where m is the number of estimated parameters). Thus if $\chi^2 > \chi^2_{k-1,1-\alpha}$ where $\chi^2_{k-1,1-\alpha}$ is the upper $1-\alpha$ quantile of the asymptotic chi-square distribution, the null hypothesis is rejected.[4] Finally, a rule of thumb to decide the number of bins is that $k \geq 3$ and $E_j \geq 5$ for all j.

Severity distribution

The severity distribution models the economic impact of operational risk loss events. Consequently, any strictly positive continuous distribution can be used to model operational losses. However, operational risk databases are often characterized by a large bulk of "high frequency/low impact" losses and a few "low frequency/high impact" losses. Leptokurtic distributions are thus most appropriate to model the severity distribution. Candidate distributions include log-normal, log-logistic, Pareto or Weibull distributions. Table 1.1 summarizes the probability distribution functions (PDF) of these distributions.[5]

To test the adequacy of the estimated distribution for the observed values, goodness-of-fit statistics can again be calculated, for example by the

Table 1.1 Severity distributions

Distribution	Probability distribution function
Lognormal(μ, σ)	$f(x) = \frac{1}{x\sqrt{2\pi}} \exp\left[\frac{-(\log x)^2}{2}\right]$
LogLogistic(α, β)	$f(x) = -\frac{\alpha(x/\beta)^{\alpha-1}}{\beta[1+(x/\beta)^{\alpha}]^2}$
Pareto(θ, α)	$f(x) = \alpha\theta^{\alpha}x^{-(\alpha+1)}$
Weilbull(α, β)	$f(x) = \alpha\beta^{-\alpha}x^{\alpha-1}\exp\left(-\left(\frac{x}{\beta}\right)^{\alpha}\right)$

Kolmogorov–Smirnov test (Kolmogorov, 1933, and Smirnov, 1939) defined by the statistics,

$$D_{KS} = \max_{i=1,\ldots,n} [|F_n(x_i) - F(x_i;\theta)|] \tag{1.5}$$

This test does not depend on the underlying CDF being tested. On the other hand, it has several drawbacks: it is only available for continuous distributions, the distribution must be fully specified and it is more sensitive near the center than at the tails, which makes it somewhat conservative. In the operational risk framework, the first two issues are not problematic but the last one should be a source of concern. Severity distributions are usually heavy-tailed and a good fit at the extreme right tail of the density is crucial.

The Cramer–von Mises test (Cramer, 1928) is quite similar to the KS test, but introduces a size-based correction. It is defined by the statistics:

$$CVM = n\int_{-\infty}^{\infty} (F_n(x) - F(x;\theta))^2 \mathrm{d}F(x) \tag{1.6}$$

which in practice can be computed as

$$CVM = \frac{1}{12n} + \sum_{i=1}^{n} (F_n(x_i) - F(x_i;\theta))^2$$

where $F(x_i)$ is the cumulative distribution function value at x_i, the i-th ordered value.

Very often, the fat-tailed behavior of operational losses makes the accurate estimation of the (regulatory required) extreme quantiles a tricky exercise. Consequently, heavy-tailed distributions such as the ones presented above are sometimes unable to correctly capture the probability of occurrence of exceptional losses (for example, the extreme right part of the severity distribution). This is especially true since loss data collection is sometimes still in its infancy at some banks, which results in internal loss databases lacking very large losses. Some recent studies indeed indicate that classical distributions are unable to fit the entire range of observations in a realistic manner (see Fontnouvelle, Rosengren and Jordan, 2004, or Chapelle, Crama, Hübner and Peters, 2005).

As a consequence, numerous authors propose to use alternative approaches to improve the accuracy of the tail modeling. One of the most common approaches relies on Extreme Value Theory (EVT), which is presented in next section.

Modeling extreme losses

Extreme Value Theory (EVT) is a powerful theoretical tool to build statistical models describing extreme events. It has been developed to answer the crucial question: if things go wrong, how wrong can they go?

Two techniques are available: the Block Maxima method and the Peak Over Threshold (POT) method. While the origins of the former date back in the early twentieth century, it has been presented in a general context by Gumbel (1958). It focuses on the modeling of maxima of different periods, such as month or year (for example, the p observations are the maximum observed value of each of the p periods considered). These extremes are then modeled with the Generalized Extreme Value (GEV) distribution. While useful in domains such as climatology, the Block Maxima approach is less attractive for financial applications.

The POT approach builds upon results of Balkema and de Haan (1974) and Pickands (1975) which state that, for a broad class of distributions, the values of the random variables above a sufficiently high threshold follow a Generalized Pareto Distribution (GPD) with location parameter μ, scale parameter β and shape parameter ξ (also called the tail index). The GPD can thus be thought of as the conditional distribution of X given $X > \mu$ (see Embrechts *et al.*, 1997, for a comprehensive review). Its cdf can be expressed as:

$$F(x;\xi,\beta,\mu) = 1 - \left(1 + \xi\frac{(x-\mu)}{\beta}\right)^{\frac{-1}{\xi}} \tag{1.7}$$

A major issue when applying POT is the determination of the threshold μ. A standard technique is based on the visual inspection of the Mean Excess Function (MEF) plot (see Davidson and Smith, 1990, or Embrechts, Klüpperberg and Mikosch, 1997, for details). This graph plots the empirical mean excess, defined as:

$$e(u) = \frac{1}{n_u}\sum_{i=1}^{n_u}(x_i - u)$$

where the x_i's are the n_u values of X such that $x_i > u$. The MEF plot is a plot of $e(u)$ against u. The method is to detect a significant shift in slope at some high point. When the empirical plot seems to follow a reasonably straight line with positive gradient above a certain value, this indicates a heavy-tailed distribution.

The visual inspection of MEF is sometimes tricky as no (or several) "break(s)" can be observed. Several authors have suggested methods to identify the optimal threshold (see, for example, Drees and Kaufmann, 1998; Dupuis, 1999; Matthys and Beirlant, 2003) but no single approach has become widely accepted. A possible solution is proposed in Chapelle, Crama, Hübner and Peters (2005) with an algorithmic procedure that builds on ideas from Huisman, Koedijk, Kool and Palm (2001) and shares some similarities with a procedure used by Longin and Solnik (2001) in a different context. The Appendix summarizes the various steps of this algorithm.

1.3 THE COLLECTION THRESHOLD

1.3.1 Selection of a threshold

A sound loss data collection process is key to operational risk management and measurement as statistical inference based on historical internal loss data *and* monitoring/reporting activities both heavily rely on the quality of the collected data. Coherence and completeness of collected data amongst business units is therefore crucial.

Selecting the most adequate collection threshold is obviously bank-specific, as each bank will examine the tradeoff between increasing the number of observations in its internal database and the associated increase in costs.

In addition to cost issues, reporting very low losses is likely to be viewed as a waste of time by the employees. When this is the case, adhesion of the employees is hard to obtain and the reliability of the collection process can be questioned. On the measurement side, this results in an incomplete database and the accuracy of the capital charge estimation is not ensured.

In contrast, however, fixing a very high threshold creates a truncation bias that can lead to an over-estimation of the severity (see Frachot, Moudoulaud and Roncalli, 2003).

To determine an adequate threshold, some banks rely on indications given by Basel II, which recommends setting the collection threshold at 10,000 EUR.[6] For banks that are members of a data collection consortium, the decision is sometimes driven by the rules of the consortium:

- The Italian initiative DIPO led by the ABI (the Italian Bankers' Association) requires banks to provide all their operational risk losses above a threshold fixed at 5,000 EUR.
- ORX is a private consortium comprising large internationally active banks. It has fixed the reporting threshold at 20,000 EUR.

For smaller banks, however, fixing a threshold at 10,000 EUR might drastically reduce the amount of data available for computing the capital requirements. A threshold of 1,000 EUR or 5,000 EUR can be more adequate.

Whatever the final choice, statistical methods used to calculate the regulatory capital charge for operational risk should be adapted to account for this threshold. This issue is discussed in the following section, while an analysis of the impact of the collection threshold on the value of the capital charge is provided in section 1.4.

1.3.2 Impact of the collection threshold on the estimated parameters

As noted by Frachot, Moudoulaud and Roncalli (2003):

> the data collection threshold affects severity estimation in the sense that the sample severity distribution (for example, the severity distribution of reported losses) is different from the "true" one (for example, the severity distribution one would obtain if all losses were reported). Unfortunately, the true distribution is the most relevant for calculating capital charge and also for being able to pool different sources of data in a proper way. As a consequence, linking the sample distribution to the true one is a necessary task.

Mathematically, this is a well-known phenomenon referred to as "truncation". More precisely, the density function $f*(x;\theta)$ of the losses in $[L;\infty)$ can be expressed as:

$$f^*(x;\theta) = \frac{f(x;\theta)}{1 - F(L;\theta)}$$

where $f(x;\theta)$ is the complete (non truncated) distribution on $[0;\ \infty)$. The corresponding log-likelihood function is:

$$\ell(x;\theta) = \sum_{i=1}^{N} \ln\left(\frac{f_i(x_i;\theta)}{1 - F(L;\theta)}\right) \tag{1.8}$$

where $(x_1,\ldots,x_N)$ is the sample of observed losses and L is the collection threshold. It must be maximized in order to estimate θ.

Usually, the quality of distribution fitting is assessed through goodness-of-fit tests. All these tests are based on a comparison between the observed cumulative distribution function and the hypothetical one. Consequently, they should be adjusted to account for the collection threshold as well. For instance, the Kolmogorov–Smirnov statistics becomes:

$$D_{KS} = \max_{i=1,\ldots,n}\left[\left|F_n(x_i) - \frac{F(x_{i;\theta})}{F(U;\theta)}\right|\right] \tag{1.9}$$

To show how spurious the estimates of the parameters or the goodness-of-fit test can be when the collection threshold is not accounted for, consider the generation of 10,000 random variables that follows a Weibull (0.001, 0.68) distribution. Table 1.2 reports three cases:

- In Case 1, the whole series is considered (for example, there is no collection threshold) and the parameters are estimated by the Maximum Likelihood technique.
- In Case 2, we only consider losses larger than 1,000. Parameters are also estimated by MLE and we do not modify the likelihood function to be optimized (for example, we ignore the collection threshold).

Table 1.2 Adjusted parameters estimation for truncated distribution

	Case 1	Case 2	Case 3
N	10,000	8,910	8,910
a	0.0011	0.0002	0.0012
b	0.673	0.815	0.668
KS	0.0060	0.0571	0.1110
KS*	–	0.0754	0.0047

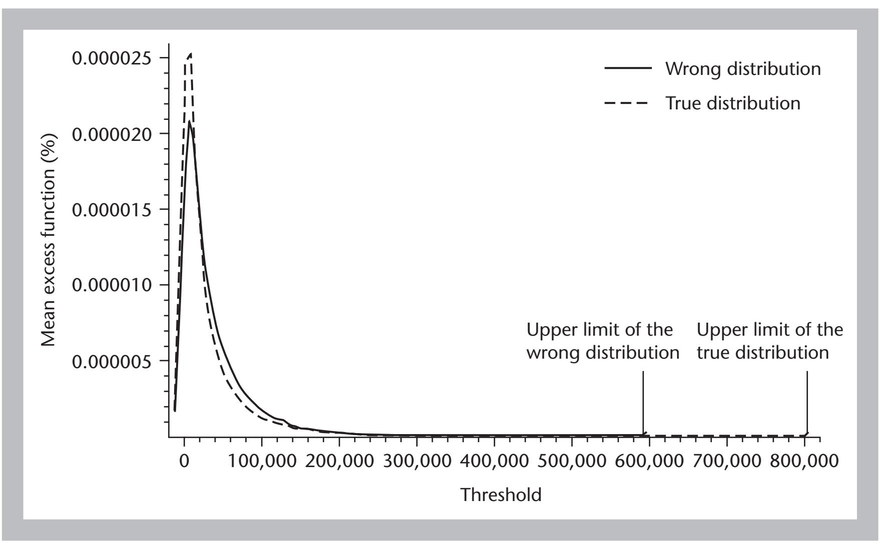


Figure 1.1 Impact of the truncation on estimated distributions

- In Case 3, we also consider losses larger than 1,000, but we adjust the likelihood function according to equation (1.8).

For each case, we also compute the Kolmogorov–Smirnov test. In Table 1.2, KS relates to the unmodified test (for example, not accounting for the collection threshold), while KS* is the modified test.

The table clearly demonstrates the importance of accurately adjusting the estimation techniques to account for the collection threshold. Without adequate changes in the likelihood function and the goodness-of-fit statistics, fallacious conclusions could be drawn as the parameters estimated in Case 2 (with KS = 0.06) could be preferred to those estimated in Case 3 (with KS = 0.11). This would in turn lead to inaccurate Monte Carlo results, as both distributions are very different. Figure 1.1 reports both distributions and clearly shows that failing to adapt the estimation procedure to account for truncation may have a significant impact. The estimated distribution in Case 2 has an upper limit that is 25 percent smaller than the true distribution, seriously impacting subsequent simulations.

1.4 EMPIRICAL ANALYSIS

1.4.1 Data

In this section, we apply the methodology outlined in the previous sections to real operational loss data provided by a large European bank. For this study, we focus our analysis on two complete business lines, regardless of the loss event type.[7] For the sake of confidentiality, we call these business lines "BL1" and "BL2". For the same reasons, we have scaled all loss amounts by a same constant. The summary statistics of losses are given in Table 1.3.

Table 1.3 Summary statistics for the operational loss database

	Business line 1	**Business line 2**
No. observations	1,666	7,841
Collection threshold	0.25	0.25
Median loss	1.35	0.94
Mean loss	118.7	20.7
Std. dev.	1,813	256
Total loss	197,707	162,034

1.4.2 Calibration of AMA

First, we consider both business lines with a collection threshold of 0.25. Preliminary analysis indicates that frequencies of both samples are well described by a Poisson process. As this distribution is characterized by a single parameter which is the average of the observed frequency, we use a Poisson (1666) and a Poisson (7841) to model frequency for Business Line 1 (hereafter BL1) and Business Line 2 (hereafter BL2), respectively.

To model severity, we start by applying a single PDF for the whole distribution. Based on the Cramer–von Mises test, the most adequate distributions to model BL1 and BL2 among those presented in Table 1.1 are a Weibull (4.9, 0.09) and a Weibull (8.5, 0.07), respectively. But as often encountered with operational risk losses, even these distributions are unable to satisfactorily capture the whole distributional form, especially at the tail level. Figure 1.2 shows the QQ-plot for both cases. Points in the tail clearly depart from the straight line that would indicate a good fit. To circumvent this problem, we adopt the approach described in the modeling extreme losses section by using EVT to model the tail of the severity distributions.

To estimate the cut-off point from which observations are used to estimate the parameters of the GPD distribution, we first take a look at the Mean Excess Plots (see Figure 1.3). Visual inspection indicates potential "breaks" around 400 for BL1 and 500 for BL2.

To validate this first impression, we apply the algorithm described in the Appendix. For both business lines, we consider all the observations above 100 to be potential threshold candidates. This means that $m = 59$ for BL1 and $m = 227$ for BL2. The threshold values for which MSE is minimized are 375 and 450, for BL1 and BL2, respectively.[8] MSE associated with each tested threshold are plotted in Figure 1.4.

Table 1.4 reports the estimation of the three parameters for the GPD. Location parameter is estimated through the algorithm described in the Appendix, while scale and shape parameters are estimated with (constrained) MLE.[9]

The next step is to model the "body" of the distribution, for example, the losses that are below the estimated extreme threshold. For BL 1, this means all the losses between the collection threshold (0.25 in this case) and 375. For BL 2, this covers all the losses between 0.25 and 450. To do so, we consider the distributions presented in Table 1.1 and we use the Cramer–von Mises statistic as a discriminant factor to compare goodness of the fit. Once the severity is fully characterized, 10,000 Monte Carlo simulations are performed to derive the aggregate loss distribution for each business line. Results are summarized in Table 1.5. The regulatory capital charge amounts to 1.3 and 0.8 million for BL1 and BL2, respectively. Additionally, it is interesting to note that these values represent 8.1 and 6.1 times the

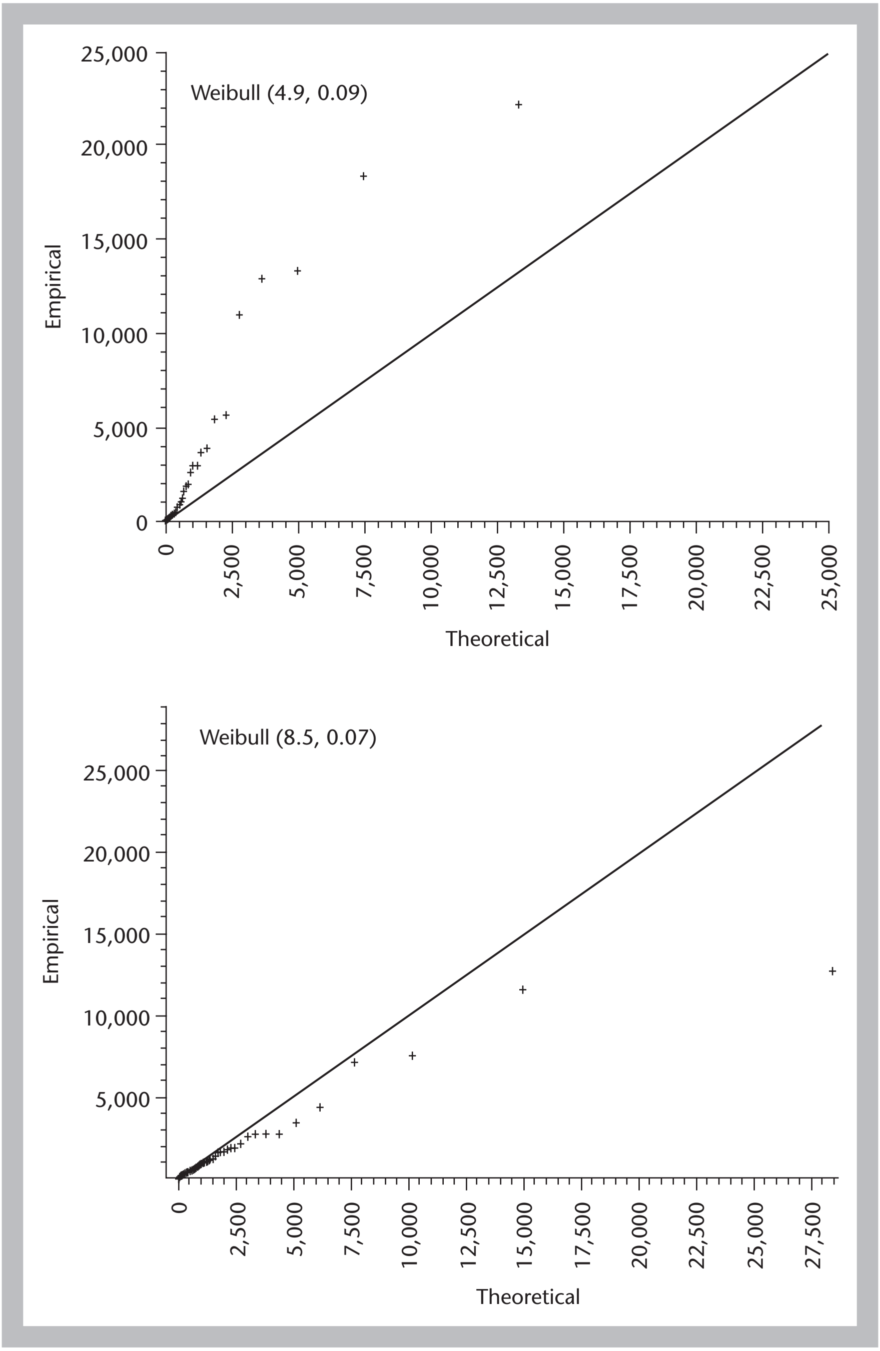


Figure 1.2 QQ-plots for BL1* (above) and BL2 (below)
*For sake of visual clarity, the largest loss has been removed from the first graph

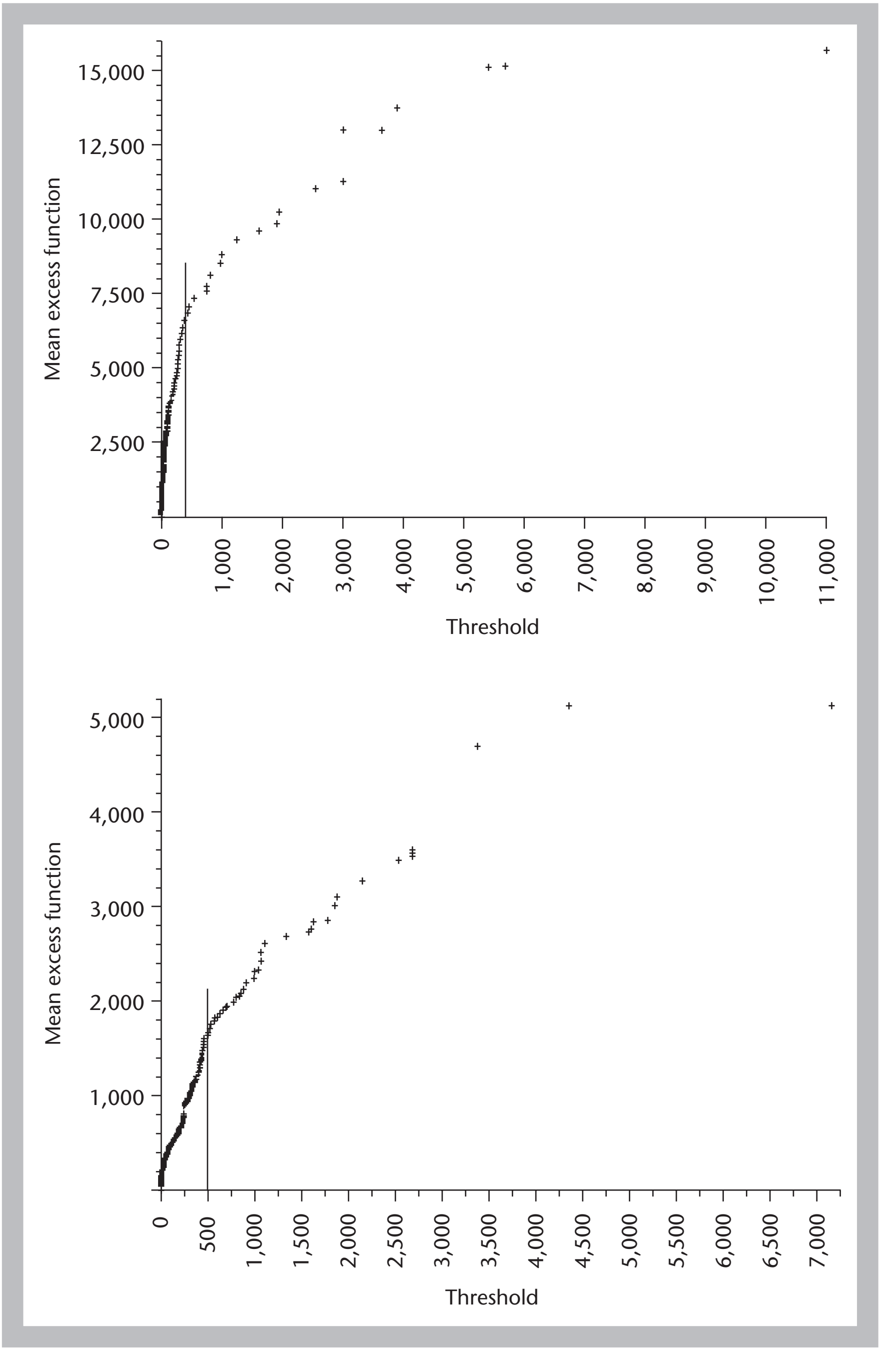


Figure 1.3 Mean excess plots for BL1 (above) and BL2 (below)*
*For the sake of visual clarity, the 5 and 2 largest losses have been removed from the MEF plots of BL1 and BL2, respectively

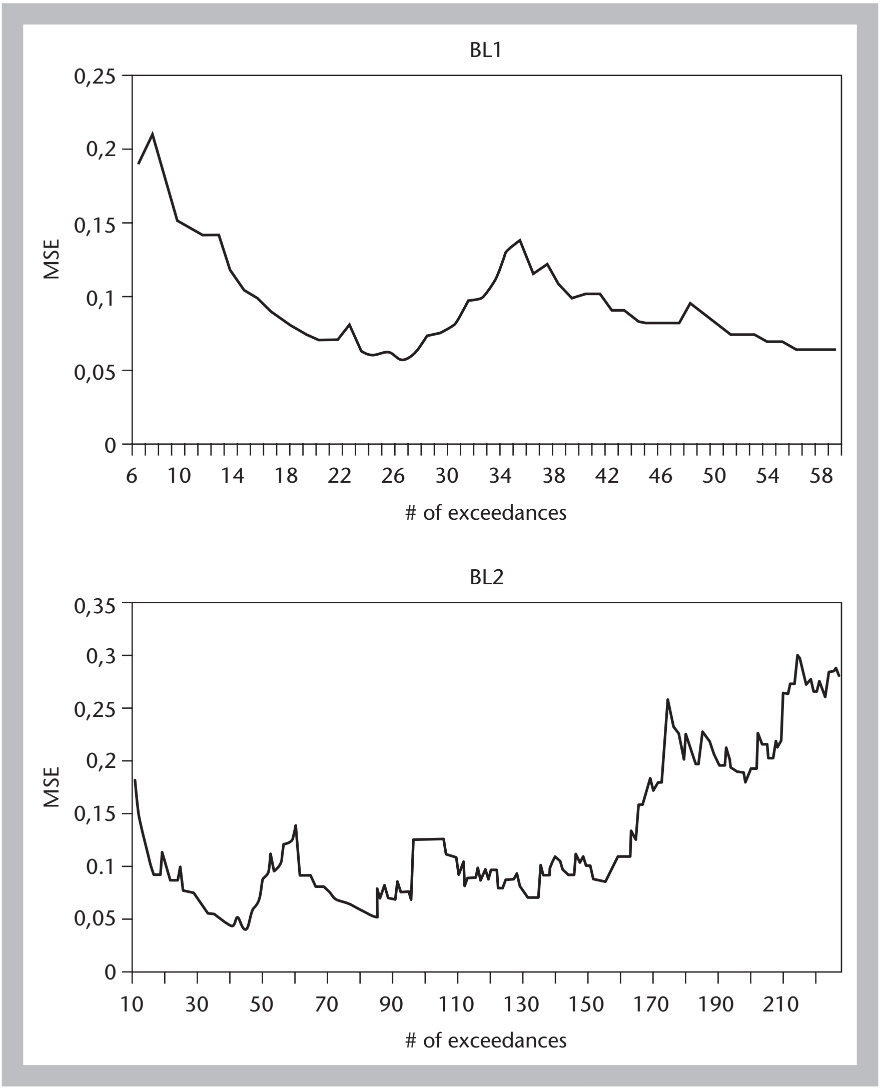


Figure 1.4 MSE for threshold candidates for BL1 (above) and BL2 (below)

Table 1.4 Estimated parameters for the GPD

	Business line 1	**Business line 2**
Optimal threshold	375	450
GPD – shape	1.041	0.750
GPD – scale	1677	601
Mean square error	0.0588	0.0402

Table 1.5 Summary results for BL1 and BL2 (collection threshold = 0.25)

	Business line 1	**Business line 2**
Frequency	*Poisson (1666)*	*Poisson (7841)*
"Body"	Weibull	Weibull
Parameter #1	3.056	11.852
Parameter #2	0.147	0.047
"Tail"	GPD	GPD
Location parameter	375	450
Shape parameter	1.041	0.750
Scale parameter	1677	601
Expected loss	282,429	179,893
Unexpected loss	1,596,046	990,471
Regulatory capital charge[10]	1,313,617	810,578

observed losses, respectively. The higher value for Business Line 1 is not surprising as BL1 is characterized by a higher proportion of severe losses. This is reflected in the aggregate loss distribution with a fatter tail and, consequently, proportionally heavier capital requirements.

1.4.3 Sensitivity analysis

Next, we move on to the analysis of the impact of the collection threshold on the regulatory capital charge. We thus apply the approach of section 1.4.2 to both business lines when considering 6 collection thresholds: 0.25, 1, 5, 10, 20 and 50. In all cases, using a single distribution to fit the entire severity distribution appears to perform badly when it comes to model its extreme right part. As a consequence, we use EVT in all 12 cases. Results are provided in Table 1.6.

With all this information, Monte Carlo simulations are performed and 10,000 years of losses are simulated. The resulting 10,000 aggregate losses form the aggregate loss distribution from which the capital charge is derived.

Table 1.7 summarizes the main results, while Figure 1.5 provides a visual overview of the fluctuation of the capital charge depending on the collection threshold.

1.5 CONCLUSION

This chapter has provided three kinds on evidence on the impact of the choice of the collection threshold for operational losses that have

Table 1.6 Estimated frequency and severity distributions

	Business Line 1					
Collection threshold	0.25	1	5	10	20	50
No. obs.	1,666	979	427	324	190	95
Frequency	Poisson	Poisson	Poisson	Poisson	Poisson	Poisson
Body	Weibull	Weibull	Log-Normal	Weibull	Pareto	Pareto
Parameter #1	3.056	2.447	1.658	14.249	20	50
Parameter #2	0.147	0.174	1.768	0.048	0.797	0.776
% of extremes	1.6%	2.7%	6.1%	8.0%	13.7%	27.4%
GPD – location	375	375	375	375	375	375
GPD – shape	1.041	1.041	1.041	1.041	1.041	1.041
GPD – scale	1677	1677	1677	1677	1677	1677
	Business Line 2					
Collection threshold	0.25	1	5	10	20	50
No. obs.	7,841	3,754	1,348	901	593	359
Severity	Poisson	Poisson	Poisson	Poisson	Poisson	Poisson
Body	Weibull	Pareto	Pareto	Pareto	Pareto	Pareto
Parameter #1	11.852	1	5	10	20	50
Parameter #2	0.047	0.551	0.514	0.492	0.402	0.375
% of extremes	0.6%	1.2%	3.3%	4.9%	7.4%	12.3%
GPD – location	450	450	450	450	450	450
GPD – shape	0.750	0.750	0.750	0.750	0.750	0.750
GPD – scale	601	601	601	601	601	601

not been previously documented in the operational risk management literature.

First, the level of the collection threshold has little impact on regulatory capital charge estimations. The own funds needed to cover operational risk are indeed stable in both business lines. For BL1, it ranges from 1.26 million to 1.43 million depending on the threshold. The variation range is narrower for BL2 as capital charge fluctuates between 0.78 and 0.85 million.

Second, this result is mainly due to the way the tail is modeled. As we rely on EVT to model the very high losses, the collection threshold has no or little impact on the fatness of the tail for the severity distribution.

Finally, the choice of the collection threshold should thus not be guided by capital requirements concerns but rather by a "pro/cons" analysis of the practical implementation issues (costs, required systems, resources…) as regulatory arbitrage seems not to be applicable in this case.

Table 1.7 Statistics from the aggregate loss distributions

			Business Line 1			
Collection Threshold	0.25	1	5	10	20	50
Observed total loss	197,707	197,362	196,162	195,469	193,577	190,613
Median	207,382	206,234	202,775	203,900	193,577	197,428
Mean (EL)	282,429	283,843	275,686	276,033	274,708	273,983
VaR_90	569,446	571,079	554,891	551,029	549,296	559,235
VaR_95	772,580	782,968	745,494	742,276	744,448	758,515
VaR_99	1,129,687	1,156,350	1,113,569	1,123,954	1,128,338	1,149,638
VaR_999 (UL)	1,596,046	1,664,351	1,540,889	1,709,335	1,624,079	1,589,754
VaR_9995	1,990,333	1,841,773	1,742,146	1,791,703	1,801,347	1,763,800
Capital Charge (UL-EL)	**1,313,617**	**1,380,508**	**1,265,203**	**1,433,302**	**1,349,371**	**1,315,771**
UL/Observed Loss	8.07	8.43	7.86	8.74	8.39	8.34
			Business Line 2			
Collection Threshold	0.25	1	5	10	20	50
Observed total loss	162,034	159,946	154,747	151,670	147,554	140,389
Median	159,668	163,497	174,317	153,049	151,411	144,668
Mean (EL)	179,893	182,426	155,161	172,279	172,345	164,182
VaR_90	242,829	243,356	235,854	233,870	234,269	226,114
VaR_95	305,361	292,105	285,729	284,993	297,339	282,444
VaR_99	520,023	513,738	505,102	526,324	548,432	526,876
VaR_999 (UL)	990,471	997,423	1,007,766	1,012,012	1,010,079	952,224
VaR_9995	1,017,129	1,108,098	1,070,175	1,077,630	1,100,504	1,017,359
Capital Charge (UL-EL)	**810,578**	**814,997**	**852,605**	**839,733**	**837,734**	**788,042**
UL/Observed total loss	6.11	6.24	6.51	6.67	6.85	6.78

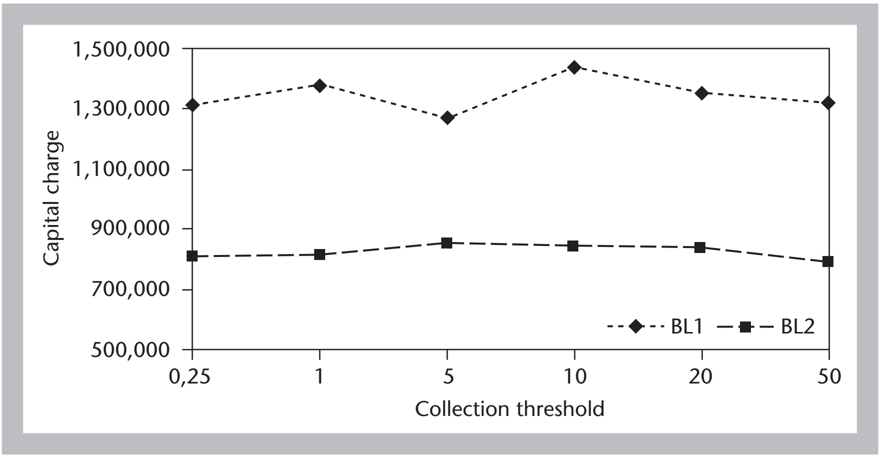


Figure 1.5 Regulatory capital charge for BL1 and BL2 with various collection thresholds

APPENDIX: FINDING THE EVT THRESHOLD

The algorithm performs the following steps:

1 Let $(x_1, \ldots, x_n)$ be the ordered sample of observations. Consider m candidate thresholds $U_1, \ldots, U_m$ such that $x_{n-i}, \ldots, x_n > U_i$ for $i = 1, \ldots, m$.

2 For each threshold U_i, use the weighted average of Hill estimators proposed in Huisman *et al.* (2001) to estimate the tail index ξ_i of the GPD distribution. This method corrects for the small-sample bias of the original Hill estimator.

3 Then compute the maximum likelihood estimator of the scale parameter β_i of the GPD, with the tail index ξ_i fixed to the value obtained in step 2.

4 For each threshold U_i, compute the Mean Squared Error statistic[11] $MSE(U_i) = \frac{1}{n_i} \sum_{k=1}^{n_i} (F_k - \hat{F}_k)^2$, where n_i is the number of losses above threshold U_i, F_k is the cdf of the GPD(ξ_k, β_k, μ_k) and $\hat{F}_k$ is the empirical cdf.

5 Identify $MSE(U_{opt}) = \min(MSE(U_1), \ldots, MSE(U_m))$; U_{opt} is retained as estimator of the cut-off threshold and the fitted distribution is the GPD$(\xi_{opt}, \beta_{opt}, U_{opt})$.

NOTES

1. The Basel Committee thus assumes perfect positive dependence between operational risks; alternatively, it also allows banks to use internally defined correlations. See paragraph 669 of BCBS (2004).
2. Basel II states: "Supervisors will require the bank to calculate its regulatory capital requirement as the sum of expected loss (EL) and unexpected loss (UL), unless the bank can demonstrate that it is adequately capturing EL in its internal business practices" (BCBS, 2004, §669).

3. *Consistent* means that for large n, the estimates converge to the true value of the parameters, *Unbiased* means that for all sample sizes the parameter of interest is calculated correctly, *Efficient* means that the ML estimate is the estimate with the smallest variance while *Sufficient* indicates that is uses all the information in the observations.
4. Note that, for the Poisson distribution, χ^2converges exactly to the chi-square distributions with $k-1$ degrees of freedom. See for instance section 6.6.2 of Law and Kelton (2000) for a discussion on this test.
5. See Appendix A of Klugman *et al.* (1998) for a wider range of continuous distributions.
6. See BCBS, 2004, § 673.
7. This approach is not Basel II compliant as it assumes independence between the various loss event types of a given business line. While this should be carefully kept in mind, it does not have an impact of the results of the present study and allows us increasing the size of samples under consideration.
8. The mean square errors associated with the optimum thresholds are 0.0588 and 0.0402 for BL1 and BL2, respectively.
9. In the MLE estimation, the location parameter is fixed to the optimized value obtained with the algorithm.
10. We assume that the bank under consideration in this study accounts for expected losses in its tariff policy. Regulatory capital charge is the difference between unexpected loss (the 0.999th quantile of the aggregate loss distribution) and expected loss (the mean of the aggregate loss distribution).
11. We choose the MSE criterion because it explicitly accounts for both the bias and inefficiency effects (see Theil, 1971).

REFERENCES

Alexander, C. (2003) *Operational Risk: Regulation, Analysis and Management* (London: FT Prentice-Hall).

Balkema, A.A. and de Haan, L. (1974) “Residual Life Time at Great Age”, *Annals of Probability*, 2: 792–804.

Basel Committee on Banking Supervision (2004) *International Convergence of Capital Measurement and Capital Standards – A Revised Framework*. Basel Committee Publications No. 107, The Bank for International Settlements, Basel, Switzerland.

Chapelle, A., Crama, Y., Hübner, G. and Peters, J.P. (2005) “Measuring and Managing Operational Risk in the Financial Sector: An Integrated Framework”, Working Paper, University of Liège, Belgium.

Cramer, H. (1928) “On the Composition of Elementary Errors”, *Skandinavisk Aktuarietidskrift*, 11: 141–80.

Cruz, M.G. (2002) *Modeling, Measuring and Hedging Operational Risk* (New York: Wiley & Sons).

Dupuis, D.J. (1999) “Exceedances over High Thresholds: A Guide to Threshold Selection”, *Extremes* 1: 251–61.

Embrechts, P., Furrer, H. and Kaufmann, R. (2003) “Quantifying Regulatory Capital for Operational Risk”, Working Paper, RiskLab, ETH Zürich.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance* (Berlin: Springer-Verlag).

de Fontnouvelle, P., Jordan, J. and Rosengren, E. (2003) “Using Loss Data to Quantify Operational Risk”, Working Paper, Federal Reserve Bank of Boston.

de Fontnouvelle, P., Rosengren, E. and Jordan, J. (2004) "Implications of Alternative Operational Risk Modeling Techniques", Working Paper, Federal Reserve Bank of Boston.

Frachot, A., Georges, P. and Roncalli, T. (2001) "Loss Distribution Approach for Operational Risk", Working Paper, Groupe de Recherche Opérationnelle, Crédit Lyonnais.

Frachot, A., Moudoulaud, O. and Roncalli, T. (2003) "Loss Distribution Approach in Practice", Working Paper, Groupe de Recherche Opérationnelle, Crédit Lyonnais.

Gumbel, E.J. (1958) *Statistics of Extremes* (New York: Columbia University Press).

Huisman, R., Koedijk, K.G., Kool, C.J. and Palm, F. (2001) "Tail-Index Estimates in Small Samples", *Journal of Business and Economic Statistics*, 19(1): 208–16.

Klugman, S.A., Panjer, H.H. and Willmot, G.E. (1998) *Loss Models – From Data to Decisions* (New York: Wiley Series on Probability and Statistics).

Kolmogorov, A.N. (1933) "Sulla Determinazione Empiricadi una Legge di Distribuzione", *Giornale dell' Istituto Attuari*, 4: 83–91.

Law, A.M. and Kelton, W.D. (2000) *Simulation Modeling and Analysis*, 3rd edn (New York: McGraw-Hill).

Longin, F. and Solnik, B. (2001) "Extreme Correlation of International Equity Markets", *Journal of Finance,* 56(2): 649–76.

Matthys, G. and Beirlant, J. (2003) "Estimating the Extreme Value Index and High Quantiles with Exponential Regression Models", *Statistica Sinica,* 13: 853–80.

Moscadelli, M. (2004) "The Modelling of Operational Risk: Experience with the Analysis of the Data Collected by the Basel Committee", Working Paper no. 517, Bank of Italy.

Picklands, J. (1975) "Statistical Inference Using Extreme Order Statistics", *The Annals of Statistics*, 3: 119–31.

Smirnov, N.V. (1939) "Sür Les Écarts de la Courbe de Distribution Empirique", *Matematiceskii Sbornik*, 6: 3–26.

Theil, H. (1971) *Applied Economic Forecasting* (Amsterdam: North-Holland).

CHAPTER 2

Incorporating Diversification into Risk Management

*Amiyatosh Purnanandam, Mitch Warachka, Yonggan Zhao and William T. Ziemba**

2.1 INTRODUCTION

Risk measurement is of fundamental importance to financial practice. Given the widespread usage of Value-at-Risk (VaR), firms actively manage their risk. Unfortunately, VaR is not derived from fundamental economic principles and may lead to sub-optimal decisions as shown by Shapiro and Basak (2001).

Substantial progress in the academic risk management literature began with Artzner, Delbaen, Eber and Heath (1999), abbreviated ADEH hereafter, who develop an axiomatic framework for risk measurement. Their axioms stem from intuitive economic principles that define a coherent risk measure. The intent of ADEH is to provide a regulator with a methodology for determining the riskfree capital requirements of a firm, conditional on their existing portfolio. Indeed, a coherent risk measure is defined as the minimum amount of riskfree capital a portfolio requires to become acceptable to the regulator.

* Mitch Warachka and Yonggan Zhao gratefully acknowledge financial support from the Wharton-SMU Research Center, Singapore Management University, and the Center for Research in Financial Services, Nanyang Technological University. William T. Ziemba acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada.

An entire literature on extensions of coherent risk measures followed ADEH. Rockafellar and Ziemba (2000) as well as Follmer and Schied (2002) introduce convex risk measures that account for market frictions by allowing risk to increase non-linearly with a portfolio's size. Jarrow (2002) enables a put option written on firm value (with zero strike price) to be coherent.

This chapter introduces a risk measure that is appropriate for the portfolio selection decisions of firms, while maintaining the ADEH concept of acceptable portfolios. To achieve this objective, we define risk on portfolio holdings, a domain conducive to having diversification reduce portfolio risk. We maintain an axiomatic structure and define the risk of a portfolio as its distance from the set of acceptable portfolios. More importantly, distance involves as many components as available assets, including but not limited to riskfree capital. As a consequence, derivative as well as insurance contracts become important tools for risk management. Thus, our approach conforms to market practice while its implementation involves quadratic programming, a technique with prior applications in finance originating from portfolio theory.

In contrast to coherent risk measures which focus on the regulator, this paper operates from the firm's perspective. In particular, we recognize that firms prefer to pursue investment opportunities that are capable of earning excess economic rents. This desire may stem from a perception of having superior information or investment ability. By implication, these ambitions result in portfolios that are not well-diversified. Intuitively, firms are unable to demonstrate investment skill by increasing their position in the riskfree asset. Thus, they are averse to adding riskfree capital to their portfolio for performance considerations, yet are constrained by an external regulator.

Balancing the demands of an external regulator and the performance objectives of firms is accomplished by introducing portfolio theory into the measurement of risk. Specifically, our proposed risk measure offers firms the ability to rebalance their portfolio. During this rebalancing, the addition of riskfree capital remains feasible, but is not the exclusive means by which a portfolio becomes acceptable. Since every asset portfolio weight may be altered, diversification is capable of reducing portfolio risk. Consequently, as discussed in Merton (1998), instruments with non-linear payoffs such as derivative and insurance contracts become important tools for risk management. In addition, market frictions may be incorporated into a firm's rebalancing decisions.

We also consider the pricing of portfolio insurance, a single contract whose addition to the existing portfolio is capable of ensuring its acceptability. This instrument provides more intuition for our risk measure, and converts the required portfolio rebalancing into a dollar-denominated quantity. The insurance contract does not reduce positive payoffs but insures against negative outcomes to avoid insolvency. Provided a firm is willing to rebalance their portfolio, only a fraction of this security is required.

The organization of this chapter is as follows. Section 2.2 details the properties of our proposed risk measure, while a simple example which illustrates our approach is given in section 2.3. Section 2.4 focuses on the implementation of our risk measure and demonstrates that coherent risk measures are contained in our framework. Section 2.5 considers the pricing of portfolio insurance while section 2.6 concludes.

2.2 RISK MEASURE WITH DIVERSIFICATION

Consider the time horizon $[0, T]$ and a finite number N of risky assets denoted x_i for $i = 0, 1, 2, \ldots, N$ with x_0 representing riskfree capital. Let P denote a $M \times (N + 1)$ payoff matrix with M rows indexed by $j = 1, 2, \ldots, M$ corresponding to the regulator's set of scenarios and $N + 1$ columns corresponding to the available assets. Elements of P are individual asset payoffs in a given scenario.

$$P = \begin{bmatrix} (1+r) & P_1(\omega_1) & \ldots & P_N(\omega_1) \\ (1+r) & P_1(\omega_2) & \ldots & P_N(\omega_2) \\ \vdots & \vdots & & \vdots \\ (1+r) & P_1(\omega_M) & \ldots & P_N(\omega_M) \end{bmatrix}$$

A vector of portfolio holdings $\eta = [\eta_0, \eta_1, \ldots, \eta_N]^\top$ represents the number of units, not dollar amounts or fractions of a portfolio, invested in the various assets. Portfolio values $P\eta$ in the M scenarios determine whether a portfolio complies with the demands of an external regulator.

Coherent risk measures evaluate a portfolio's risk according to its value in the worst possible scenario or under the probability measure that produces the largest negative outcome. Mathematically, these risk measures are defined in terms of terminal portfolio values, $X = P\eta$, as

$$\rho(X) = \max_j \frac{E^{P_j}[-X|P_j \in P]}{1+r} \tag{2.1}$$

with P representing a set of scenarios and r the riskfree rate of interest. In our framework, $E^{P_j}[-X]$ is replaced by $P\eta_j^-$, the jth row of $P\eta^- = -\min\{0, P\eta\}$ as each row of $P\eta$ corresponds to a regulator's scenario. Note that each scenario represents a probability measure. However, the expected value of the portfolio across multiple scenarios is not computed by the regulator. Instead, the worst outcome across the scenarios defines the risk of a coherent risk measure.

It is important to emphasize that coherent risk measures do not account for diversification. Although ADEH have a subadditivity axiom that parallels a property implied by our risk measure, their definition of risk applies to terminal portfolio values. Thus, the portfolio weights of the underlying assets cannot be altered to exploit the benefits of diversification. Instead, according to the stricter version of ADEH's acceptance set which invokes Axiom 2.2′, ADEH focus solely on the amount of riskfree capital required to ensure the portfolio has non-negative terminal values in the scenarios considered relevant by the regulator. This exclusive focus on riskfree capital is overcome by our methodology which operates on a different domain. Specifically, define $M \subset R^{N+1}$ as the space of portfolio holdings with the subset of acceptable portfolios denoted $A_\eta \subset M$.

As in ADEH, the definition of acceptable portfolios ensures the firm cannot become insolvent in any of the regulator's M scenarios. However, firms may supplement this set of scenarios to obtain additional protection against insolvency.

Definition 2.2.1 The set of acceptable portfolio holdings $A_\eta \subset M$ contains all portfolios that have non-negative outcomes, $P\eta \geq 0$, in all M scenarios evaluated by the regulator.

Clearly, the acceptance set A_η depends on the payoff matrix P with the regulator controlling the number of scenarios (rows). Moreover, the regulator focuses on preventing insolvency but does not consider *variability* in a firm's portfolio value when measuring risk. With respect to risk factors such as market, interest rate or foreign exchange movements, these variables may define the M scenarios. For example, the Standard Portfolio Analysis of Risk (abbreviated SPAN) risk management system employed by most international exchanges such as the CBOT, CME, NYBOT, NYMEX and LIFFE, investigates 16 scenarios defined by changes in price and volatility. In each of these scenarios, the firm's portfolio is required to be non-negative.

Proposition 2.2.1 The acceptance set A_η has the following two properties:

1. Closed under multiplication by $\gamma \geq 0$.
2. Convexity.

Proof: First, it must be shown that if $\eta \in A_\eta$, then $\gamma\eta \in A_\eta$ for $\gamma \geq 0$. This property follows from $\eta \in A_\eta$ being equivalent to $P\eta \geq 0$ and the property $P[\gamma\eta] = \gamma P\eta$ which is non-negative since both γ and $P\eta$ are non-negative.

Second, if $\eta_1, \eta_2 \in A_\eta$, implying $P\eta_1 \geq 0$ and $P\eta_2 \geq 0$, then $\gamma\eta_1 + (1-\gamma)\eta_2 \in A_\eta$ for $0 \leq \gamma \leq 1$ since $P[\gamma\eta_1 + (1-\gamma)\eta_2] = \gamma P\eta_1 + (1-\gamma)P\eta_2 \geq 0$.

Therefore, as in the ADEH framework, unless each element of $P\eta$ is non-negative, the portfolio η is unacceptable. In this instance, an optimal acceptable η^* is found based on its *proximity* to η as we assume firms prefer to engage in as little portfolio rebalancing as possible given their initial preference for η. Quadratic programming solves for the portfolio η^* in section 2.4.

Define a trivial acceptable portfolio η_c consisting of \$1 invested only in riskfree capital, in other words, a $N+1$ vector with one as the first element and zero in the remaining elements. This portfolio has the property $P\eta_c = (1+r)1 > 0$ where 1 is a $N+1$ vector of ones. Given the acceptance set A_η in Definition 2.2.1, portfolio risk is defined in terms of the l_2 norm, $||x-y||_2$ equals $\sqrt{\sum_{i=0}^{N}(x_i-y_i)^2}$, on M. Our risk function $\rho(\eta)$ maps from the domain of portfolio holdings, M, into the non-negative real line, $\rho(\eta)$: $M \to R^1_+$.

Definition 2.2.2 Given A_η defined by the payoff matrix P, the risk of a portfolio η equals

$$\rho(\eta) = \inf\{||\eta - \eta'||_2 : \eta' \in A_\eta\}$$

Observe the fundamental difference between our approach and that of ADEH, instead of defining risk on terminal portfolio values, risk is defined on portfolio holdings. Thus, although both measures of risk are defined by a *distance* from an acceptance set, our concept of distance has $N+1$ variables (one for each asset) instead of only one (riskfree capital). Different objective functions may be utilized with the important property in Definition 2.2.2 being the measurement of risk in terms of distance. The l_2 norm is chosen for tractability and because of its prior applications in portfolio theory.

At this stage, we state three important clarifications regarding our risk measure. First, although $\rho(\eta)$ is Euclidean distance, it has an immediate dollar-denominated interpretation given prices for each of the assets as discussed in section 2.5. Second, firm preferences are easily incorporated into our risk measure. Third, the dollar-denominated value of the original portfolio differs from its acceptable counterpart denoted η' in Definition 2.2.2.

As elaborated in section 2.4, our methodology recognizes that the firm is not necessarily less willing to rebalance assets with higher prices. Instead, deviations in the portfolio weights of the original portfolio are minimized since relatively inexpensive assets such as out-of-the-money options or futures contracts (with zero value after being market-to-market) may be crucial to both the desirability and riskiness of a firm's investment strategy. Indeed, forward and swap contracts have zero initial value but potentially large positive or negative payoffs. In contrast, the firm may be willing to alter their holdings of expensive instruments such as Treasury bonds. Thus, our primary objective is minimizing perturbations to the firm's original portfolio η, which is assumed to be its preferred allocation. Instead, the firm is able to specify the cost of rebalancing each individual asset from their perspective when finding the acceptable portfolio's solution. As seen in section 2.4, one possibility simply has the rebalancing cost for each asset being equal to its price. This special case minimizes the dollar-denominated amount of rebalancing. However, such a formulation is not necessarily compatible with our objective of including derivatives or other (leveraged) assets capable of offering negative correlation. Irrespective of this complication, the dollar-denominated risk measure implied by Definition 2.2.2 is presented in subsection 2.5.3.

In section 2.4, an enhanced methodology which accounts for firm preferences also becomes explicit, and addresses the second issue. This extension is identical to the incorporation of market frictions such as transaction costs and illiquidity into portfolio rebalancing decisions. The third observation also applies to the ADEH risk measure as the addition of riskfree capital increases the dollar-denominated value of the original portfolio. Indeed, no existing risk management system reduces portfolio risk while preserving its original value.

If η already comprises an acceptable portfolio, then its associated risk equals zero. For example, the portfolio η_c has zero risk, $\rho(\eta_c) = 0$. Otherwise, portfolio risk is determined by the amount of rebalancing a portfolio requires to become acceptable. This illustrates a major advantage of our risk measure. A firm may rebalance their portfolio by purchasing derivative instruments, insurance contracts, or simply reducing their exposure to certain risky assets. In summary, portfolio rebalancing may include, but is not limited to, increasing the amount of riskfree capital.

As a final observation, generalized scenarios considered in ADEH are also addressed in our methodology. A generalized scenario is defined as a combination of multiple scenarios aggregated by a specified probability measure. For example, one element of the SPAN procedure considers a 30 percent chance of an extreme scenario in conjunction with a 70 percent chance of another base scenario. In our context, the {0.3, 0.7} probability forms the generalized scenario:

$$\begin{array}{l} \quad 0.3 \times \text{Payoff vector in extreme scenario} \\ \underline{+\,0.7 \times \text{Payoff vector in the base scenario}} \\ = \text{Payoff vector of generalized scenario} \end{array}$$

which is no different than any other row of the P matrix. However, whether or not the scenarios underlying the generalized scenario are themselves included as distinct rows of P is immaterial to our analysis. For example, the extreme scenario in SPAN is not evaluated as an individual scenario.

2.2.1 Properties of risk measure

The next proposition summarizes the properties of our risk measure. Interestingly, all but one of ADEH's coherence axioms are preserved. However, removal of the translation invariance axiom results in an important generalization by eliminating the strict dependence on riskfree capital to reduce risk. To clarify, the operations $\eta_1 \pm \eta_2$ are applied componentwise to signify operations on two vectors representing portfolio holdings.

Proposition 2.2.2 The proposed risk measure with diversification has the following properties:

1 Subadditivity $\rho(\eta_1 + \eta_2) \leq \rho(\eta_1) + \rho(\eta_2)$

2 Monotonicity $\rho(\eta_1) \leq \rho(\eta_2)$ if $P\eta_1 \geq P\eta_2$

3 Positive homogeneity $\rho(\gamma\eta) = \gamma\rho(\eta)$ for $\gamma \geq 0$

4 Riskfree capital monotonicity $\rho(\eta + \gamma\eta_c) \leq \rho(\eta)$ for $\gamma \geq 0$

5 Relevance $\rho(\eta) > 0$ if $\eta \notin A_\eta$

6 Shortest path For every $\eta \notin A_\eta$ and for $0 \leq \gamma \leq ||\eta - \eta^*||_2$:

$$\rho(\eta + \gamma \cdot \tilde{u}) = \rho(\eta) - \gamma$$

where $\tilde{u}$ is the unit vector in the direction $\eta^* - \eta$ defined as $\eta^* - \eta/||\eta^* - \eta||_2$ given a portfolio η^* that lies on the boundary of A_η and minimizes the distance $||\eta - \eta^*||_2$.

The proof is contained in Appendix A. The shortest path property imposes cardinality on the risk measure with $\tilde{u}$ representing a unit of rebalancing. Observe that *riskier* portfolios are farther from the acceptance set with larger associated risk measures $\rho(\eta)$. Versions of the subadditivity, monotonicity, and positive homogeneity properties found in the original ADEH paper remain with subadditivity responsible for incorporating diversification into our framework. The second and third properties, monotonicity and positive homogeneity, are discussed in ADEH. Monotonicity guarantees that a portfolio whose terminal payoffs are larger than another portfolio in every scenario has lower risk than its counterpart. Positive homogeneity allows a firm to scale an acceptable portfolio up or down with the resulting portfolio remaining acceptable. To account for market frictions, Follmer and Schied (2002) replace positive homogeneity and subadditivity with a convexity axiom. In our framework, market frictions influence the solution for η^* as demonstrated in section 2.4.

The key distinction arises from ADEH's translation invariance axiom. Our risk measure with diversification employs a weaker concept manifested in the riskfree capital monotonicity and shortest path properties. The relevance property ensures the risk function is positive if there exists a scenario, considered relevant by the regulator, where the terminal value of the portfolio is negative. Consequently, the relevance property ensures unacceptable portfolios have positive risk.

When $\rho(\eta) = 0$, an amount γ^* of riskfree capital may be removed from the portfolio according to $\sup_{\gamma^*} \rho(\eta - \gamma^*\eta_c) = 0$, which is unique by the monotonicity of riskfree capital property. Since A_η is closed, there exists a boundary point which minimizes the required amount of riskfree capital. Although quadratic programming is capable of solving for γ^*, this issue is not elaborated on further as our focus concerns unacceptable η portfolios.

2.2.2 Economic motivation

An unacceptable portfolio may initially be chosen by a firm which believes it has superior information or investment skill. Moreover, additional riskfree capital does not permit a firm to exhibit investment ability or skill. Provided firms pursue excess economic rents and fail to maintain well diversified portfolios, a coherent risk measure may overestimate portfolio risk. Fundamentally, a tradeoff exists between preventing insolvency and maximizing the expected value of the portfolio. This tension motivates our risk management framework since the firm is able to maintain an acceptable portfolio as close as possible to their original positions while complying with the external regulator.

To enhance the motivation behind our risk measure, we introduce a non-negative function $R(\eta) \geq 0$ to determine the aggregate desirability of a portfolio. This function is only intended as an example to illustrate the desirability of rebalancing a portfolio in comparison to the addition of riskfree capital. Therefore, neither the functional form nor the exact specification of $R(\eta)$ are necessary for our analysis. Since the selection criteria and perceived desirability of individual assets are highly variable across firms, very little structure is imposed on $R(\eta)$. For illustration, we assume:

$$R(\eta) = \frac{\sum_{i=0}^{N} \eta_i \cdot c_i}{\sum_{i=0}^{N} \eta_i} \tag{2.2}$$

where c_i implicitly denotes a ranking of the assets. For example, c_i may represent numerical weightings associated with *strong outperform*, *weak outperform*, or *hold* among other possibilities. Equation (2.2) allows several variables, including expected returns and variances, to influence a portfolio's desirability. However, covariances are not considered in equation (2.2) as diversification is reserved for our subsequent discussion of the proposed risk measure.

Regardless of the exact functional form for $R(\eta)$, the c_i elements may be derived from an infinite number of scenarios, not only those evaluated by the regulator. Indeed, the regulator is primarily concerned with a small subset of *extreme* scenarios. In contrast, the firm's investment criteria is comprised of more frequently occurring scenarios. This disparity reflects the diverging interests of the regulator and firm which our proposed risk measure attempts to bridge.

In the absence of portfolio rebalancing, define the amount of additional riskfree asset required to ensure the portfolio η becomes acceptable as $\alpha \geq 0$. This quantity equals

$$\begin{aligned} \alpha &= \inf\{\gamma : \eta + \gamma \eta_c \in A_\eta\} \\ &= -\min\{0, P\eta\} \end{aligned} \tag{2.3}$$

and depends on η but is written as α rather than $\alpha(\eta)$ for notational simplicity.

Overall, for $\eta \notin A_\eta$, diversification is beneficial from the firm's perspective whenever there exists an $\eta' \in M$ (not necessarily acceptable) such that:

$$\text{Condition 1: } \eta + \eta' \in A_\eta \tag{2.4}$$

$$\text{Condition 2: } R(\eta + \eta') \geq R(\eta + \alpha\eta_c) \tag{2.5}$$

The first condition ensures that η', when added to η, is capable of constituting an acceptable portfolio. A solution for η' that satisfies the first condition is provided in section 4. The second condition states that portfolio rebalancing is preferred to the addition of riskfree capital when complying with the regulator. Indeed, setting $\eta' = \alpha\eta_c$ results in equality for the second condition. The existence of a portfolio η' is motivated by the inability of η_c to generate excess economic rents. Other functions besides equation (2.2) are possible, with the property $c_i \geq c_0$ for $i \geq 1$ ensuring the second condition is satisfied.

In practice, the regulator may impose a fine denoted f on firms that continue to hold unacceptable portfolios. Thus, the second condition expressed in equation (2.5) may be extended to

$$R(\eta + \eta') \geq \max\left\{R(\eta + \alpha\eta_c), R(\eta) - f1_{\{\eta \notin A_\eta\}}\right\} \tag{2.6}$$

Assuming the fine is large enough to satisfy both:

1 $f \geq R(\eta) - R(\eta + \alpha\eta_c)$

2 $f \geq R(\eta) - R(\eta + \eta')$

firms strive to be in compliance with the regulator. Indeed, the firm is better-off rebalancing the portfolio than adding riskfree capital or paying the fine and maintaining their original portfolio. Since $R(\eta + \eta') \geq R(\eta + \alpha\eta_c)$, the two requirements above reduce to the first statement:

$$f \geq R(\eta) - R(\eta + \alpha\eta_c).$$

Hence, the required fine is a function of both η and the firm's aversion to adding riskfree capital expressed via $R(\eta)$. Intuitively, firms which are less averse to holding riskfree capital require smaller fines to induce compliance.

Observe that the addition of riskfree capital increases a portfolio's payoffs in all scenarios, even those for which the original portfolio already has non-negative values. Indeed, the portfolio payoff increases in scenarios that are not even considered by the regulator. Therefore, the addition of riskfree capital is a very conservative approach to risk management, one suitable from the perspective of a regulator but not firms. Section 2.5 investigates the pricing of portfolio insurance, a security which only increases payoffs in scenarios that prevent the portfolio from being acceptable. Furthermore, firms are able to evaluate scenarios beyond those considered by the regulator if their internal risk management procedures are designed to be more stringent.

The next section considers a simple example to differentiate our risk measure from coherent risk measures.

2.3 NUMERICAL EXAMPLE

Consider an economy with two risky assets and riskfree capital. Uncertainty in the economy is captured by a coin toss. For the first risky asset, the payoff is \$4 if heads and −\$2 if tails, while their counterparts are \$0 and \$2 respectively for the second risky asset. The rate of interest is assumed to be zero ($r = 0$) implying risk-free capital is worth \$1 at time T.

The two risky assets are negatively correlated. For emphasis, no probability measure is required for the occurrence of the two states as the regulator is not concerned with their likelihood. Instead, preventing insolvency in each scenario is the regulator's task, which is entirely independent of the portfolio's expected value across the scenarios. Indeed, the second risky asset resembles a put option on the first security. Furthermore, the market is complete since a portfolio weight of $-\frac{1}{3}$ in the first asset combined with $\frac{4}{3}$ of risk-free capital replicates the put option. Intuitively, the put option provides negative correlation, facilitating greater diversification, by shorting the first asset while indirectly providing additional riskfree capital. As illustrated in the remainder of this example, beyond serving as an effective means to hedge risk and ensure portfolio acceptability, the derivative reduces the amount of riskfree capital required to be held by the firm.

The space of acceptable portfolio holdings whose terminal values are non-negative in both scenarios is characterized by:

$$1\eta_0 + 4\eta_1 + 0\eta_2 \geq 0 \qquad \text{Heads} \tag{2.7}$$

$$1\eta_0 - 2\eta_1 + 2\eta_2 \geq 0 \qquad \text{Tails} \tag{2.8}$$

Consider the portfolio $\eta = [1, 1, 0]^\top$ consisting of one unit of riskfree capital, one unit of the first risky asset and none of the second. The portfolio η is not acceptable since the payoff is negative if the coin toss results in tails.

In the coherent risk measure framework, η requires an additional unit of riskfree capital resulting in $\eta^*_{ADEH} = [2, 1, 0]^\top$.

Solving for our optimal portfolio η^* involves minimizing the distance between $\eta = [1, 1, 0]^\top$ and $\eta^* \in A_\eta$ under the l_2 norm using quadratic programming (QP). The portfolio η^* equals $[1.11,\ 0.78,\ 0.22]^\top$ with details pertaining to its solution found in the next section. MATLAB code which solves for η^* is available from the authors.

As demonstrated above, a coherent risk measure evaluates the risk of η as 1 due to the negative payoff when the coin toss is tails. However, the portfolio $[1.11,\ 0.78,\ 0.22]^\top \in A_\eta$ implies the portfolio's risk in our framework is $||\eta^* - \eta||_2 = \sqrt{(1.11-1)^2 + (0.78-1)^2 + (0.22-0)^2} = 0.33$. Thus, our proposed risk measure evaluates the risk of η at one third that of a coherent risk measure. However, the rebalanced portfolio has non-negative payoffs in both scenarios and therefore satisfies the regulator.

Table 2.1 Asset payoffs at time T in both scenarios

Asset	Heads	Tails
Riskfree capital	1	1
Risky asset #1	4	−2
Risky asset #2	0	2

Table 2.2 Payoffs at time T in both scenarios for the unacceptable portfolio η

Portfolio	Heads	Tails
$\eta = [1,1,0]^\top$	5	−1

Table 2.3 Payoffs at time T in both scenarios for the η^*_{ADEH} portfolio

Acceptable portfolio – ADEH	Heads	Tails
$\eta^*_{ADEH} = [2,1,0]^\top$	6	0

Table 2.4 Payoffs at time T in both scenarios for the η^* portfolio

Optimal Portfolio – QP	Heads	Tails
$\eta^* = [1.11,0.78,0.22]^\top$	4.22	0

The distinction between the riskfree asset and its risky counterparts is not central to our proposed risk management framework. In contrast to traditional portfolio allocation decisions, the important issue in our context is the tradeoff between additional risk-free capital versus rebalancing the risky assets. A portfolio's rebalancing may include the purchase of options, futures contracts, insurance or reducing the portfolio's exposure to certain troublesome assets.

For example, to comply with the regulator, suppose the firm is faced with two choices; adding \$1,000,000 in riskfree capital (lowering their expected return) versus acquiring positions in forward/futures/swap contracts (which are costless at initiation) or inexpensive out-of-the-money options. Naturally, deviating from the existing asset allocation between the risky assets is essential since maintaining its exact composition would be clearly inefficient. Subsection 2.2.2 contains additional economic motivation in favor of rebalancing.

To clarify, portfolio theory selects portfolio weights to exploit diversification *before* choosing the desired amount of riskfree capital. These decisions are independent and sequential since the risky portfolio is assumed to already be fully diversified. However, this assumption is not present in our methodology. As demonstrated in the next section, we recognize that a fully diversified portfolio results in our risk measure being identical to that of ADEH. Furthermore, to minimize the assumptions and structure imposed on firm preferences, our risk measure's objective is to perturb the firm's original portfolio by the least amount possible while complying with the regulator.

Section 2.4 also reveals that the firm may ignore all the flexibility offered by our approach and preserve their original risky asset allocation. As alluded to earlier, we provide the firm with $N+1$ degrees of freedom to satisfy the regulator in contrast to ADEH who only allow the portfolio weight of riskfree capital to be manipulated. Thus, fixing the original positions in the risky assets or focusing one's attention on the division between the risky portfolio and riskfree capital eliminates any possibility of diversification since portfolio theory requires an optimal solution to be expressed in terms of portfolio weights.

Finding more general solutions for η^* that incorporate market frictions into the rebalancing decision is addressed in section 2.4. In our previous example, the positive portfolio payoff in the heads scenario was reduced. Section 2.5 computes the value of a portfolio insurance contract which eliminates negative terminal values without reducing their positive counterparts.

To summarize, this section offers an illustration of how firms may comply with the demands of a regulator while holding less riskfree capital. Indeed, regulators may adopt our risk measure without compromising their original role of preventing insolvency in each scenario.

2.4 IMPLEMENTATION

If $P\eta$ has any negative elements, then the regulator deems the portfolio to be unacceptable. This section is concerned with implementing our risk measure by solving for the portfolio $\eta^* \in A_\eta$ such that $P\eta^* \geq 0$ and η^* is "as close as possible" to the firm's original portfolio η.

Definition 2.4.1 Allowing g to represent the l_2 norm, the portfolio $\eta^* \in A_\eta$ is the solution to the optimization problem:

$$\begin{aligned} &min_{\eta^* \in R^{N+1}} \quad g(\eta^* - \eta) \\ &\text{subject to} \quad P\eta^* \geq 0 \end{aligned} \tag{2.9}$$

Equation (2.9) solves for the minimum amount of portfolio rebalancing, which is not a dollar-denominated quantity. Indeed, the properties of our risk measure described in Proposition 2.2.2 apply to portfolio weights.

Altering equation (2.9) to minimize a function of $q^T(\eta - \eta^*)$, where q denotes a vector of asset prices, contradicts the initial portfolio η being preferred by the firm. Specifically, there is no reason why the firm would be less willing to rebalance assets with higher initial prices. At inception, forward, swap and futures contracts have zero value but potentially large positive or negative payoffs, while inexpensive out-of-the-money options have a similar property. Consequently, deviations from the original portfolio are minimized since relatively inexpensive assets may be crucial to a firm's investment strategy or have higher market frictions as discussed in the next subsection. The dollar-denominated price of portfolio insurance is addressed in section 2.5.

In a financial context, quadratic programming, implied by the l_2 norm, is equivalent to the mean-variance analysis underlying much of portfolio theory. Since the objective function g is twice differentiable and strictly convex and the feasible region is also convex, the Kuhn Tucker conditions imply a unique solution. Although this problem cannot be solved analytically, very efficient numerical solutions are available. In particular, the problem is well suited for a pivoting scheme described in Luenberger (1990).

Proposition 2.4.1 Let $y = P\eta^*$ and $g(\eta^*) = \frac{1}{2}(\eta^* - \eta)^\top(n^* - \eta)$. The optimal solution to equation (2.9) is given by $\eta^* = \eta + P^\top\lambda$ where λ solves the linear complementarity conditions:

$$\begin{cases} y - PP^\top\lambda = P\eta \\ y \geq 0, \quad \lambda \geq 0, \quad \lambda^\top y = 0 \end{cases} \tag{2.10}$$

Proof: The Kuhn Tucker conditions are:

$$\begin{cases} \eta^* - P^\top\lambda = \eta \\ P\eta^* \geq 0, \quad \lambda \geq 0, \quad \lambda^\top P\eta^* = 0 \end{cases}$$

since the gradient of the objective function, $g(\eta^*) - (P\eta^*)^\top\lambda$, equals $\eta^* - \eta - P^\top\lambda$. Hence, with $y = P\eta^*$, the above conditions become:

$$\begin{cases} y - PP^\top\lambda = P\eta \\ y \geq 0, \quad \lambda \geq 0, \quad \lambda^\top y = 0 \end{cases}$$

which completes the proof.

Hence, the optimization problem in equation (2.9) is reduced to solving the linear complementary conditions in (2.10). Furthermore, the optimal portfolio η^* is a linear function of the vector λ which satisfies these linear complementary conditions. However, there may exist multiple solutions to (2.10), raising the question whether all possible solutions yield the same optimal portfolio η^* in Definition 2.4.1. This issue is addressed in the following proposition whose proof is found in Appendix B.

Proposition 2.4.2 All solutions to the linear complementary conditions in (2.10) yield the same optimal portfolio η^* in Definition 2.4.1.

The λ parameters have interesting interpretations as each element corresponds to a specific regulator scenario. If the constraint $P\eta \geq 0$ is not binding in scenario i with $(P\eta)_i \geq 0$, then the corresponding λ_i equals 0. Otherwise, the optimal λ_i is a positive number representing the *cost* of preventing insolvency.

If $P\eta \geq 0$, then (2.10) has an obvious solution; $\lambda = 0$ and $y = P\eta$, implying η is optimal. Otherwise, the general pivoting approach transforms (2.10) to optimality. After finitely many pivots, bounded above by the number of rows (scenarios), the vector $P\eta^*$ is non-negative. In terms of computational complexity, a total of M linear equations are solved for each pivot operation. The algorithm stops when $P\eta^* \geq 0$, providing the optimal solution to (2.10).

2.4.1 Incorporating market frictions and firm preferences

In general, the objective function g may be defined with respect to a positive definite matrix A as in $(\eta^* - \eta)^\top A(\eta^* - \eta)$. Consider a diagonal matrix of positive elements a_i:

$$A = \begin{bmatrix} a_0 & & & \\ & a1 & & \\ & & \ddots & \\ & & & a_N \end{bmatrix}$$

representing the associated market friction (illiquidity and transaction costs) of the *ith* asset as well as the firm's unwillingness to alter their position in this asset. Larger a_i values correspond to larger *penalties* for altering that element of the portfolio. Even if riskfree capital has the smallest corresponding penalty, the addition of riskfree capital may still be sub-optimal. Indeed, a portfolio may require a large amount of additional riskfree capital to become acceptable, but only minor modifications to positions with larger a_i penalties. This issue is re-examined in the next section when pricing portfolio insurance.

Alternatively, the price of each asset could define the a_i elements. In this circumstance, the dollar-denominated amount of rebalancing is minimized as assets with higher prices are more expensive to rebalance. However, as alluded to earlier, inexpensive futures contracts or out-of-the-money options often provide large future payoffs and are crucial to a firm's investment strategy, while expensive instruments such as high-coupon bonds are not. Thus, the price of an individual security is not necessarily representative of firm preferences towards rebalancing. Nonetheless, the potential to incorporate prices into the solution of η^* is apparent.

Observe that the c_i elements of the $R(\eta)$ function in equation (2.2) are not incorporated into A. Indeed, solving for the optimal acceptable portfolio that maximizes $R(\eta)$ is well-beyond the scope of this paper and would require far greater structure on firm preferences, information and beliefs. Since η represents the firm's optimal portfolio in the absence of the regulator, we merely assume any deviation from η is disliked by the firm.

Proposition 2.4.1 has an immediate corollary when the positive definite matrix A is inserted into the objective function which alters the values of both η^* and λ.

Corollary 2.4.1 Let $y = P\eta^*$ and $g(\eta^*) = \frac{1}{2}(\eta^* - \eta)^\top A(\eta^* - \eta)$ where A is a positive definite matrix. The optimal portfolio η^* equals $\eta + A^{-1}P^\top\lambda$, where λ satisfies the modified linear complementarity conditions:

$$\begin{cases} y - PA^{-1}P^\top\lambda = P\eta \\ y \geq 0, \quad \lambda \geq 0, \quad \lambda^\top y = 0 \end{cases} \tag{2.11}$$

Given Corollary 2.4.1 above, we now reconsider the example in section 2.3 for different A matrices and their corresponding optimal acceptable portfolios.

2.4.2 Continuation of example

Once again, the original unacceptable portfolio $\eta = [1, 1, 0]^\top$ is considered. Suppose a firm is extremely adverse to adding riskfree capital to their portfolio. This preference is expressed through the matrix:

$$A_1 = \begin{bmatrix} \infty & & \\ & 1 & \\ & & 1 \end{bmatrix}$$

which implies η_1^* equals $[1, 0.75, 0.25]^\top$. When implementing the numerical examples presented in this paper, ∞ is replaced with 1000. The portfolio η_1^* is acceptable with $P\eta_1^*$ being non-negative in both scenarios. Therefore, our proposed risk measure generates an acceptable portfolio without any additional riskfree capital by reducing the firm's exposure to the first risky asset and purchasing a portion of the second risky asset as a hedge.

Interestingly, one may begin with the portfolio $\overline{\eta} = \eta - \eta_c = [0, 1, 0]^\top$ and find $\overline{\eta}_1^*$, with the prevailing A_1 matrix, without utilizing any additional riskfree capital. Indeed, $[0,\ 1,\ 1]^\top$ consists entirely of risky assets and is acceptable.

Furthermore, suppose the firm also has a strong desire to maintain their position in the first risky asset. Returning to the original η portfolio, the A matrix:

$$A_2 = \begin{bmatrix} \infty & & \\ & \infty & \\ & & 1 \end{bmatrix}$$

generates an optimal portfolio $\eta_2^* = [1, 1, 0.50]^T$. As expected, only the position in the second risky asset is modified.

Finally, we examine an A matrix capable of replicating the optimal ADEH portfolio:

$$A_{ADEH} = \begin{bmatrix} 1 & & \\ & \infty & \\ & & \infty \end{bmatrix}$$

which implies $\eta_{ADEH}^* = [2, 1, 0]^\top$. In this situation, only additional riskfree capital is chosen. Overall, by eliminating the possibility of rebalancing the risky assets, the ADEH risk measure implicitly has $a_{i \neq 0} = \infty$.

The above examples illustrate the ability of our methodology to find optimal acceptable portfolios that reflect market frictions, as well as an aversion to additional riskfree capital or altering positions in specific risky assets. In summary, implementing our framework reduces to solving a quadratic programming problem, a situation encountered in many financial applications involving portfolio theory.

2.5 PRICING PORTFOLIO INSURANCE

This section determines the price of portfolio insurance, a single contract whose combination with the original portfolio satisfies the regulator. Consistent with the goal of incorporating derivative contracts into our risk management framework, we assume the economy admits no arbitrage opportunities. For notational simplicity, we assume A in the previous section is an $N+1$ identity matrix although incorporating this extension into our analysis is immediate.

Conceptually, the insurance contract summarizes the amount of rebalancing required to satisfy the regulator and provides a dollar-denominated measure of risk. In particular, the insurance contract itself represents a portfolio whose combination with the original portfolio may be interpreted as rebalancing the latter. Furthermore, as expected, the negative correlation introduced by the insurance contract in relation to the original portfolio ensures that exploiting the benefits of diversification is feasible.

Let IC denote the non-negative price of the contract in circumstances where $P\eta$ contains at least one negative value. Denote $X^+ = \max\{0, X\}$ and $X^- = -\min\{0, X\}$. To become acceptable, the firm requires a contract with a payoff profile equal to $(P\eta)^-$. In addition, we ensure the portfolio, when combined with the insurance contract, continues to provide $(P\eta)^+$ in scenarios with positive values. Thus, the insurance contract does not reduce positive terminal values, it only increases negative terminal values to zero. Hence, in contrast to riskfree capital, portfolio insurance only provides a positive payoff in scenarios where it is necessary.

We endogenously determine the value of portfolio insurance by equating the dollar value of the optimal portfolios at time zero with and without this contract. This indifference stems from portfolio insurance being redundant since an acceptable portfolio may be obtained via rebalancing. Indeed, portfolio insurance provides an economically intuitive *short-cut* to acceptability by serving as a customized put option on the portfolio's terminal value.

2.5.1 Insurance without rebalancing

Let q denote the price vector of the $N+1$ assets at time zero which is assumed to be free of arbitrage. The proposition below solves for the price of portfolio insurance under the assumption that no additional rebalancing is conducted after its introduction.

Proposition 2.5.1 The price of the portfolio insurance, without additional portfolio rebalancing, equals

$$IC_{wo} = q^\top P^\top \lambda_{wo}$$

where λ_{wo} is determined by the resulting linear complementary conditions.

Proof: Consider the alternative to purchasing an insurance contract. The firm must rebalance their portfolio to obtain η^* which satisfies $P\eta^* \geq (P\eta)^+$. The optimization problem which solves for η^* is:

$$\begin{aligned} \min_{\eta^*} \quad & g(\eta^* - \eta) \\ \text{subject to} \quad & P\eta^* \geq (P\eta)^+ \end{aligned} \tag{2.12}$$

The Kuhn-Tucker conditions imply that the optimal solution is given by the solution to the following linear complementarity conditions:

$$\begin{cases} y - PP^\top \lambda = -(P\eta)^- \\ y \geq 0, \quad \lambda \geq 0, \quad \lambda^\top y = 0 \end{cases} \tag{2.13}$$

where $y = P\eta^* - (P\eta)^+$. The property $P\eta = -(P\eta)^- + (P\eta)^+$ implies $y - PP^\top \lambda = -(P\eta)^-$ in (2.13) is equivalent to $P\eta^* - P\eta - PP^\top \lambda = 0$ in Proposition 2.4.1.

Denote the solution to (2.13) by $(\eta_{wo}, \lambda_{wo})$ where η_{wo} represents the optimal portfolio without the insurance contract. The following linear relationship between η_{wo} and the original portfolio η holds:

$$\eta_{wo} = \eta + P^\top \lambda_{wo} \tag{2.14}$$

With firms indifferent between buying the contract or rebalancing their portfolio, the dollar values of the two acceptable portfolios at time zero are equated. Thus, the price of the insurance contract equals $IC_{wo} + q^\top \eta = q^\top \eta_{wo}$, implying

$$IC_{wo} = q^\top(\eta_{wo} - \eta) = q^\top P^\top \lambda_{wo} \tag{2.15}$$

which completes the proof.

The value of IC_{wo} is positive since the payoff $(P\eta)^-$ is non-negative in each scenario and strictly positive in at least one scenario. Specifically, the property $P\eta \geq 0$ with strict inequality in at least one scenario implies the initial cost of the portfolio $q^\top \eta$ is positive. The condition $y \geq 0$ in (2.13) yields $P\eta_{wo} - (P\eta)^+ \geq 0$ which implies that $P(\eta_{wo} - \eta) \geq 0$ with strict inequality in at least one scenario provided $(P\eta)^- \neq 0$. Therefore, no arbitrage implies $IC_{wo} = q^\top(\eta_{wo} - \eta) > 0$.

2.5.2 Insurance with rebalancing

The following analysis has firms willing to engage in additional rebalancing to exploit the diversification benefit offered by the availability of portfolio insurance. Let the insurance contract be the $N + 2nd$ security resulting in an additional column being appended to P to form $Q = [P(P\eta)^-]$. This column increases negative terminal values in scenarios that previously implied insolvency. In addition, enhanced portfolios with and without portfolio insurance are defined as:

$$\delta_1 = \begin{bmatrix} \eta \\ 1 \end{bmatrix} \quad \text{and} \quad \delta_0 = \begin{bmatrix} \eta \\ 0 \end{bmatrix}$$

While δ_0 is not acceptable, δ_1 is acceptable since $Q\delta_1 = P\eta + (P\eta)^- = (P\eta)^+ \geq 0$. However, we later prove that δ_1 is not optimal when there are fewer scenarios than available assets.

Proposition 2.5.2 The price of portfolio insurance, with additional portfolio rebalancing, equals

$$IC_w = \frac{q^\top P^\top(\lambda_{wo} - \lambda_w)}{((P\eta)^-)^\top \lambda_w}$$

with λ_{wo} previously determined in Proposition 2.5.1 and λ_w by the resulting linear complementary conditions.

Proof: Denote $\delta^* = \begin{bmatrix} \eta_w \\ x_w \end{bmatrix}$. The optimal solution defined over the $N + 2$ assets is

$$\begin{aligned} \min_{\delta^* \in R^{N+2}} \quad & g(\delta^* - \delta_0) \\ \text{subject to} \quad & Q\delta^* \geq (P\eta)^+ \end{aligned} \tag{2.16}$$

with linear complementarity conditions:

$$\begin{cases} y - QQ^\top \lambda = -(P\eta)^- \\ y \geq 0, \quad \lambda \geq 0, \quad \lambda^\top y = 0 \end{cases} \tag{2.17}$$

for $y = Q\delta^* - (P\eta)^+$. Denote the optimal solution to (2.17) by (η_w, x_w, λ_w) which yields:

$$\begin{bmatrix} \eta_w \\ x_w \end{bmatrix} = \begin{bmatrix} \eta \\ 0 \end{bmatrix} + \begin{bmatrix} P^\top \\ ((P\eta)^-)^\top \end{bmatrix} \lambda_w \tag{2.18}$$

Therefore, the second equation of (2.18) implies the optimal amount of insurance to purchase equals:

$$x_w = ((P\eta)^-)^\top \lambda_w \geq 0 \tag{2.19}$$

Hence, conditional on additional rebalancing from η to η_w, the price of the insurance contract is $IC_w \cdot x_w + q^\top \eta_w = q^\top \eta_{wo}$ which is equivalent to

$$IC_w = \frac{q^\top P^\top (\lambda_{wo} - \lambda_w)}{((P\eta)^-)^\top \lambda_w} \tag{2.20}$$

by equation (2.19) and the relationship $\eta_{wo} - \eta_w = \eta + P^\top \lambda_{wo} - \eta - P^\top \lambda_w = P^\top (\lambda_{wo} - \lambda_w)$.

The magnitude of x_w in equation (2.19) quantifies the importance of diversification. Additional portfolio rebalancing reduces the required amount of portfolio insurance contract from 1 to x_w when P is of full row rank as proved in the next corollary.

With little loss of generality, the matrix P is of full row rank with the available N risky assets exceeding the number of scenarios M. For example, consider a collection of futures contracts and options ranging across different maturities and strike prices. Although the payoffs of these derivative securities are correlated, it is important to clarify the distinction between linear dependencies in the columns of P versus its rows. In particular, correlation between the N risky securities influences the column rank of this payoff matrix but not its row rank. Indeed, diversification implies the more linearly dependent securities included in the optimization problem, the less drastic is the necessary portfolio rebalancing to achieve acceptability. More importantly, the row rank of P is a function of how "close" the M scenarios are to one another. However, since the scenarios involve extreme events, redundancy in the rows of P is not anticipated since this would imply the scenarios produce identical payoffs for each asset.

Corollary 2.5.1 The optimal amount of portfolio insurance to purchase, x_w, is strictly less than one unit if P is of full row rank.

Proof: The inequality $x_w \leq 1$ follows from $\lambda_w^\top QQ^\top \lambda_w = \lambda_w^\top (P\eta)^-$ by (2.17), which is equivalent to $\lambda_w^\top PP^\top \lambda_w + (\lambda_w^\top (P\eta)^-)^2 = \lambda_w^\top (P\eta)^-$. When P is of full row rank, $PP^\top$ is positive definite. This property implies $\lambda_w^\top PP^\top \lambda_w \geq 0$ which yields $(\lambda_w^\top (P\eta)^-)^2 \leq \lambda_w^\top (P\eta)^-$ and proves that

$$x_w = \lambda_w^\top (P\eta)^- \leq 1 \tag{2.21}$$

Thus, the optimal amount of insurance to purchase is strictly less than one unit.

The strict inequality in the above corollary reinforces the importance of diversification. Specifically, we are able to diversify risk more effectively once the insurance contract becomes available.

To summarize, it is not necessary for firms to purchase the entire insurance contract provided they engage in subsequent portfolio rebalancing. As indicated in the next corollary, fewer dollars are also required to be spent on portfolio insurance in this circumstance, a result that is later reinforced by Proposition 2.5.3.

Corollary 2.5.2 The dollar value of required insurance is less with portfolio rebalancing, $x_w IC_w < IC_{wo}$, if P is of full row rank.

Proof: This result follows from equations (2.19) and (2.20),

$$x_w IC_w = IC_{wo} - q^\top P^\top \lambda_w$$

and the fact that the last term $q^\top P^\top \lambda_w = q^\top (\eta_w - \eta)$ is positive. Indeed, $q^\top (\eta_w - \eta) > 0$ is a consequence of the condition $y = Q\delta^* - (P\eta)^+ = P\eta_w + (P\eta)^- x_w - (P\eta)^+ \geq 0$ from (2.17) which implies $P\eta_w + (P\eta)^- - (P\eta)^+ \geq 0$ since $1 > x_w \geq 0$. Therefore, $P\eta_w - P\eta \geq 0$ with strict inequality in at least one scenario and by the assumption of no arbitrage, $q^\top (\eta_w - \eta) > 0$.

In addition, equation (2.18) implies that neither δ_1 nor η_{wo} are optimal in the presence of the insurance contract. These statements are formalized in the following corollary.

Corollary 2.5.3 If η is an unacceptable portfolio and P is of full row rank, then neither

$$\delta_{wo} = \begin{bmatrix} \eta_{wo} \\ 0 \end{bmatrix} \quad \text{nor} \quad \delta_1 = \begin{bmatrix} \eta \\ 1 \end{bmatrix}$$

are optimal in the presence of the insurance contract.

Proof: If δ_{wo} is acceptable, then it is also acceptable in the presence of the insurance contract. But if δ_{wo} is optimal, then (2.14) and (2.18) jointly imply that

$$P^\top (\lambda_{wo} - \lambda_w) = 0 \quad \text{and} \quad ((P\eta)^-)^\top \lambda_w = 0$$

Hence, with the payoff matrix P being of full row rank, it follows that $\lambda_{wo} = \lambda_w$ with (2.13) implying $\lambda_{wo}^\top PP^\top \lambda_{wo} = 0$ which contradicts $PP^\top$ being positive definite since $(P\eta)^-$ is strictly greater than 0 in at least one scenario.

Hence δ_{wo} is not optimal. A similar contradiction is obtained if one assumes δ_1 is optimal.

The next proposition states that the two portfolio insurance prices, IC_{wo} and IC_w, is identical when the market is arbitrage-free.

Proposition 2.5.3 If η is an unacceptable portfolio, then the prices IC_{wo} and IC_w are equal.

Proof: The binding properties of the constraints in equations (2.12) and (2.16) imply:

$$\begin{cases} P\eta_{wo} = P\eta + (P\eta)^- \\ P\eta_w + (P\eta)^- x_w = P\eta + (P\eta)^- \end{cases}$$

It follows that η plus the insurance contract, η_{wo} and $\delta^* = \begin{bmatrix} \eta_w \\ x_w \end{bmatrix}$ all have the same payoff, $P\eta + (P\eta)^-$. By no arbitrage, their values at time zero are also equal with

$$\begin{cases} q^\top \eta_{wo} = q^\top \eta + IC_{wo} \\ q^\top \eta_w + IC_w \cdot x_w = q^\top \eta + IC_w \end{cases}$$

implying $IC_{wo} - IC_w = q^\top \eta_{wo} - q^\top \eta_w - IC_w \cdot x_w = 0$ which completes the proof.

In summary, prices for portfolio insurance without portfolio rebalancing and with portfolio rebalancing are given by Propositions 2.5.1 and 2.5.2 respectively. Additional portfolio rebalancing exploits the diversification benefit offered by the introduction of the insurance contract. As a result, the firm is able to purchase strictly less than one unit of the contract. However, with or without portfolio rebalancing, the price for one unit of portfolio insurance is identical according to Proposition 2.5.3. More intuition behind Proposition 2.5.3 is given in the next subsection.

2.5.3 Insurance and dollar-denominated risk

We now demonstrate that although the risk measure $\rho(\eta)$ is defined on portfolio weights, our results may be interpreted in terms of a dollar-denominated quantity. Furthermore, the dollar-denominated amount of rebalancing equals the price of portfolio insurance.

Specifically, the difference between η^* and η equals $\eta^* - \eta = P^T\lambda$, producing a dollar-denominated amount of risk equal to

$$\begin{aligned} q^T(\eta^* - \eta) &= q^T P^T \lambda \\ &= IC_{wo} \end{aligned} \tag{2.22}$$

Therefore, although risk is defined in terms of the l^2 norm on portfolio weights, it may be converted into the more traditional dollar-based domain and coincides with the price of portfolio insurance (with or without rebalancing).

As a consequence of equation (2.22), minimizing the distance in portfolio weights between η and the acceptance set is equivalent to minimizing the dollar-denominated amount of rebalancing. Therefore, the price of portfolio insurance equals the amount of rebalancing, in dollars, required to ensure the portfolio η becomes acceptable.

2.5.4 Example revisited

Returning to the example in section 2.3, let the price vector equal $q = [1, 1.3, 0.9]^\top$. As discussed in section 2.3, the second asset's price is obtained by no arbitrage as $-\frac{1.3}{3} + \frac{4}{3} = 0.90$.

Existing specifications imply $(P\eta)^- = [0, 1]^\top$, and $\eta = [1, 1, 0]^\top$ along with the payoff matrix P illustrates the results in Propositions 2.5.1 and 2.5.2. The vector λ_{wo} equals $[0.0673, 0.1635]^\top$, implying a price for portfolio insurance of $IC_{wo} = q^\top P^\top \lambda_{wo}$ which equals \$0.45. The λ_{wo} parameters are associated with two restrictions; preventing negative terminal values and not reducing positive terminal values.

The second optimization in equation (2.16) based on δ_0 and Q yields $\lambda_w = [0.0579, 0.1405]^\top$. According to Proposition 2.5.2, the price IC_w equals \$0.45, in accordance with Proposition 2.5.3.

However, the optimal amount of portfolio insurance to purchase is $x_w = \lambda_w^\top (P\eta)^- = (\lambda_w)_2 = 0.1405$, a quantity strictly less than one since P is of full row rank. Thus, with additional portfolio rebalancing, the dollar-denominated reduction in the amount of portfolio insurance that is required equals $IC_{wo} - x_w IC_w = (1 - 0.1405) \times 0.45 = \$0.39$.

2.6 CONCLUSION

A risk measure defined on the space of portfolio holdings rather than terminal values is proposed which enables diversification to reduce portfolio risk. Consequently, derivative and insurance contracts have important roles in risk management. Through portfolio rebalancing, our risk measure offers firms greater flexibility than coherent risk measures when complying with an external regulator. Indeed, our approach allows every asset in the portfolio, including riskfree capital, to be adjusted. Thus, as in the existing literature, risk is defined as the *distance* to an acceptance set. However, to incorporate diversification, the concept of distance is extended to include the risky assets as well as riskfree capital.

Our analysis incorporates market frictions such as illiquidity and transaction costs into the portfolio rebalancing decision. The price of portfolio insurance is also derived. When combined with the original portfolio, this contract ensures non-negative portfolio values in every scenario considered by the regulator. Furthermore, the amount of required portfolio insurance is determined by the firm's willingness to rebalance their portfolio once this contract is available.

APPENDIX A: PROOF OF PROPOSITION 2.2.2

Recall the properties of Proposition 2.2.1 regarding the acceptance set A_η.

Consider two portfolios η_1 and η_2 and let η_1^* be the closest portfolio on the acceptance set A_η. In other words, $\eta_1^* = \eta'$ such that $\inf\{\|\eta_1 - \eta'\|_2 : \eta' \in A_\eta\}$. Similarly define η_2^* as the equivalent quantity for η_2. Therefore, by definition,

$$\rho(\eta_1) = \|\eta_1 - \eta_1^*\|_2$$

$$\rho(\eta_2) = \|\eta_2 - \eta_2^*\|_2$$

and the following holds by the triangle inequality property of norms:

$$\|\eta_1 + \eta_2 - \eta_1^* - \eta_2^*\|_2 \leq \|\eta_1 - \eta_1^*\|_2 + \|\eta_2 - \eta_2^*\|_2 = \rho(\eta_1) + \rho(\eta_2)$$

However, the quantity $\eta_1^* + \eta_2^*$ is also in the acceptance set since A_η is convex and closed under multiplication by $\gamma \geq 0$. For two portfolios η_1^*, $\eta_2^* \in A_\eta$ convexity implies $\eta^* = \frac{1}{2}\eta_1^* + \frac{1}{2}\eta_2^* \in A_\eta$ while $2\eta^* = \eta_1^* + \eta_2^* \in A\eta$ as a consequence of A_η being closed under multiplication of positive scalars. Therefore,

$$\rho(\eta_1 + \eta_2) \leq \|\eta_1 + \eta_2 - \eta_1^* - \eta_2^*\|_2 = \|(\eta_1 + \eta_2) - (\eta_1^* + \eta_2^*)\|_2$$

since $\eta_1^* + \eta_2^*$ is an element of A_η but need not be optimal. Hence, $\rho(\eta_1 + \eta_2) \leq \rho\ (\eta_1) + \rho\ (\eta_2)$ and subadditivity is proved.

Consider two portfolios η_1 and η_2 and let $P\eta_1 \geq P\eta_2$ a.s. The proof for monotonicity follows by recognizing that $\eta_1 = \eta_1 - \eta_2 + \eta_2$ and $\rho(\eta_1 - \eta_2) = 0$ since the portfolio $\eta_1 - \eta_2$ always generates a non-negative payoff implying $\eta_1 - \eta_2 \in A_\eta$. Applying subadditivity, $\rho(\eta_1) = \rho(\eta_1 - \eta_2 + \eta_2) \leq \rho(\eta_2)$, demonstrates that $\rho(\eta_1) \leq \rho(\eta_2)$ and monotonicity is proved.

Consider a portfolio η and a scalar $\gamma \geq 0$. Define η^* as in the proof of subadditivity. The function $\rho(\eta)$ is defined as $\|\eta - \eta^*\|_2$ which implies that:

$$\gamma\rho(\eta) = \gamma\|\eta - \eta^*\|_2 = \|\gamma\eta - \gamma\eta^*\|_2 \geq \|\gamma\eta - (\gamma\eta)^*\|_2 = \rho(\gamma\eta)$$

since $\gamma\eta^*$ is in the acceptance set but need not be optimal in terms of minimizing the distance to the acceptable set. The reverse direction is proved by defining $\rho(\gamma\eta)$ as $\|\gamma\eta - (\gamma\eta)^*\|_2 = \gamma\|\eta - \frac{(\gamma\eta)^*}{\gamma}\|_2 \geq \gamma\rho(\eta)$ since $\frac{1}{\gamma}(\gamma\eta)^*$ is an element of A_η but need not be optimal. Thus, $\rho(\gamma\eta)$ and $\gamma\rho(\eta)$ are equal and positive homogeneity is proved.

Consider two portfolios η_1 and η_2 that differ only in terms of the riskfree asset with $\eta_{2,0} > \eta_{1,0}$. It suffices to show that $\rho(\eta_2) \leq \rho(\eta_1)$. Consider a portfolio that is a combination of η_1 and another portfolio $\gamma\eta_c$ for $\gamma \geq 0$ that consists entirely of an amount $\eta_{2,0} - \eta_{1,0}$ in riskfree capital. This new portfolio is equivalent to η_2 and implies that:

$$\eta_2 = \eta_1 + \gamma\eta_c \Rightarrow \rho(\eta_2) = \rho(\eta_1 + \gamma\eta_c) \leq \rho(\eta_1) + \rho(\gamma\eta_c)$$

using subadditivity. However, $\rho(\gamma\eta_c)$ equals zero since this portfolio is accepted by the regulator, $\gamma\eta_c \in A_\eta$. Hence, $\rho(\eta_2) \leq \rho(\eta_1)$ and the monotonicity of riskfree capital is proved.

Consider a portfolio $\eta \notin A_\eta$ such that $P\eta_i^- < 0$ for some i. It must be proved that $\rho(\eta) > 0$. Proceed by contradiction by supposing that $\rho(\eta) = 0$ which implies that $\eta \in A_\eta$ by Definition 2.2.2. However, Definition 2.2.1 requires that $P\eta \geq 0$ for $\eta \in A_\eta$, contradicting $P\eta_i^- < 0$ for any i. Hence, relevance is proved.

Consider a portfolio η that does not belong to the acceptance set. By the Separating Hyperplane Theorem (the acceptance set A_η is a convex subset of R^{N+1} (according to Proposition 2.2.1) and non-empty), there exists a point η^* on the boundary of A_η such that $\|\eta - \eta^*\|_2$ is the unique minimum distance of η from set A_η. Now consider any scalar γ and let $\tilde{u}$ be the unit directional vector in the direction $\eta^* - \eta$. The vector $\eta + \gamma \cdot \tilde{u}$ is a point along the path of minimum distance and proves the shortest path property:

$$
\begin{aligned}
\rho(\eta + \gamma \cdot \tilde{u}) &= \|\eta + \gamma \cdot \tilde{u} - \eta^*\|_2 \\
&= \left\| \eta + \gamma \cdot \left(\frac{\eta^* - \eta}{\|\eta - \eta^*\|_2} \right) - \eta^* \right\| \\
&= \left\| \eta^* - \eta - \gamma \cdot \left(\frac{\eta^* - \eta}{\|\eta - \eta^*\|_2} \right) \right\|_2 \\
&= \left(1 - \frac{\gamma}{\|\eta - \eta^*\|_2} \right) \|\eta - \eta^*\|_2 \\
&= \|\eta - \eta^*\|_2 - \gamma \\
&= \rho(\eta) - \gamma
\end{aligned}
$$

APPENDIX B: PROOF OF PROPOSITION 2.4.2

It is sufficient to prove that any two solutions to the linear complementary conditions in (2.10) yield the same optimal portfolio η^*. Therefore, our procedure is optimal. Let (y_1, λ_1) and (y_2, λ_2) denote two solutions to (2.10) with the following conditions:

$$
\begin{cases}
y_1 - PP^\top \lambda_1 = P\eta \\
y_2 - PP^\top \lambda_2 = P\eta \\
\lambda_1 \geq 0, \lambda_2 \geq 0, y_1 \geq 0, y_2 \geq 0 \\
\lambda_1^\top y_1 = 0, \lambda_2^\top y_2 = 0.
\end{cases}
\tag{2.23}
$$

We proceed to show:

$$
P^\top \lambda_1 = P^\top \lambda_2
$$

with both solutions generating the same optimal portfolio $\eta^* = \eta + P^\top \lambda_i$ for $i = 1, 2$. From (2.23):

$$
\begin{cases}
\lambda_1^\top P\eta = -\lambda_1^\top PP^\top \lambda_1 \\
\lambda_2^\top P\eta = -\lambda_2^\top PP^\top \lambda_2
\end{cases}
$$

Therefore,

$$\begin{aligned}(\lambda_1 + \lambda_2)^\top (y_1 + y_2) &= \lambda_1^\top y_2 + \lambda_2^\top y_1 \\ &= \lambda_1^\top PP^\top \lambda_2 + \lambda_2^\top PP^\top \lambda_1 + \lambda_1^\top P\eta + \lambda_2^\top P\eta \\ &= -(\lambda_1 - \lambda_2)^\top PP^\top (\lambda_1 - \lambda_2) \\ &\leq 0\end{aligned}$$

Since $\lambda_1 \geq 0$, $\lambda_1 \geq 0$, $y_1 \geq 0$, and $y_2 \geq 0$, it follows that:

$$(\lambda_1 - \lambda_2)^\top PP^\top (\lambda_1 - \lambda_2) = 0$$

which implies

$$P^\top (\lambda_1 - \lambda_2) = 0$$

Therefore, the optimal solution to equation (2.9) is

$$\eta^* = \eta + P^\top \lambda_1 = \eta + P^\top \lambda_2$$

which completes the proof.

REFERENCES

Artzner, P., Delbaen, F., Eber, J.M. and Heath, D. (1999) "Coherent Measures of Risk", *Mathematical Finance*, 9(3): 203–28.

Basak, S. and Shapiro, A. (2001) "Value-at-Risk-Based Risk Management: Optimal Policies and Asset Prices", *Review of Financial Studies*, 14(2): 371–405.

Follmer, H. and Schied, A. (2002) "Convex Measures of Risk and Trading Constraints", *Finance and Stochastics*, (6)4: 429–47.

Jarrow, R. (2002) "Put Option Premiums and Coherent Risk Measures", *Mathematical Finance*, 12(2): 125–34.

Luenberger, D.G. (1990) *Introduction to Linear and Nonlinear Programming,* 2nd edn (Reading, MA: Addison-Wesley).

Merton, R.C. (1998) "Applications of Option-Pricing: Twenty-Five Years Later", *American Economic Review*, 88(3): 323–49.

Rockafellar, R.T. and Ziemba, W.T. (2000) "Modified Risk Measures and Acceptance Sets", Discussion Paper, University of British Columbia.

CHAPTER 3

Sensitivity Analysis of Portfolio Volatility: Importance of Weights, Sectors and Impact of Trading Strategies

*Emanuele Borgonovo and Marco Percoco**

3.1 INTRODUCTION

This chapter discusses the application of a new method to the Sensitivity Analysis (SA) of portfolio properties and proposes an SA scheme that is capable of assessing the joint impact of changes in portfolio composition on portfolio volatility (σ_p).

Recent years have seen the fast development of models for the estimation of volatility. Studies in this field has moved from both a theoretical and an empirical need to explain evidence on volatility behavior (for example, volatility smile) that is not captured by constant volatility models as the Black and Scholes one (Black and Scholes, 1973; Duan, 1995; Shephard, 2005).

* We would like to thank Simone Manganelli for providing us with data and basic MatLab codes, and participants at EWFM (Brescia, May 2005) for useful comments on an earlier draft. We also thank Guillermo Baquero for precious comments at the EFMA conference (Milan, 2005). Financial support from Bocconi University is gratefully acknowledged.

Such models are usually categorized in the literature as Stochastic Volatility models[1] and autoregressive models, namely ARCH, GARCH and their generalizations (Bollerslev, 1986; Bollerslev and Engle, 1993; Engle, 1982). The rapid development of the computation technology has enabled the utilization of increasingly complex models. At the same time, some recent studies have shown that, as the use of these models becomes widespread, it is felt the need for the development of appropriate SA techniques capable of providing analysts with tools that fully exploit the information embedded in the models (Drudi, Generale and Majnoni, 1997; Manganelli, 2004; Manganelli, Ceci and Vecchiato, 2002; McNeal and Frey, 2000; Saltelli, 2003).

In a recent paper, Saltelli (2003) demonstrates how SA can be thought of as an essential ingredient in portfolio management. McNeal and Frey (2000) and Gourieroux, Laurent and Scaillet (2000) use partial derivatives (PD) to study the sensitivity of the Value at Risk (VaR) models. These authors derive analytically the expressions for the first and second derivatives of the VaR, and explain how they can be used to simplify statistical inference and to perform a local analysis of the VaR. A similar application of this technique can be found in Drudi, Generale and Majnoni (1997), where the sensitivity of risk assessment is tested with respect to the number of factors employed, the measures of volatility (conditional versus unconditional) and correlations (stable versus unstable), and the linearization of non-linear payoffs.

Manganelli, Ceci and Vecchiato (2002) propose a tool based on the calculation of the PDs of σ_p estimated via the GARCH model to help "risk managers to find out what the major sources of risk are, or allow them to evaluate the impact on the portfolio variance of a certain transaction." In a more recent paper by Manganelli the implications of the approach in asset allocation are discussed (Manganelli, 2004).

Recent studies in the SA literature have highlighted that PD-based SA suffers of several limitations when used for parameter impact evaluation and risk management purposes (Borgonovo and Apostolakis, 2001a; Borgonovo and Apostolakis, 2001b; Borgonovo and Peccati, 2004; Borgonovo and Peccati, 2005; Cheok, Parry and Sherry, 1998). More precisely, these studies show that utilizing a PD-based SA to evaluate the impact of parameter changes with respect to the generic model output:

1. is equivalent to neglecting the relative parameter changes, or, equivalently, to impose that all the parameters are varied in the same way;
2. does not allow the appreciation of the model sensitivity to changes in groups of parameters.

One could think of replacing the pure PD with the parameter Elasticity (E) (Simon and Blume, 1994). In this case Limitation 2 would still be in place, as E is not additive, and Limitation 1 would be replaced by the following (Borgonovo and Apostolakis, 2001a; Borgonovo and Apostolakis, 2001b; Borgonovo and Peccati, 2004; Borgonovo and Peccati, 2005):

3 utilizing E is equivalent to impose that all the parameters are changed by the same proportion.

These mathematical considerations translate into shortfalls in using PDs or E for the evaluation of the impact of weights on trading/reallocation strategies (TRS):[2] due to Limitations 1 and 3 it is not possible to evaluate the impact of generic portfolio composition relative changes, and due to limitation 2, it is not possible to test the impact of simultaneous changes in groups of weights. In TRS, however, simultaneous changes in more than one weight are involved and the relative changes are generally not uniform/proportional.

In this study, we show that the use of an alternative SA technique, namely the Differential Importance Measure (D), leads to overcome the two above mentioned limitations. D generalizes other local SA techniques, and, in particular, contains PDs and E as particular cases. D shares two important properties – (i) additivity and (ii) relative changes consideration. With reference to TRS analysis, we show that property (i) makes the computation of the sensitivity of σ_p on groups of weights straightforward, and property (ii) enables the analyst to accommodate any relative portfolio composition changes.

The empirical part of this chapter begins with the derivation of the analytical expression of individual portfolio weights D for the SA of σ_p estimated via Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models.[3] Thanks to D additivity, we obtain the importance of group of weights straightforwardly. As a result a method for the SA of σ_p with respect to weight groups (Sectors) is provided.

We apply the approach to a portfolio composed of 30 stocks of the Dow Jones index. We present numerical results for the impact of weights on σ_p in TRS involving uniform, proportional and optimal weight changes.[4] We show how the utilization of Savage Score Correlation Coefficients (SSCC) (Campolongo and Saltelli, 1997) can serve as a quantitative measure of similarity among TRS. We then analyse the portfolio with respect to its sectorial composition, examining how the results can be interpreted in terms of diversification across sectors.

In section 3.2, the definition of D and some SA background related to the recent developments in this field are discussed, and in section 3 analytical considerations on the SA of portfolio models highlighting the limitation of PDs and E are presented. In section 3.4 results for the SA of σ_p estimated via GARCH models are derived. Section 3.5 presents numerical results focusing on financial management aspects and their implication in the analysis of TRS. Section 3.6 offers conclusions and future research perspectives.

3.2 SENSITIVITY ANALYSIS BACKGROUND

Recently, the activity in the scientific field of SA of Model Output has been steadily growing, due to the increasing complexity of numerical models, whereby SA has acquired a key role in testing the correctness and corroborating the robustness of models in several disciplines. This has led to the development and application of several new SA techniques (Borgonovo and Apostolakis, 2001a; Saltelli, 1997; Saltelli, 1999; Saltelli, Tarantola and Chan, 1999; Turany and Rabitz, 2000). Most of the recent literature in portfolio management has proposed SA approaches based on PDs (Drudi, Generale and Majnoni, 1997; Gourieroux, Laurent and Scaillet, 2000; Manganelli, 2004; McNeal and Frey, 2000). In the next paragraphs we present the Differential Importance Measure (D) and discuss in detail its relation to PDs and Elasticity.

Let us consider the generic model output:

$$Y = f(\mathbf{x}) \tag{3.1}$$

where $\mathbf{x} = \{x_i, i = 1, 2, \ldots, n\}$ is the set of the input parameters. Let also

$$d\mathbf{x} = [dx_1, dx_2, \ldots, dx_n]^T$$

denote the vector of changes.

If $f(\mathbf{x})$ is differentiable, then the differential importance of x_s at $\mathbf{x}^0$ is defined as (Borgonovo and Peccati, 2004):

$$D_s(\mathbf{x}^0, d\mathbf{x}) = \frac{df_s(\mathbf{x}^0)}{df(\mathbf{x}^0)} = \frac{f_s(\mathbf{x}^0)\, dx_s}{\sum_{j=1}^{n} f_j(\mathbf{x}^0)\, dx_j} \tag{3.2}$$

D can be interpreted as the ratio of the (infinitesimal) change in Y caused by a change in x_s and the total change in Y caused by a change in all the parameters. Thus, D is the normalized change in Y provoked by a change in parameter x_s. It can be shown that (Borgonovo and Apostolakis, 2001a; Borgonovo and Apostolakis, 2001b; Borgonovo and Peccati, 2004; Borgonovo and Peccati, 2005):

A D shares the additivity property with respect to the various inputs, for example, the impact of the change in some set of parameters coincides with the sum of the individual parameter impacts. More formally, let $S \subseteq \{1, 2, \ldots, n\}$ identify some subset of interest of the input set. We have:

$$D_S(\mathbf{x}^0, d\mathbf{x}) = \frac{\sum_{s \in S} f_s(\mathbf{x}^0)\, dx_s}{\sum_{j=1}^{n} f_j(\mathbf{x}^0)\, dx_j} = \sum_{S \in S} D_s(\mathbf{x}^0, d\mathbf{x}) \tag{3.3}$$

As a consequence,

$$\sum_{s=1}^{n} D_s(\mathbf{x}^0, d\mathbf{x}) = 1 \tag{3.4}$$

for example, the sum of the D_i $(i=1,\ldots,n)$ of all parameters is always equal to unity.

B Equation (3.2) shows that D accounts for the relative parameters changes through the dependence on $d\mathbf{x}$. In fact, equation (3.2) can be rewritten as:

$$D_s(\mathbf{x}^0, d\mathbf{x}) = \frac{f_s(\mathbf{x}^0)}{\sum_{j=1}^{n} f_j(\mathbf{x}^0)\frac{dx_j}{dx_s}} \tag{3.5}$$

In the hypothesis of uniform parameter changes (H1) $(dx_j = dx_s \; \forall j, s)$, one finds:

$$D1_s(\mathbf{x}^0) = \frac{f_s(\mathbf{x}^0)}{\sum_{j=1}^{n} f_j(\mathbf{x}^0)} \tag{3.6}$$

In the hypothesis of proportional changes (H2) $\left(\frac{dx_j}{x_j^0} = \omega \forall_j\right)$, one finds:

$$D2_s(\mathbf{x}^0) = \frac{f_s(\mathbf{x}^0) \cdot x_s^0}{\sum_{j=1}^{n} f_j(\mathbf{x}^0) \cdot x_j^0} \tag{3.7}$$

It can be shown that D generalizes other local SA techniques as the Fussell–Vesely importance measure and Local Importance Measures based on normalized partial derivatives, also known as Criticality Importance or E.[5] More specifically, in case H2 it holds that (Borgonovo and Peccati, 2004):

$$D2_s(\mathbf{x}^0) = \frac{E_s(\mathbf{x}^0)}{\sum_{j=1}^{n} E_j(\mathbf{x}^0)} \tag{3.8}$$

where $E_s(\mathbf{x}^0)$ is the elasticity of Y with respect to x_s at $\mathbf{x}^0$. Equation (3.8) shows that E produces the importance of parameters for proportional changes in their. In the next section, we examine how these results affect the SA of portfolio properties.

3.3 EFFECT OF RELATIVE WEIGHT CHANGES

We now show a first portfolio management implication of equations (3.6) and (3.7): relative weight changes cannot be neglected when evaluating the impact of TRS on portfolio properties. We begin with a simple example.

Example 1, let

$$v = a_1 v_1 + a_2 v_2 \tag{3.9}$$

be the value of a portfolio at a certain point in time. Let also $a_1 = 100$, $a_2 = 9900$, $v_1 = 10$EUR and $v_2 = 5$EUR. The total value of the portfolio is then $v = 50500$EUR. Let us undertake the SA of v with respect to the weights, with reference to two hypothetical trading strategies. The first TRS is to buy one additional stock of 1 and 2. In this case we have a unitary change in a_1 and a_2, for example, $da_1 = da_2 = 1$. Applying equation (3.6), one gets: $D1_1 = 0.667$ and $D1_2 = 0.333$. This result means that asset 1 is the most influential if a TRS involving uniform weight changes is considered. Let us consider the case in which the trader opts for a proportional change in the two assets, for example, he buys (or sells) ϖ% in each of them. Applying equation (3.7), one gets: $D2_1 = 0.02$ and $D2_2 = 0.98$. In this case asset 2 would be the most influential one on the portfolio value.

The above example clearly shows that to evaluate the impact of changes in portfolio composition one must consider not only the rate of change (PD) of the portfolio with respect to the weights, but also the relative way in which the weights are changed [equation (3.5)].[6]

We now extend the meaning of "example" showing that evaluating the impact of portfolio changes by means of the sole *PD* is equivalent to make the implicit assumption of a TRS involving uniform weight changes.

Proposition 1 Ranking weights based on PDs is equivalent to considering TRS involving uniform weight changes.

Proof: Let $Y^0 = f(\mathbf{a}^0)$ denote a n-asset, differentiable portfolio property as a function of the current allocation $\mathbf{a}^0$. We use the symbol $a_i \succ a_j$ to indicate that weight a_i is more important than a_j (Borgonovo, 2001b). If one utilizes *PD* to rank weights, then one says that a_i is more important than a_j when the magnitude of the change in Y^0 provoked by a change in a_i namely $|f_i(\mathbf{a}^0)|$, is greater than the magnitude of the change in Y^0 provoked by a change in a_j:

$$a_i \succ a_j \Leftrightarrow |f_i(\mathbf{a}^0)| > |f_j(\mathbf{a}^0)| \tag{3.10}$$

Nothing changes in $|f_i(\mathbf{a}^0)| > |f_j(\mathbf{a}^0)|$ if one multiplies and divides both sides for $|\sum_{k=1}^{n} f_k(\mathbf{a}^0)|$. One gets:

$$a_i \succ a_j \Leftrightarrow |f_i(\mathbf{a}^0)| > |f_j(\mathbf{a}^0)| \Leftrightarrow \frac{|f_i(a^0)|}{|\sum_{k=1}^{n} f_k(a^0)|} > \frac{|f_j(a^0)|}{|\sum_{k=1}^{n} f_k(a^0)|} \tag{3.11}$$

The above is then equivalent to stating:

$$a_i \succ a_j \Leftrightarrow |D1_i(\mathbf{a}^0)| > |D1_j(\mathbf{a}^0)| \tag{3.12}$$

proving that ranking based on PDs is equivalent to ranking based on $D1_s(\mathbf{a}^0)$, for example, to stating an assumption of uniform parameter changes. Q.E.D.

We now demonstrate that, if instead of PDs, the E of the weights were considered, this would be equivalent to make the implicit assumption of a TRS involving proportional weight changes.

Proposition 2 Ranking weights based on "elasticities" is equivalent to consider TRS involving proportional weight changes.

Proof: In the same settings as in the above proof, let us assume that one ranks weights utilizing Elasticity; for example:

$$a_i \succ a_j \Leftrightarrow |E_i(\mathbf{a}^0)| > |E_j(\mathbf{a}^0)| \tag{3.13}$$

Nothing changes in $|E_i(\mathbf{a}^0)| > |E_j(\mathbf{a}^0)|$ if one multiplies and divides both sides for $\left|\sum_{k=1}^{n} E_k(\mathbf{a}^0)\right|$; one gets:

$$a_i \succ a_j \Leftrightarrow |E_i(\mathbf{a}^0)| > |E_j(\mathbf{a}^0)| \Leftrightarrow \frac{|E_i(a^0)|}{\left|\sum_{k=1}^{n} E_k(a^0)\right|} > \frac{|E_j(a^0)|}{\left|\sum_{k=1}^{n} E_k(a^0)\right|} \tag{3.14}$$

Utilizing equation (3.8), the above is then equivalent to stating:

$$a_i \succ a_j \Leftrightarrow |D2_i(\mathbf{a}^0)| > |D2_j(\mathbf{a}^0)| \tag{3.15}$$

for example, analysing the influence of weights based on E is equivalent to stating an assumption of proportional parameter changes. Q.E.D.

As mentioned in section 3.1, the implication on the analysis of TRS of Propositions is that PDs/E are appropriate SA measures on the subset of strategies involving a uniform/proportional change in the portfolio composition. As discussed in section 3.2, these limitations can be overcome by applying D. In fact, equation (3.2) accommodates any TRS. Furthermore it includes PDs and E as particular cases, as equations (3.6) and (3.7) show. In the next section, we illustrate the application of these concepts to the SA of σ_p estimated via GARCH models.

3.4 IMPORTANCE OF PORTFOLIO WEIGHTS IN GARCH VOLATILITY ESTIMATION MODELS

Models of time-varying volatility have been popular since the early 1990s in empirical research area of finance, following the influential papers by Engle (1982). GARCH models are well-known in the time series Econometrics literature. From the initial concern with an economic phenomenon, for example, time-varying and autoregressive variance of inflation (Engle, 1982), the utilization of these models has become widespread, ranging from asset management (Manganelli, 2004), to derivatives pricing (Duan and Zhang, 2001) and risk management (Manganelli, Ceci and Vecchiato, 2002).

A stochastic process is called *GARCH*(p,q) if its time-varying conditional variance is heteroscedastic with both autoregression and moving average (Bollerslev and Engle, 1993):

$$y_t = \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2) \tag{3.16}$$

$$h^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}^2 \tag{3.17}$$

In equation (3.17) autoregression in the GARCH(p,q) process squared residuals has an order of q, and the moving average component has an order of p.

One of the features that has traditionally made GARCH models popular is the fact that parameters of the model can be straightforwardly estimated, since construction of the ML function is made direct from the fact that the model is formulated "in terms of the distribution of the one step ahead prediction error" (Shephard, 2005). The conditional log-likelihood of y_{t+1} is (Campbell, Lo and McKinley, 1997; Hull, 1999; Noh, 1997):

$$L_T(y_1, \ldots, y_T) =_{t=1}^{T} l_t(y_{t+1}; q) \tag{3.18}$$

where

$$l_t(y_{t+1}; q) = \log \left[N\left(\frac{y_{t+1}}{h_t} \right) \right] - \frac{\log (h_t^2)}{2}$$

is the one-step conditional log-likelihood function, θ is the vector of the parameters of the model and N(·) is a standard normal density function. Parameters can then be estimated by maximization of equation (3.18).

Throughout our discussion we consider the following GARCH(p,q) process for a portfolio of n assets (Manganelli, 2004):

$$y_t = \sqrt{h_t} \varepsilon_t \quad \varepsilon_t \sim N(0, 1) \tag{3.19}$$

$$h_t = z_t \theta \tag{3.20}$$

where y_t is the return of a portfolio composed by $n+1$ assets calculated as $y_t = \sum_{i=1}^{n+1} a_i y_{t,i}$, where a_i and $y_{t,i}$ are the weight and the return respectively of asset i; $z_t = (1, y_{t-1}^2, \ldots, y_{t-q}^2, h_{t-1}, \ldots, h_{t-q})$ and $\theta = (a_0, a_1, \ldots, a_q, b_1, \ldots, b_p)$

We now derive the expression of the differential importance of weights on portfolio volatility.

Proposition 3 The differential importance of weight a_i with respect to σ_p^{GARCH} for any change in portfolio composition is given by:

$$D_i(\mathbf{a}^0, \mathbf{da}) = \left. \frac{\left(\dfrac{\partial z_t}{\partial a_i}\theta + z_t \dfrac{\partial \theta}{\partial a_i} \right) da_i}{\sum_{j=1}^{n} \left(\dfrac{\partial z_t}{\partial a_j}\theta + z_t \dfrac{\partial \theta}{\partial a_j} \right) da_j} \right|_{a^0} \tag{3.21}$$

Proof: The proof is in the Appendix.

Equation (3.21) determines the analytical expression of the importance of portfolio weights with respect to σ_p estimated via a GARCH model for the generic TRS. From equation (3.21), it is then straightforward to estimate the importance of weights for strategies that foresee a uniform or a proportional change in weights.

Proposition 4 The importance of individual weights with respect to σ_p^{GARCH} for a TRS that assumes of uniform weight changes is:

$$D1_i(\mathbf{a}^0, \text{d}\mathbf{a}) = \left. \frac{\left(\dfrac{\partial z_t}{\partial a_i}\theta + z_t \dfrac{\partial \theta}{\partial a_i} \right)}{\sum_{j=1}^{n} \left(\dfrac{\partial z_t}{\partial a_j}\theta + z_t \dfrac{\partial \theta}{\partial a_j} \right)} \right|_{a^0} \tag{3.22}$$

Proof: Combine equation (3.21) with equation (3.6).

Recalling Proposition 1, equation (3.22) shows that utilizing PDs [equation (3.27)] to evaluate the impact of weights on σ_p^{GARCH}, one would not evaluate the impact of any transaction, but only of TRS involving a uniform change in the portfolio weights.

Proposition 5 The importance of individual weights with respect to σ_p^{GARCH} for a TRS that assumes proportional weight changes is:

$$D2_i(\mathbf{a}^0) = \left. \frac{\left(\dfrac{\partial z_t}{\partial a_i}\theta + z_t \dfrac{\partial \theta}{\partial a_i} \right) a_i^0}{\sum_{j=1}^{n} \left(\dfrac{\partial z_t}{\partial a_j}\theta + z_t \dfrac{\partial \theta}{\partial a_j} \right) a_j^0} \right|_{a^0} \tag{3.23}$$

Proof: Combine equation (3.21) with equation (3.7).

Recalling Propositions 3.23, equation (3.23) states that utilizing elasticity to evaluate the importance of weights with respect to σ_p^{GARCH} would be equivalent to consider only the TRS involving proportional relative weight changes.

Suppose now that the analyst wants to evaluate the impact of changing group A vs. group B of portfolio weights. Groups A and B could be, for instance, the set of assets belonging to two selected Sectors. Let

$S_A = \{a_{1A}, a_{2A}, \ldots, a_{kA}\}$ and $S_B = \{a_{1B}, a_{2B}, \ldots, a_{mB}\}$, with kA and mB lower than n be the assets belonging to group A and B respectively. Then:

Proposition 6 The influence of a change in set A of weights with respect to σ_p^{GARCH} is determined by:

$$D_{S_A}(\mathbf{a}^0, d\mathbf{a}) = \left. \frac{\sum_{i=1,\ldots,kA} \left(\frac{\partial z_t}{\partial a_{iA}} \theta + z_t \frac{\partial \theta}{\partial a_{iA}} \right) da^0_{Ai}}{\sum_{j=1}^{n} \left(\frac{\partial z_t}{\partial a_j} \theta + z_t \frac{\partial \theta}{\partial a_j} \right) a_j^0} \right|_{a^0} \tag{3.24}$$

for example, it is the sum of the importance of the weights in set A.

Proof: Combine equation (3.21) with equation (3.3).

The above result is a consequence of D additivity property and cannot be obtained utilizing the PDs or E as a means for computing the sensitivity of σ_p on the portfolio weights.

Similarly, the influence of a change in set B weights is determined by:

$$D_{S_B}(\mathbf{a}^0, d\mathbf{a}) = \left. \frac{\sum_{i=1,\ldots,mB} \left(\frac{\partial z_t}{\partial a_{iB}} \theta + z_t \frac{\partial \theta}{\partial a_{iB}} \right) da^0_{iB}}{\sum_{j=1}^{n} \left(\frac{\partial z_t}{\partial a_j} \theta + z_t \frac{\partial \theta}{\partial a_j} \right) a_j^0} \right|_{a^0} \tag{3.25}$$

Thus, if $|D_{S_B}(\mathbf{a}^0, da)| > |D_{S_A}(\mathbf{a}^0, da)|$ then set B is more influential or as influential as set A on σ_p. Proposition 6 enables to perform the joint sensitivity of σ_p on sets of portfolio weights in a straightforward way.

The next section is devoted to the illustration of empirical results and insights found by application of the results in Propositions 1–6 and to a portfolio composed by the 30 stocks of the Dow Jones index.

3.5 EMPIRICAL RESULTS: TRADING STRATEGIES THROUGH SENSITIVITY ANALYSIS

In this section we present the implications of Propositions 1–6 in the analysis of trading/reallocation strategies. We consider a portfolio of 30 stocks composing the Dow Jones Industrial Average index (Table 3.1), as of 11 March 2002. Daily returns cover the period ranging from 2 January 1992 through 11 March 2002.

The classical portfolio choice problem (Campbell, Lo and McKinley, 1997; Taggart, 1996) in its dual form is written as:

$$\begin{aligned} &\mathbf{a}^* = \arg\mathbf{a}\min[\text{Var}(\mathbf{a}'\mathbf{y})] \\ &\qquad \text{s.t.} \\ &\qquad \text{E}(\mathbf{a}'\mathbf{y}) = \mu \end{aligned} \tag{3.26}$$

Table 3.1 Composition of the Dow Jones Industrial Average Index

Stock	Sector
Alcoa In.	Manufacturing
American Express Co.	Financial services
AT&T Corp.	Telecommunication and ICT
Boeing Co.	Manufacturing
Caterpillar Inc.	Other
Citigroup Inc.	Manufacturing
Coca-Cola Co.	Manufacturing
Walt Disney Co.	Others
E.I. DuPont de Nemours & Co.	Manufacturing
Eastman Kodak Co.	Manufacturing
Exxon Mobil Corp.	Energy
General Electric Co.	Energy
General Motors Corp.	Manufacturing
Hewlett-Packard Co.	Telecommunication and ICT
Hope Depot Inc.	Others
Honeywell International Inc.	Telecommunication and ICT
Intel Corp.	Telecommunication and ICT
International Business Machines Corp.	Telecommunication and ICT
International Paper Co.	Manufacturing
Johnson&Johnson	Manufacturing
J.P. Morgan Chase & Co.	Financial Services
McDonald's Corp.	Others
Merck & Co.	Others
Microsoft Corp.	Telecommunication and ICT
3M Co.	Telecommunication and ICT
Altria Group Inc.	Others
Procter & Gamble	Manufacturing
SBS Communications Inc.	Telecommunication and ICT
United Technologies Corp.	Others
Wal-Mart Stores Inc.	Others

Problem (3.26) states that the optimal portfolio composition ($\mathbf{a}^*$) is the one that minimizes volatility for a given return. Since asset values change with time, portfolio composition should be re-assessed in order to match $\mathbf{a}^*$. Given a suboptimal portfolio, an ideal strategy might consist in the fastest or cheapest way to reach the minimum variance (as of March 2002). Manganelli (2004) proposes a strategy based on volatility sensitivity analysis (VSA) to achieve efficient reallocation. We refer to this strategy as to the "optimal" strategy.[7]

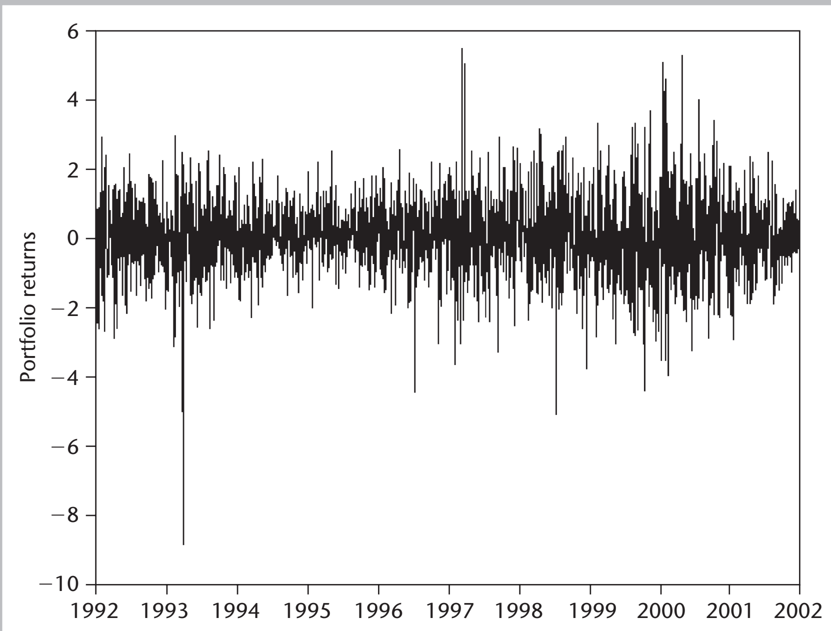


Figure 3.1 Portfolio returns

Figures 3.1 and 3.2 depict daily returns of the EWMA optimal portfolio and the time series of the estimated variance respectively.

In order to assess the importance and impact of weights on σ_p, we applied D. As we are to illustrate, this gives one the possibility of calibrating strategies with respect to a specific stock or a group of assets, in consideration of both the relative size of the imposed relative weight change [equation (3.5)].

Table 3.2 shows the SA results for the considered stocks under the hypothesis of uniform and proportional changes and for the optimal strategy. As expected, the results imply that the impact of weights on the portfolio volatility varies depending on the TRS (see also Figure 3.3).

This is also evident from Table 3.3 that shows the ranking of assets and Figure 3.3 that compares the asset importance in the three strategies.

Let us analyse the ranking in somewhat greater detail. Honeywell International Inc. is the most influential stock for uniform and proportional strategies (for example, it would rank first if one used *PDs* or *E*); it is the fourth most influential for the optimal strategy. Alcoa In. is the most important asset in the optimal strategy; it ranks 7th and 5th for uniform and proportional strategies respectively. Hewlett-Packard Co. is the least influential weight in the case of uniform changes; it ranks however 3rd for proportional changes and 7th in the optimal strategy. We note that Table 3.2

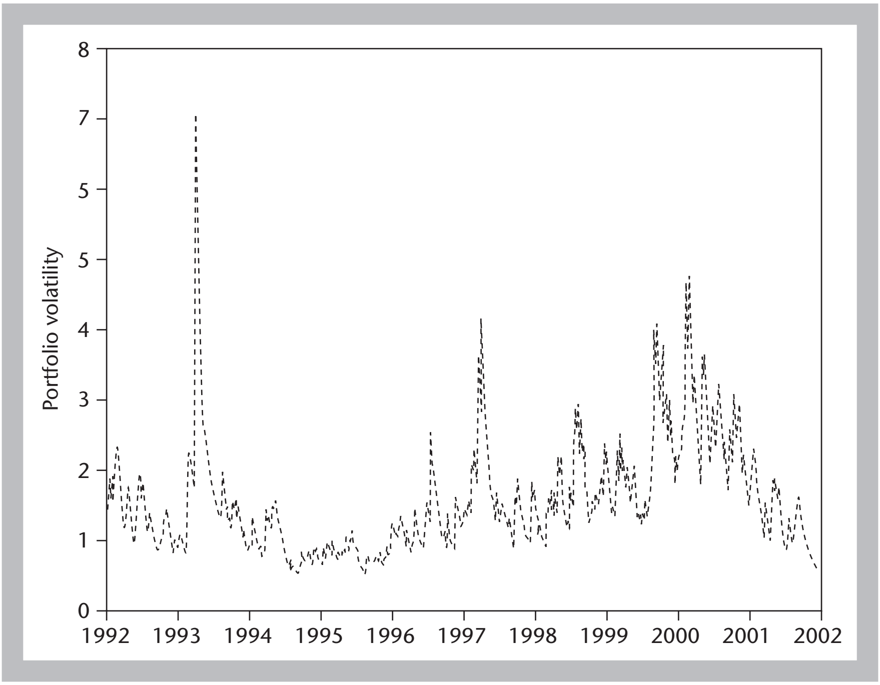


Figure 3.2 Volatility of the portfolio

columns 1 and 2 are the SA results that would be obtained if an analyst made use of PDs or E respectively (see Propositions 1 and 2). The difference in ranking shows that conclusions on σ_p sensitivity obtained by making use of PD/E cannot be extended to TRS other than uniform/proportional ones respectively.

Let us now discuss how the use of Savage Score correlation coefficients (SSCC) can enable the analyst to have a quantitative measure of TRS similarity. We refer the reader to Campolongo and Saltelli (1997) for the mathematical definition of SSCC's. For the purposes of this paper, we need only to recall that a high SSCC between two rankings means that the most and least influential assets tend to be the same.

Table 3.4 displays the Savage Score correlation matrix for the three strategies for the portfolio at hand.

The interpretation of Table 3.4 is as follows. Rankings in proportional and optimal strategy tend to be more correlated than rankings of the uniform and optimal strategy. This suggests that the optimal strategy is closer to a proportional rather than a uniform one from an SA viewpoint.

Finally a note on the information that one can infer from the quantification of the total change in volatility provoked by the three strategies (Table 3.2). In the case shown in Table 3.2, the TRS imposing uniform changes will result in a decrease of −7.0 in the volatility and the optimal strategy would reduce

Table 3.2 DIMs for different strategies

Stock	Uniform changes	Proportional changes
Alcoa In.	0.051073	0.15398633
American Express Co.	0.069683	0.12372132
AT&T Corp.	0.044879	0.1134629
Boeing Co.	0.056514	0.05225033
Caterpillar Inc.	0.061969	0.03523841
Citigroup Inc.	0.049297	−0.0453762
Coca-Cola Co.	0.014888	−0.1265676
Walt Disney Co.	0.043174	−0.0264934
E.I. DuPont de Nemours & Co.	0.040858	−0.1400711
Eastman Kodak Co.	0.029535	−0.0074914
Exxon Mobil Corp.	0.037164	0.13972337
General Electric Co.	0.0334	0.0073784
General Motors Corp.	0.046271	0.0660631
Hewlett-Packard Co.	−0.03921	0.23467628
Hope Depot Inc.	0.038471	0.0291163
Honeywell International Inc.	0.091107	0.45881233
Intel Corp.	0.035303	−0.0385615
International Business Machines Corp.	0.054383	0.20823972
International Paper Co.	0.035303	−0.0466493
Johnson&Johnson	0.005725	−0.0539873
J.P. Morgan Chase & Co.	0.039949	−0.077956
McDonald's Corp.	−0.00473	0.00894133
Merck & Co.	0.026424	−0.0438889
Microsoft Corp.	0.023015	−0.0013181
3M Co.	0.035772	−0.1002449
Altria Group Inc.	−0.01725	0.26155141
Procter & Gamble	0.01172	0.01668587
SBS Communications Inc.	0.027916	−0.1251667
United Technologies Corp.	0.057395	−0.076075
Total Differential	**−7.039**	**0.1720617**

volatility by −0.15 while the proportional changes strategy will increase volatility by 0.17. The observations are: (i) clearly the proportional strategy is the worst one, since it increases volatility, leading to higher risk for the same return level; (ii) the uniform strategy seems the one that diversifies risk the most, but it is not optimal since it would not respect the return constraint; (iii) the magnitude of the changes (−0.15 vs. 0.17) shows that the effect of a proportional strategy is closer to that of a proportional strategy, and seems consistent with the SSCC results.

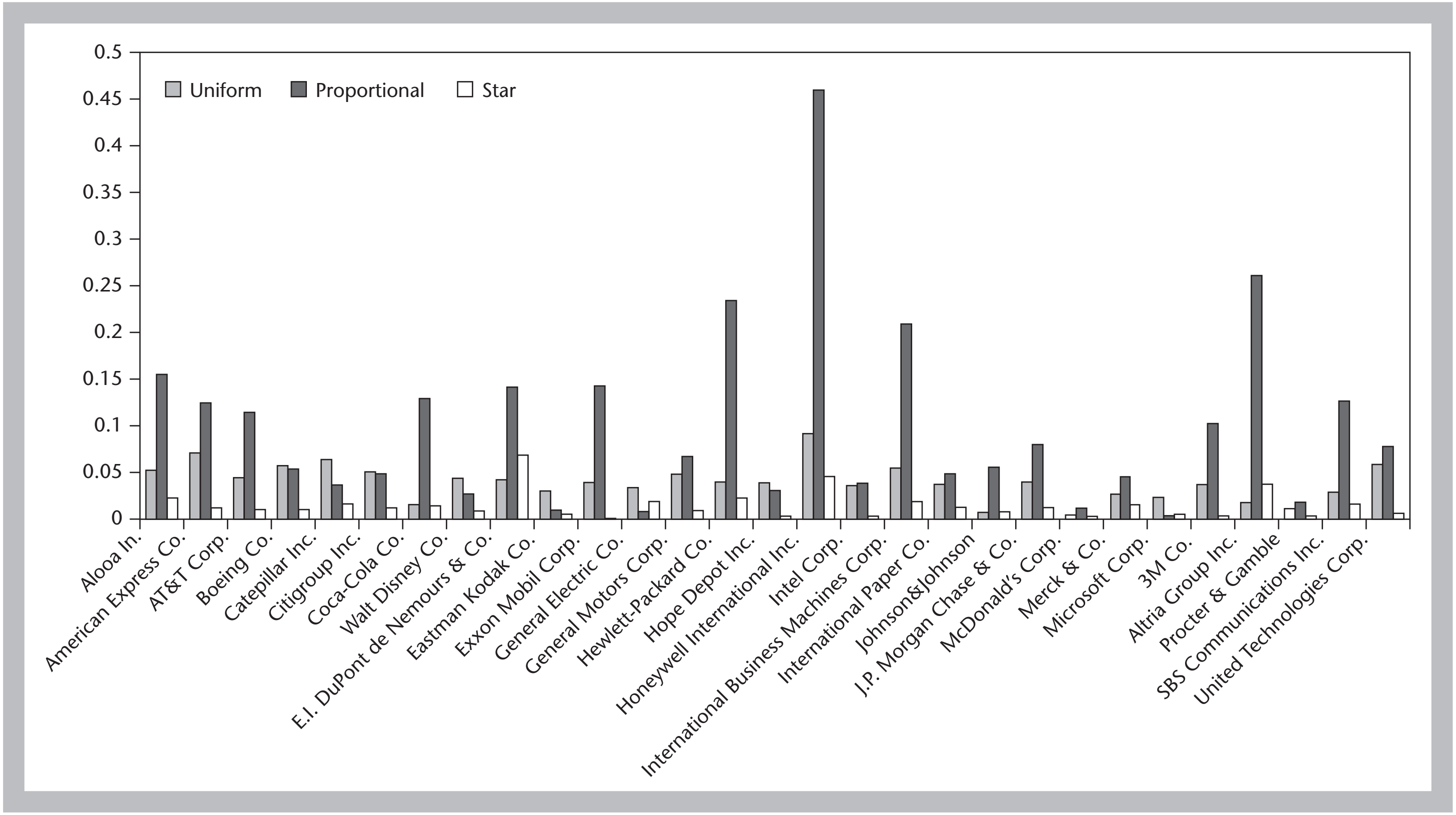


Figure 3.3 Absolute value of importance measures for stocks

Table 3.3 Rankings of stocks according to the DIM

Rank	Uniform changes	Proportional changes
1	Honeywell International Inc.	Honeywell International Inc.
2	American Express Co.	Altria Group Inc.
3	Caterpillar Inc.	Hewlett-Packard Co.
4	United Technologies Corp.	International Business Machines Corp.
5	Boeing Co.	Alcoa In.
6	International Business Machines Corp.	Exxon Mobil Corp.
7	Alcoa In.	American Express Co.
8	Citigroup Inc.	AT&T Corp.
9	General Motors Corp.	General Motors Corp.
10	AT&T Corp.	Boeing Co.
11	Walt Disney Co.	Caterpillar Inc.
12	E.I. DuPont de Nemours & Co.	Hope Depot Inc.
13	J.P. Morgan Chase & Co.	Procter & Gamble
14	Hope Depot Inc.	McDonald's Corp.
15	Exxon Mobil Corp.	General Electric Co.
16	3M Co.	Microsoft Corp.
17	Intel Corp.	Eastman Kodak Co.
18	International Paper Co.	Walt Disney Co.
19	General Electric Co.	Intel Corp.
20	Eastman Kodak Co.	Merck & Co.
21	SBS Communications Inc.	Citigroup Inc.
22	Merck & Co.	International Paper Co.
23	Microsoft Corp.	Johnson&Johnson
24	Coca-Cola Co.	United Technologies Corp.
25	Procter & Gamble	J.P. Morgan Chase & Co.
26	Johnson&Johnson	3M Co.
27	McDonald's Corp.	SBS Communications Inc.
28	Altria Group Inc.	Coca-Cola Co.
29	Hewlett-Packard Co.	E.I. DuPont de Nemours & Co.

These results are also influenced by the following two main factors:

- initial conditions (in our case the EWMA optimal portfolio);
- the choice of the feasible adjustment strategy.

Regarding the second point, even if it is not the proper object of the present paper, it should be stated that by solving equation (3.26), no information are provided on the optimal strategy to reach the optimum point. Computing D provides information on *how* the optimum point is reached, by calculating

Table 3.4 DIMs by sector and strategy

Sector	Strategy		
	Uniform change	Proportional change	Optimal strategy
Manufacturing	0.353857	−0.050542	0.346473
Financial services	0.158929	0.000389	0.088251
Tcom and ICT	0.273164	0.7499	0.077969
Energy	0.070564	0.147101	0.318412
Others	0.143486	0.153151	0.168886

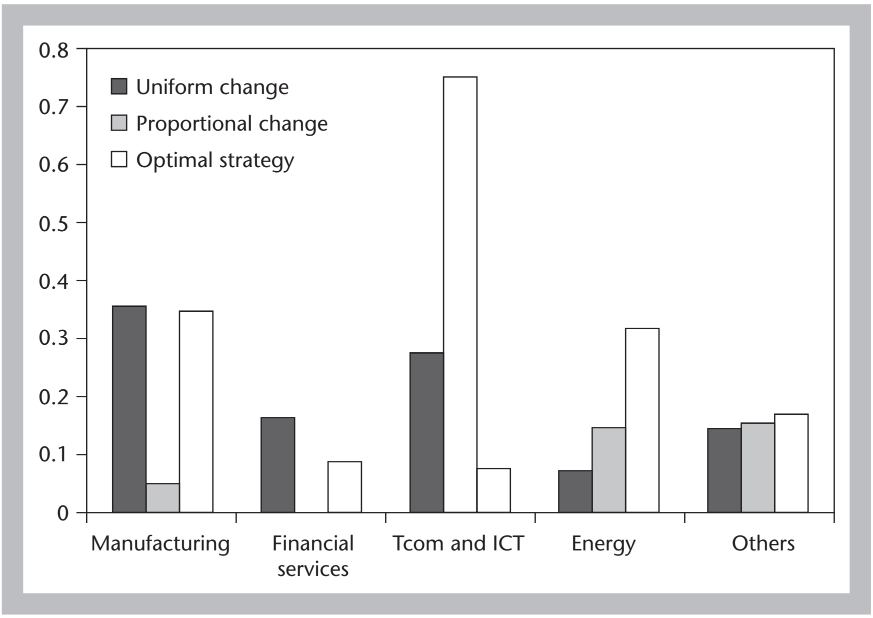


Figure 3.4 DIM by sector, absolute values

both the sign and magnitude of the impact of a given stock and by ranking stocks according to that impact.

In all the previous cases, D provides the importance of a single asset in inducing changes in volatility. However, it allows also finding the aggregate importance in a straightforward way. More precisely, having equations (3.24) and (3.25) in mind, in the uniform changes case, stocks ranking from 1–10 account for 46 percent of the result, whilst the lowest ranking 10 only for 18.38 percent. This implies that the former assets are the most important for the considered strategy.

The additive property of the D is then useful in calibrating TRS aiming at diversification across sectors. In our simple 30-stock portfolio, let us consider five types of industries: manufacturing, financial services, telecommunication and information and communication technologies, energy, and others (as a residual category). As Table 3.1 shows, stocks have been divided in those groups by following a simple criterion over the corresponding company main area of activity. Of course, the categorization may result as loose or not adequate in some cases, but it has the advantages of considering a small number of industries and it does fit our scope of showing the empirical application of the D to a portfolio GARCH(1,1).

Table 3.4 shows the D for the five sectors. By additivity, they are the sum of the D of the assets belonging to the group, over different strategies. Figure 3.4 presents the absolute values of D in Table 3.4.

It is interesting to note that:

- Manufacturing is the most important sector for the uniform and optimal strategies. However, it is among the least influential (ranking 4th) for a proportional strategy;
- Telecommunication and ICT is the most influential sector if a proportional strategy is assumed; it ranks second for the uniform strategy and least for the proportional one;
- Financial services assets rank 3rd, 5th and 4th respectively; hence they tend to have a low impact with respect to the change in portfolio volatility in the various strategies.
- Energy assets rank 5th, 3rd and 2nd in the three strategies: their influence varies significantly across trading/reallocation strategies, as they are the least influential for a uniform strategy, while ranking 2nd for the optimal strategy;
- Others assets rank 4th, 2nd and 3rd for the three strategies respectively: they tend to maintain an intermediate relevance across the strategies.

A final note, when recalling the definition of D and the interpretation of volatility, the above results can be readily interpreted in terms of risk. For example, considering the optimal strategy, most of the diversification would be due to the Manufacturing and Energy categories, since the corresponding assets are responsible for 66 percent of the change in σ_p.

3.6 CONCLUSION

In this study we have illustrated the sensitivity analysis (SA) of portfolio volatility (σ_p) estimation models. We have suggested that performing the

SA based on partial derivatives (PD) or elasticity (E) leads to limitations in testing:

1 the sensitivity of a portfolio property with respect to simultaneous changes in several parameters;

2 the sensitivity of a portfolio property for a strategy involving relative weight changes other than uniform or proportional ones.

In particular, we have proven that (i) utilizing PDs to rank weights is equivalent to state the assumption that all the weights are changed by same amount; and (ii) utilizing E would be equivalent to state that all the weights are changed by the same proportion.

We have illustrated that these limitations can be overcome if one utilizes the Differential Importance Measure (D) as SA method. We have shown that D: (i) makes the evaluation of the impact of changes in several weights straightforward thanks to the additivity property; and (ii) allows to take into account the effect of arbitrary relative changes in portfolio weights.

We have discussed the SA of σ_p estimated through GARCH models. We have found the analytical expression of the D of portfolio weights with respect to σ_p. We have examined the differences in the expressions for three possible strategies: the uniform weight change strategy – equivalent to utilizing PDs, the proportional change case – equivalent to utilizing E, and the "optimal" strategy. We have also provided the expression for the joint importance of weights with respect to σ_p by exploiting D additivity property.

Empirical results have been obtained by applying the proposed approach to the SA of the portfolio composed by the 30 stocks of the Dow Jones index – the same portfolio as in Manganelli (2004). We have analysed the importance of weights for strategies involving uniform and proportional changes in the weights, utilizing D in cases H1 [equation (3.22)] and H2 [equation (3.23)]. We then focused on the "optimal" strategy, and estimated quantitatively the degree of similarity of the three TRS by making use of Savage Score Correlation Coefficients. The corresponding strategy correlation matrix has shown that the optimal strategy resulted closer to a proportional strategy than to a uniform one. Such similarity is confirmed also by the magnitude of total change in σ_p provoked by the two strategies. The numerical findings confirm the fact that the effect of assets varies according to the adopted strategy and that conclusions obtained applying PDs or E cannot be extended to the examination of generic strategies.

We then studied the effect of diversification by examining the change in σ_p with respect to the portfolio composition. This required the computation of the sensitivity of σ_p to groups of assets [equation (3.24)]. Each group reflected the industry of operation (sector) of the firm. We have seen that Manufacturing assets resulted the most important for the uniform and optimal strategies, while "Telecommunication and ICT" assets resulted as the most influential in a proportional strategy. In the optimal strategy, most of the diversification would be due to the Manufacturing and Energy categories, since the corresponding assets are responsible for 66 percent of the change in σ_p.

Future research by the authors will involve the application of the proposed method to the SA of other portfolio properties (for example, VaR), and the examination of its role in asset allocation with dynamic portfolio optimization.

APPENDIX: PROOF OF PROPOSITION 3

The PDs of σ_t with respect to the portfolio weights are (Theorem 1 of Manganelli, 2004):

$$\frac{\partial h_t}{\partial a_i} = \frac{\partial z_t}{\partial a_i} \cdot \theta + z_t \cdot \frac{\partial \theta}{\partial a_i} \tag{3.27}$$

where $\partial z_t/\partial a_i$ is immediately derived due to the linear dependence, and $\partial\theta/\partial a_i$ denotes the ith row of the matrix:

$$\left[\frac{\P q}{\P a}\right] = -L_{aq}^{T} \times L_{qq}^{-1} \tag{3.28}$$

where $\partial\theta/\partial a = [\partial\theta_j/\partial a_i]$ $(i = 1, \ldots, n,\ \ j = 1, \ldots, m)$, $L_{a\theta} = [\partial^2 L/\partial a_i \theta_j]$ $(i = 1, \ldots, n,$ $j = 1, \ldots, m)$ and $L_{\theta\theta} = [\partial^2 L/\partial\theta_s \theta_r]$ $(s, r = 1, \ldots, m)$ and L is given by equation (3.18). Equation (3.28) is found by implicit differentiation from the solution to the set of conditions determining the parameters:

$$\left.\frac{\partial L_T}{\partial \theta_i}\right|_{\theta} = 0 \qquad i = 1, 2, \ldots, m \tag{3.29}$$

namely, $\theta = \{\theta_1, \theta_2, \ldots, \theta_m\}$, which can be regarded as an implicit function of the portfolio weights:

$$\theta = g(a^0) \tag{3.30}$$

where $\mathbf{g}$ is an m-dimensional vector of n-dimensional functions at $\mathbf{a}^0$. Combining equation (3.28) with equations (3.27) and (3.2), one finds the result.

NOTES

1. See the monography Shephard (2005) for an overview of the literature on volatility estimation.
2. By TRS we mean any change in portfolio composition.
3. After the seminal works of Bollerslev (1986) and Engle (1982), GARCH models have become widespread tools in investment management (Duan, 1995; Manganelli, 2004).
4. We consider the VSA strategy proposed by Manganelli (2004).

5. The discussion of the relationship between DIM and the Fussel–Vesely importance can be found in Borgonovo and Apostolakis (2001a), the discussion on the relationship between DIM and Elasticity can be found in Borgonovo and Peccati (2004) and Borgonovo and Peccati (2005).
6. Geometrically this would correspond to considering the direction of the change, not only its projection on the cartesian axes.
7. Some technical notes on computation. The optimal strategy in Manganelli (2004) implies the change $\mathbf{da}^* = \mathbf{a}^* - \mathbf{a}^0$, where $\mathbf{a}^0$ is the initial point. More specifically, we consider as a starting point $\mathbf{a}^0$ the result of the estimation of the weights for a given portfolio as defined by the first-order conditions of an exponentially weighted moving average (EWMA) (Manganelli, 2004). The choice of that particular stochastic process as a generator of the weights has been made as Manganelli (2004) demonstrates that the volatility of the 30 stocks estimated by an EWMA is only 7.34% lower than the one estimated by a GARCH(1,1) model in $\mathbf{a}^*$, where $\mathbf{a}^*$ is the vector of optimal weights minimizing the variance of the portfolio. This result is particularly interesting for our purposes as the EWMA optimal portfolio can be thought as a local deviation, in terms of volatility, from the GARCH(1,1) computed in the optimum. Furthermore, Following Wang's (1998) arguments, we do not impose any constraint on portfolio weights as to run the analysis under maximum attainable efficiency.

REFERENCES

Black, F. and Scholes, M. (1973) "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy*, 81(3): 637–54.

Bollerslev, T. (1986) "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, 31: 307–27.

Bollerslev, T. and Engle, R.F. (1993) "Common Persistence in Conditional Variances", *Econometrica*, 61(1): 167–86.

Borgonovo, E. (2001) "Risk-Informed Decision Making: Sensitivity Analysis and Applications", Ph.D thesis, Massachusetts Institute of Technology.

Borgonovo, E. and Apostolakis, G.E. (2001) "A New Importance Measure for Risk-Informed Decision-Making", *Reliability Engineering and System Safety*, 72(2): 193–212.

Borgonovo, E. and Apostolakis, G.E. (2001) "On Local Sensitivity Measures", *Proceedings of Sensitivity Analysis of Model Output*, Madrid, Spain.

Borgonovo, E. and Peccati, L. (2004) "Sensitivity Analysis in Investment Project Evaluation", *International Journal of Production Economics*, 90: 17–25.

Borgonovo, E. (2001) "Importance Relations and Sensitivity Analysis in Decision Theory", *Proceedings of the Tenth International Conference on Foundations and Applications of Utility, Risk and Decision Theory*, Torino, 30 May–2 June 2001, pp. 20–3.

Borgonovo, E. and Peccati, L. (2004) "The Importance of Assumptions in Investment Evaluation", *International Journal of Production Economics*, 90: 17–25.

Campbell, J.Y., Lo, A.W. and MacKinlay, A.C. (1997) *The Econometrics of Financial Markets* (Princeton, NJ: Princeton University Press).

Campolongo, F. and Saltelli, A. (1997) "Sensitivity Analysis of an Environmental Model: an Application of Different Analysis Methods", *Reliability Engineering and System Safety*, 57: 49–69.

Cheok, M.C., Parry, G.W. and Sherry, R.R. (1998) "Use of Importance Measures in Risk-Informed Regulatory Applications", *Reliability Engineering and System Safety*, 60: 213–26.

Drudi, F., Generale, A. and Majnoni, G. (1997) "Sensitivity of VaR Measures to Different Risk Models", Banca d'Italia, *Temi di discussione*, no. 317.

Duan, J.-C. (1995) "The GARCH Option Pricing Model", *Mathematical Finance*, 5: 13–32.
Duan, J.-C. and Zhang, H. (2001) "Pricing Hang Seng Index Options around the Asian Financial Crisis – A GARCH Approach", *Journal of Banking and Finance*, 25(11): 1989–2014.
Engle, R.F. (1982) "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of the United Kingdom Inflation", *Econometrica*, 50: 987–1007.
Engle, R.F. and Merzich, J. (1996) "GARCH for Groups", *RISK*, August: 36–40.
Gourieroux, C., Laurent, J.P. and Scaillet, O. (2000) "Sensitivity Analysis of Values at Risk", *Journal of Empirical Finance*, 7: 225–45.
Hull, J.C. (1999) *Options, Futures and Other Derivatives* (Upper Saddle River, NJ: Prentice-Hall).
Manganelli, S. (2004) "Asset Allocation by Variance Sensitivity Analysis", *Journal of Financial Econometrics*, 2(3): 370–89.
Manganelli, S., Ceci, V. and Vecchiato, W. (2002) "Sensitivity Analysis of Volatility: A New Tool for Risk Management", European Central Bank, Working Paper Series, Working Paper no. 194.
McNeil A.J. and Frey, R. (2000) "Estimation of Tail Related Risk Measures for Heteroskedastic Financial Time Series: An Extreme Value Approach", *Journal of Empirical Finance*, 7(3–4): 271–300
Noh, J. (1997) "Small Sample Properties of GARCH(1,1) Estimator Under Non-normality", *Economics Letters*, 55: 161–4.
Saltelli, A. (1997) "The Role of Sensitivity Analysis in the Corroboration of Models and its Link to Model Structural and Parametric Uncertainty", *Reliability Engineering and System Safety*, 57: 1–4.
Saltelli, A. (1999) "Sensitivity Analysis: Could Better Methods be Used?", *Journal of Geophysical Research*, 104: 3789–93.
Saltelli, A. (2003) "Introduction to Variance-Based Sensitivity Analysis", *JRC*, mimeo.
Saltelli, A., Tarantola, S. and Chan, K.P.S. (1999) "A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output", *Technometrics*, 41(1): 39–56.
Shephard, N. (2005) "Stochastic Volatility", in N. Shephard (ed.), *Advanced Texts in Econometrics* (Oxford: Oxford University Press).
Simon, C.P. and Blume, L. (1994) *Mathematics for Economists* (New York: W.W. Norton & Co.).
Taggart, R.A. (1996) *Quantitative Analysis for Investment Management* (Upper Saddle River, NJ: Prentice-Hall).
Turanyi, T. and Rabitz, H. (2000) "Local Methods", in A. Saltelli, K. Chan and E.M. Scott (eds), *Sensitivity Analysis* (Chichester, UK: John Wiley & Sons).
Wang, Z. (1998) "Efficiency Loss and Constraints on Portfolio Holding", *Journal of Financial Economics*, 48(3): 359–75.

CHAPTER 4

Managing Interest Rate Risk under Non-Parallel Changes: An Application of a Two-Factor Model

Manuel Moreno

4.1 INTRODUCTION

Two decades ago, fixed-income markets experienced a great increase in the volatility of assets dealt in those markets.[1] Because of this academics and market participants developed and implemented tools and techniques to manage the interest rate risk. In particular, we will consider default-free securities and liquid markets. We will distinguish two types of risk: market risk and the yield curve one, associated to parallel and non-parallel changes in the yield curve, respectively.

The classic solution in managing market risk is to use duration to immunize a certain bond portfolio. The main assumption of duration is that the yields of different securities change in a parallel way. Hence, duration can be an appropriate tool to manage market risk.

However, non-parallel movements[2] in the yield curve limit the use of duration.[3] Several duration measures associated to non-parallel movements in the yield curve have been proposed and tested in different papers, for example in Bierwag, Kaufman and Toevs (1983), Elton, Gruber and Michaely (1990), Klaffky, Ma and Nozari (1992), Ho (1992) and Reitano (1992, 1996).

These measures do not depart from a continuous-time model for interest rates but they are arbitrarily specified. Several exceptions are Ingersoll, Skelton and Weil (1978), Cox, Ingersoll and Ross (1979), Chen (1996), Munk (1999), Jeffrey (2000) and Wu (2000).

Our main goal is to define and apply duration measures based on the continuous-time model presented in Moreno (2003). Similarly to Chen (1996), we will obtain several measures of generalized duration to reflect the changes in the stochastic factors of this model. We can analyse the sensitivity of a bond portfolio to changes in the yield curve and we can compute hedging ratios to immunize a bond portfolio. Last but not least, we can solve the limitations of the conventional duration.[4]

4.2 THE MODEL

We briefly present the two-factor model for interest rates that has been proposed and analysed in Moreno (2003).

This model assumes that the price, at a certain time, of a default-free discount bond depends only on the time to maturity and two state variables: the long-term interest rate, denoted by L, and the spread (the difference between the short-term (instantaneous) riskless interest rate, r, and the long-term interest rate), denoted by s. This selection of state variables allows us to use the assumption that both variables are orthogonal.[5]

After choosing the state variables, we assume that their dynamics over time are given by the following system of stochastic differential equations:

$$\begin{aligned} ds &= \beta_1(s,L)dt + \sigma_1(s,L)dw_1 \\ dL &= \beta_2(s,L)dt + \sigma_2(s,L)dw_2 \end{aligned} \tag{4.1}$$

where t denotes calendar time, and w_1 and w_2 are Wiener processes with $E[dw_1] = E[dw_1] = 0$, $dw_1^2 = dw_1^2 = dt$ and (by the orthogonality assumption) $E[dw_1 dw_2] = 0$. β_1 and β_2 are the expected instantaneous rates of change in the state variables and σ_1^2 and σ_2^2 are the instantaneous variances of changes in these two variables.

Let $P(s, L, t, T) \equiv P(s, L, \tau)$ be the price, at time t, of a default-free discount bond that pays \$1 at maturity $T = t + \tau$. Applying Itô's lemma, setting up a hedge portfolio, and assuming no-arbitrage conditions, we obtain the partial differential equation (PDE) that the price of this bond for all maturities must satisfy:

$$\begin{aligned} &\frac{1}{2}[\sigma_1^2(.)P_{ss} + \sigma_2^2(.)P_{LL}] + [\beta_1(.) - \lambda_1(.)\sigma_1(.)]P_s \\ &\quad + [\beta_2(.) - \lambda_2(.)\sigma_2(.)]P_L + P_t - rP = 0 \end{aligned} \tag{4.2}$$

where subscripts denote partial derivatives.

The solution of this equation, subject to the terminal condition given by the payment to be received by the bondholder at maturity, allows us to price discount bonds and, thereafter, infer the term structure of interest rates. To solve this valuation equation, we must make some assumptions about the market prices of risk and the dynamics of the state variables. Since a constant market price of risk implies strong restrictions on the preferences of investors, we establish the following:

Assumption 1 The market price of each state variable risk is linear in this variable, that is,

$$\lambda_1(.) = a + bs, \;\; \lambda_2(.) = c + dL \tag{4.3}$$

Assumption 2 Each of the state variables follows a diffusion process,

$$\begin{aligned} ds &= k_1(\mu_1 - s)dt + \sigma_1 dw_1 \\ dL &= k_1(\mu_2 - L)dt + \sigma_2 dw_2 \end{aligned} \tag{4.4}$$

Under Assumptions 1 and 2, we can rewrite equation (4.2) as:

$$\begin{aligned} &\frac{1}{2}\left[\sigma_1^2 P_{ss} + \sigma_2^2 P_{LL}\right] + q_1[\hat{\mu}_1 - s]P_s + [\hat{\mu}_2 - L]P_L \\ &+P_t - (L+s)P = 0 \end{aligned} \tag{4.5}$$

subject to the terminal condition:

$$P(s, L, T, T) = 1, \quad \forall s, \; L \tag{4.6}$$

where

$$\begin{aligned} q_1 &= k_1 + b\sigma_1, \hat{\mu}_1 = \frac{k_1\mu_1 - a\sigma_1}{q_1} \\ q_2 &= k_2 + d\sigma_2, \hat{\mu}_2 = \frac{k_2\mu_2 - c\sigma_2}{q_2} \end{aligned} \tag{4.7}$$

Solving the PDE (4.5), we obtain the following proposition:

Proposition 1 The value at time t of a discount bond[6] that pays \$1 at time T, $P(s, L, t, T) \equiv P(s, L, \tau)$, is given by:

$$P(s, L, \tau) = A_1(\tau)e^{-B(\tau)s - C(\tau)L} \tag{4.8}$$

where $\tau = T - t$ and

$$A(\tau) = A_1(\tau)A_2(\tau)$$

$$
\begin{aligned}
A_1(\tau) &= \exp\left(-\frac{\sigma_1^2}{4q_1}B^2(\tau) + s^*\left(B(\tau)-\tau\right)\right) \\
A_2(\tau) &= \exp\left(-\frac{\sigma_2^2}{4q_2}C^2(\tau) + L^*\left(C(\tau)-\tau\right)\right) \\
B(\tau) &= \frac{1-e^{-q_1\tau}}{q_1} \\
C(\tau) &= \frac{1-e^{-q_2\tau}}{q_2}
\end{aligned}
\tag{4.9}
$$

with

$$
\begin{aligned}
q_1 &= k_1 + b\sigma_1,\ s^* = \hat{\mu}_1 - \frac{1}{2}\frac{\sigma_1^2}{q_1^2},\ \hat{\mu}_1 = \frac{k_1\mu_1 - a\sigma_1}{q_1} \\
q_2 &= k_2 + d\sigma_2,\ L^* = \hat{\mu}_2 - \frac{1}{2}\frac{\sigma_2^2}{q_2^2},\ \hat{\mu}_2 = \frac{k_2\mu_2 - c\sigma_2}{q_2}
\end{aligned}
\tag{4.10}
$$

4.3 GENERALIZED DURATION AND CONVEXITY

We will generalize the concepts of conventional duration and convexity using the above two-factor model. Hence, we can measure the interest rate risk with respect to both stochastic factors.

The price, at time t, of a default-free zero-coupon bond that pays \$1 at maturity, $T = t + \tau$, is given by:

$$
P(s, L, t, T) = P(t, T) = e^{-(T-t)Y(s,L,t,T)} \tag{4.11}
$$

where $Y(s, L, t, T) \equiv Y(s, L, \tau)$ is the (continuously compounded) yield to maturity of this bond.

Applying Itô's lemma, using the closed-form expression (4.8) given by Proposition 1 and the dynamics of the state variables (see (4.4)), the instantaneous change in the bond price is given by:

$$
dP(t,T) = \mu_P(.)dt - (T-t)P(t,T)\left[\frac{\partial Y(t,T)}{\partial s}\sigma_1 dw_1 + \frac{\partial lY(t,T)}{\partial L}\sigma_2 dw_2\right] \tag{4.12}
$$

with

$$
\mu_P(.) = P_s k_1(\mu_1 - s) + P_L k_2(\mu_2 - L) + P_t + \frac{1}{2}P_{ss}\sigma_1^2 + \frac{1}{2}P_{LL}\sigma_2^2 \tag{4.13}
$$

Next, we consider a coupon bond paying n coupons c_i at times t_i, $i = 1, 2, \ldots, n$. This bond has a nominal value equal to \$1 and matures at time $T = t_n$. Let $P^*(s, L, t, T) \equiv P^*(s, L, \tau)$ be the price, at time t, of this bond.

Since this bond can be interpreted as a portfolio of n discount bonds, we get:

$$P^*(t,T) = \sum_{i=1}^{n} c_i P(t,t_i) \tag{4.14}$$

From (4.12), it is verified that the instantaneous percentage change in the price of this coupon bond is given by:

$$\frac{dP^*(t,T)}{P^*(t,T)} = \frac{1}{P^*(t,T)} \sum_{i=1}^{n} c_i \mu_P(.)dt - D_s \sigma_1 dw_1 - D_L \sigma_2 dw_2 \tag{4.15}$$

where

$$\begin{aligned} D_s &= \frac{1}{P^*(t,T)} \sum_{i=1}^{n} c_i (t_i - t) P(t,t_i) \frac{\partial Y(t,t_i)}{\partial s} \\ D_L &= \frac{1}{P^*(t,T)} \sum_{i=1}^{n} c_i (t_i - t) P(t,t_i) \frac{\partial Y(t,t_i)}{\partial L} \end{aligned} \tag{4.16}$$

The values D_s and D_L represent the generalized duration measures and reflect the sensitivity of the bond yield to changes in the factors s and L. In comparison with the conventional duration, we have two duration measures, one for each factor. Moreover, there is an additional term:

$$\frac{\partial Y(t,t_i)}{\partial s}, \frac{\partial Y(t,t_i)}{\partial L}, \quad i = 1, 2, \ldots, n \tag{4.17}$$

that reflects the sensitivity of the yield to maturity to changes in each factor. Formally, we establish the following definition:

Definition (generalized duration) The "generalized durations" D_s and D_L of a bond that pays n coupons c_i at times t_i, $i = 1, 2, \ldots, n$ with respect to the factors s and L are given by the expressions:

$$\begin{aligned} D_s &= \frac{1}{P^*(t,T)} \sum_{i=1}^{n} c_i (t_i - t) P(t,t_i) \frac{\partial Y(t,t_i)}{\partial s} \\ D_L &= \frac{1}{P^*(t,T)} \sum_{i=1}^{n} c_i (t_i - t) P(t,t_i) \frac{\partial Y(t,t_i)}{\partial L} \end{aligned} \tag{4.18}$$

where $P(t, t_i)$ is the price, at time t, of a zero-coupon bond that matures at time t_i (see Proposition 1).

It is easily shown that these measures, for a zero-coupon bond, become:

$$\begin{aligned} D_s &= B(t,T) = B(\tau) \\ D_L &= C(t,T) = C(\tau) \end{aligned} \tag{4.19}$$

where we have used equations (4.8) and (4.11).

Hence, $B(\tau)$ and $C(\tau)$ indicate the sensitivity of a zero-coupon bond to changes in the spread and in the interest rates level, respectively. Then, after assessing the behavior of the bond portfolio in the presence of both changes, these measures can be an adequate tool for portfolio management. Investors who want to immunize a portfolio against these changes must equate its generalized durations to those of the asset to be replicated.

Convexity is a measure that can complement the estimation (obtained from duration) of the change in the bond price when interest rates change in a large amount. Similarly to duration, we generalize this measure:

Definition (generalized convexity) The generalized convexities δ_s and δ_L of a bond that pays n coupons c_i at times t_i, $i=1,2,\ldots,n$ with respect to the factors s and L are given by the expressions:

$$\begin{aligned} \delta_s &= \frac{1}{P^*(t,T)} \sum_{i=1}^{n} c_i (t_i - t) P(t,t_i) \frac{\partial^2 Y(t,t_i)}{\partial s^2} \\ \delta_L &= \frac{1}{P^*(t,T)} \sum_{i=1}^{n} c_i (t_i - t) P(t,t_i) \frac{\partial^2 Y(t,t_i)}{\partial L^2} \end{aligned} \tag{4.20}$$

where $P(t, t_i)$ is the price, at time t, of a zero-coupon bond that matures at time t_i (see Proposition 1).

For a zero-coupon bond, we have:

$$\begin{aligned} \delta_s &= B^2(t,T) = B^2(\tau) \\ \delta_L &= C^2(t,T) = C^2(\tau) \end{aligned} \tag{4.21}$$

4.4 HEDGING RATIOS

An alternative technique to duration for managing the interest rate risk may be performed with bond options. Since duration measures the sensitivity of a present value to changes in interest rates, it can be applied not only to bonds but it can be extended to options. Thus, it is possible to define measures of the sensitivity of different interest rate derivatives to several factors and, then, to construct the corresponding hedging strategy.

We consider a European call option on a zero-coupon bond. Let K be its strike price. If this option is exercised at expiration, T_c, the call holder pays K and receives a discount bond that matures at time $T_b > T_c$.

The price at time t, $C(s, L, t, T_c; K, T_b)$, of this option (see Moreno, 2003) is given by:

$$C(s,L,t,T_c;K,T_b) = P(t,T_b)\Phi(h+\sigma_{\tilde{P}}) - KP(t,T_c)\Phi(h) \tag{4.22}$$

where $P(t, T_i)$ is the price, at time t, of a zero-coupon bond that matures at time T_i (see Proposition 1), $\Phi(.)$ denotes the distribution function of a standard normal variable, and

$$h = \frac{\ln(P(t,T_b)) - \ln(KP(t,T_c))}{\sigma_{\tilde{P}}} - \frac{1}{2}\sigma_{\tilde{P}},$$

$$\sigma^2_{\tilde{P}} = Var(\ln(\tilde{P})),\ \tilde{P} = P(T_c, T_b) \quad (4.23)$$

Computing the derivative of (4.22), applying the chain's rule, we obtain that the generalized duration of this option to spread changes[7] is given by:

$$\frac{\partial C(.)}{\partial s} = [B(t,T_c) - B(t,T_b)]\left[P(t,T_b)\frac{f(h+\sigma_{\tilde{P}})}{\sigma_{\tilde{P}}} - KP(t,T_c)\left(\frac{f(h)}{\sigma_{\tilde{P}}} - \Phi(h)\right)\right] - B(t,T_b)C(t,T_c;K,T_b) \quad (4.24)$$

where $f(.)$ denotes the density function of a standard normal variable.

Similarly, the generalized duration of the call option to changes in the long-term interest rate is given by:

$$\frac{\partial C(.)}{\partial L} = [C(t,T_c) - C(t,T_b)]\left[P(t,T_b)\frac{f(h+\sigma_{\tilde{P}})}{\sigma_{\tilde{P}}} - KP(t,T_c)\left(\frac{f(h)}{\sigma_{\tilde{P}}} - \Phi(h)\right)\right] - C(t,T_b)C(t,T_c;K,T_b) \quad (4.25)$$

We will apply a relationship[8] that links the generalized durations of the bond option with that of its underlying asset: this relationship establishes that the generalized duration of an option is equal to the elasticity of this option times the generalized duration of the bond where the elasticity of the option is given by:

$$\frac{P(t,T_b)}{C(t,T_c;K,T_b)}\frac{\partial C(t,T_c;K,T_b)}{\partial P(t,T_b)} \quad (4.26)$$

The elasticity of the option is the product of two terms, the leverage of the option and the hedging ratio. Therefore, from equations (4.19) and (4.26), it is deduced that the delta of the option is:

$$\Delta = \frac{\partial C(.)}{\partial s}\frac{1}{B(t,T_b)}\frac{C(t,T_c;K,T_b)}{P(t,T_b)} \quad (4.27)$$

where $\partial C(.)/\partial s$ is as given in (4.24).

4.5 A PROPOSAL OF A SOLUTION FOR THE LIMITATIONS OF THE CONVENTIONAL DURATION

We consider two portfolios with the same generalized durations with respect to the above factors. Both portfolios differ in yield and convexity. We will

Table 4.1 Bonds included in portfolios 1 and 2

Bond	A	B	C	D
Coupon (%)	5.5	10	12	9
Maturity (years)	5	15	20	10
Yield (%)	5.5	10	12	9
Duration with respect to s (D_s)	0.5085	0.5499	0.6344	0.5346
Duration with respect to L (D_L)	0.8714	0.9342	1.0750	0.9123
Convexity with respect to s (δ_s)	0.3686	0.3924	0.4519	0.3826
Convexity with respect to L (δ_L)	1.1002	1.1644	1.3354	1.1439

see that, because of the equality between their generalized durations, the relative behavior of these portfolios does not depend on the magnitude or on the type of changes in the yield curve.

Portfolio 1 consists of bonds A, B and C, and portfolio 2 includes only the bond D. All these bonds have a nominal value equal to \$100 and pay coupons on a semi-year basis. Additionally, we have the features listed in Table 4.1.

We choose the bond proportions in portfolio 1 to equate the generalized durations (per a 100-basis-points change in yield) and the market values for both portfolios, and obtain the following system of equations:

$$\begin{aligned} &x_A D_s^A + x_B D_s^B + x_C D_s^C = D_s^D \\ &x_A D_L^A + x_B D_L^B + x_C D_L^C = D_L^D \\ &x_A + x_B + x_C = 1 \end{aligned} \tag{4.28}$$

where $x_j, j = A, B, C$ is the proportion invested on each bond and D_i^j, $i = s, L$, $j = A, B, C$ represents the generalized duration of the j-th bond with respect to the i-th factor. Solving this system of equations, we obtain that the proportions for the three bonds are 59.930 percent, 11.203 percent, and 28.866 percent, respectively.

We compute the generalized convexities and the yield for both portfolios. For portfolio 1, we compute the weighted average of the convexity and the yield of its bonds; the weights are the proportions of the portfolio invested on each bond. The results are as shown in Table 4.2.

As expected by design, both portfolios have the same generalized durations with respect to both factors. It can be seen that the generalized duration with respect to the short-term interest rate and the yield of portfolio 2 are greater than those of portfolio 1. The difference in yields suggests that the best strategy consists of buying portfolio 2 and selling portfolio 1. Thus, we

Table 4.2 Computations of the generalized convexities and the yield for both portfolios

	Portfolio 1	Portfolio 2
Duration with respect to s (D_s)	0.5346	0.5346
Duration with respect to L (D_L)	0.9123	0.9123
Duration with respect to r (D_r)	2.8476	3.1155
Convexity with respect to s (δ_s)	0.3848	0.3826
Convexity with respect to L (δ_L)	1.1451	1.1439
Yield (%)	7.5272	9

can obtain a gain of 147.2 basis points. On the other hand, the generalized convexities of portfolio 2 are slightly lower than those of portfolio 1. This fact suggests that portfolio 1 can provide a greater yield than portfolio 2 if there are certain shifts in the yield curve.

From now on, we will assume a shift in the yield curve instantaneously after the acquisition of these portfolios and we will analyse the relative behavior of both portfolios, measured by the difference between their yields. We will consider three possibilities: a parallel change and two types of twist in the slope of the yield curve: flattening and steepening.

For each case, at the end of the investment horizon, we obtain a certain value for both portfolios. This value is the sum of three terms: the market value of the portfolio, the coupons paid by the bonds, and the reinvestment gain generated by the coupons. Changes in interest rates have two effects on this value and, hence, on the portfolio yield: if interest rates increase, the market value of the bond decreases (price risk) but the coupons generate a higher amount of money (reinvestment risk). The opposite situation happens if interest rates fall. Therefore, the final gain will depend on the changes in interest rates and the combined effect of both types of risk.

We consider an investment horizon of six months. Therefore, we need no assumptions on the reinvestment rate of the coupons because each bond is sold just after providing the first (and last) coupon. Hence, the final (accumulated) value for each bond is its market price plus its coupon.

Table 4.3 shows this value and the yield (in annual terms) for the bonds included in the portfolio 1 when there is a parallel change in the yield curve. The first column of this table includes the change size. The coupons, in percentage terms, paid by the bonds A, B and C are 2.75, 5 and 6, respectively. We can observe that, when interest rates are increasing, each bond yield decreases. The bond C provides the highest yield because it has the longest maturity. On the contrary, the bond A, with the shortest maturity, shows the narrowest interval of yields.

Table 4.3 Relative behavior of the bonds included in the portfolio 1 with respect to a parallel change in the yield curve

Yield Change	Bond A		Bond B		Bond C	
	Accumulated value	Yield (%)	Accumulated value	Yield (%)	Accumulated value	Yield (%)
5	85.176	−29.646	75.759	−48.480	77.809	−44.381
4.5	86.757	−26.485	78.042	−43.915	79.966	−40.067
4	88.374	−23.250	80.444	−39.110	82.242	−35.514
3.5	90.029	−19.940	82.974	−34.051	84.648	−30.703
3	91.722	−16.554	85.638	−28.722	87.191	−25.616
2.5	93.455	−13.088	88.447	−23.105	89.883	−20.233
2	95.229	−9.541	91.409	−17.181	92.735	−14.529
1.5	97.044	−5.911	94.534	−10.931	95.758	−8.482
1	98.901	−2.196	97.833	−4.333	98.967	−2.064
0.5	100.803	1.606	101.317	2.635	102.376	4.752
0	102.75	5.5	105	10	106	12
−0.5	104.742	9.485	108.892	17.785	109.856	19.713
−1	106.782	13.565	113.010	26.021	113.964	27.928
−1.5	108.871	17.743	117.369	34.738	118.343	36.687
−2	111.010	22.020	121.983	43.967	123.017	46.034
−2.5	113.200	26.401	126.872	53.744	128.008	56.016
−3	115.443	30.887	132.053	64.107	133.344	66.688
−3.5	117.740	35.481	137.547	75.095	139.054	78.108
−4	120.093	40.186	143.376	86.753	145.168	90.337
−4.5	122.502	45.005	149.563	99.127	151.723	103.447
−5	124.971	49.942	156.133	112.267	158.756	117.512

Table 4.4 includes the results for both portfolios. The accumulated value of portfolio 1 is the weighted average of those of its bonds. As in Table 4.3, the first column shows the changed size in the yield curve. The last column shows the difference between both yields. A negative (positive) value means that portfolio 1 generates a higher (lower) yield than portfolio 2.

The strategy to follow with these portfolios depends on the sign of the shift in the yield curve. As the last column of this table shows, portfolio 1 outperforms portfolio 2 if interest rates increase. On the other hand, we can see that portfolio 2 provides a greater yield than portfolio 1 when interest rates fall.

This result arises because the generalized duration with respect to the short-term interest rate of portfolio 1 is lower than that of portfolio 2. Hence,

Table 4.4 Relative behavior of two portfolios with respect to a parallel change in the yield curve

Yield Change	Portfolio 1		Portfolio 2		Difference (%)
	Accumulated value	Yield (%)	Accumulated value	Yield (%)	
5	81.632	−36.734	78.661	−42.677	−5.943
4.5	83.480	−33.038	80.802	−38.395	−5.356
4	85.398	−29.202	83.030	−33.938	−4.736
3.5	87.390	−25.219	85.349	−29.301	−4.081
3	89.459	−21.081	87.762	−24.474	−3.392
2.5	91.609	−16.780	90.275	−19.448	−2.668
2	93.847	−12.305	92.892	−14.215	−1.909
1.5	96.175	−7.648	95.617	−8.764	−1.115
1	98.600	−2.798	98.457	−3.085	−0.287
0.5	101.128	2.256	101.41	2.832	0.576
0	103.763	7.527	104.5	9	1.472
−0.5	106.513	13.027	107.714	15.429	2.402
−1	109.385	18.770	111.066	22.133	3.363
−1.5	112.385	24.771	114.562	29.125	4.354
−2	115.523	31.046	118.209	36.419	5.373
−2.5	118.806	37.612	122.014	44.029	6.417
−3	122.243	44.487	125.985	51.971	7.483
−3.5	125.846	51.692	130.130	60.261	8.568
−4	129.623	59.247	134.457	68.915	9.668
−4.5	133.588	67.176	138.976	77.953	10.776
−5	137.751	75.503	143.696	87.392	11.888

an increase (decrease) in interest rates implies that the final value of portfolio 2 decreases (increases) more than that of portfolio 1 and, so, portfolio 1 performs better (worse) than portfolio 2. The difference between these generalized durations also implies that the longer the change in interest rates, the bigger the difference in the portfolio yields.

Next, we analyse the effects of two alternative changes in the slope of the yield curve. Results are shown in Tables 4.5 to 4.6 and Tables 4.7 to 4.8, respectively.

The first non-parallel change implies a decrease in the slope of the yield curve. Specifically, we assume that the change in the 5 (15) [20]-year interest rate is equal to the change in the 10-year interest rate + 1% (−0.5%) [−1%].

Table 4.5 Relative behavior of the bonds included in the portfolio 1 with respect to a decrease in the slope of the yield curve

Yield Change	Bond A		Bond B		Bond C	
	Accumulated value	Yield (%)	Accumulated value	Yield (%)	Accumulated value	Yield (%)
5	82.121	−35.757	78.042	−43.915	82.242	−35.514
4.5	83.631	−32.737	80.444	−39.110	84.648	−30.703
4	85.176	−29.646	82.974	−34.051	87.191	−25.616
3.5	86.757	−26.485	85.638	−28.722	89.883	−20.233
3	88.374	−23.250	88.447	−23.105	92.735	−14.529
2.5	90.029	−19.940	91.409	−17.181	95.758	−8.482
2	91.722	−16.554	94.534	−10.931	98.967	−2.064
1.5	93.455	−13.088	97.833	−4.333	102.376	4.752
1	95.229	−9.541	101.317	2.635	106	12
0.5	97.044	−5.911	105	10	109.856	19.713
0	98.901	−2.196	108.892	17.785	113.964	27.928
−0.5	100.803	1.606	113.010	26.021	118.343	36.687
−1	102.75	5.5	117.369	34.738	123.017	46.034
−1.5	104.742	9.485	121.983	43.967	128.008	56.016
−2	106.782	13.565	126.872	53.744	133.344	66.688
−2.5	108.871	17.743	132.053	64.107	139.054	78.108
−3	111.010	22.020	137.547	75.095	145.168	90.337
−3.5	113.200	26.401	143.376	86.753	151.723	103.447
−4	115.443	30.887	149.563	99.127	158.756	117.512
−4.5	117.740	35.481	156.133	112.267	166.308	132.616
−5	120.093	40.186	163.113	126.227	174.424	148.849

Table 4.5 shows the results for the bonds included in portfolio 1. The first column shows the change in 10-year interest rates. The conclusions are similar to those of the above change: an increase in interest rates leads to a decrease in the bond yield and the highest (lowest) yields correspond to the bond with the longest (shortest) maturity.

Table 4.6 includes the accumulated value and the yield for both portfolios. As in Table 4.5, the first column shows the changes in the bond D. Results are analogous to those obtained with a parallel change: portfolio 1 outperforms portfolio 2 if interest rates rise and vice versa if they fall. As above, the difference between the generalized durations with respect to the short-term interest rate in both portfolios implies that (a) in percentage terms, an increase (decrease) in interest rates harms (benefits) portfolio 2 (1) more, and

Table 4.6 Relative behavior of two portfolios with respect to a decrease in the slope of the yield curve

Yield Change	Portfolio 1		Portfolio 2		Difference (%)
	Accumulated value	Yield (%)	Accumulated value	Yield (%)	
5	80.957	−38.085	78.661	−42.677	−4.592
4.5	82.825	−34.349	80.802	−38.395	−4.045
4	84.766	−30.466	83.030	−33.938	−3.471
3.5	86.784	−26.430	85.349	−29.301	−2.870
3	88.884	−22.231	87.762	−24.474	−2.242
2.5	91.069	−17.860	90.275	−19.448	−1.588
2	93.346	−13.307	92.892	−14.215	−0.907
1.5	95.718	−8.562	95.617	−8.764	−0.201
1	98.193	−3.612	98.457	−3.085	0.527
0.5	100.776	1.552	101.416	2.832	1.279
0	103.473	6.946	104.5	9	2.053
−0.5	106.292	12.584	107.714	15.429	2.844
−1	109.240	18.481	111.066	22.133	3.652
−1.5	112.326	24.652	114.562	29.125	4.473
−2	115.557	31.115	118.209	36.419	5.304
−2.5	118.944	37.889	122.014	44.029	6.139
−3	122.497	44.995	125.985	51.971	6.975
−3.5	126.227	52.454	130.130	60.261	7.806
−4	130.145	60.290	134.457	68.915	8.624
−4.5	134.264	68.528	138.976	77.953	9.424
−5	138.598	77.197	143.696	87.392	10.194

(b) the additional gain for each portfolio is monotonic with the size of the change in interest rates.

The second non-parallel change reflects an increase in the slope of the yield curve. We assume that the change in the 5 (15) [20]-year interest rate is equal to the change in the 10-year interest rate −1% (+0.5%) [+1%]. Results for the bonds included in portfolio 1 are shown in Table 4.7 while Table 4.8 contains the results for both portfolios.

Table 4.7 shows similar results to those obtained in the two previous changes: an increase in interest rates decreases the bond yield and the longest bond provides the highest yields. Looking at Table 4.8, we obtain the same conclusion as in previous changes: portfolio 2 outperforms portfolio 1 if

Table 4.7 Relative behavior of the bonds included in the portfolio 1 with respect to an increase in the slope of the yield curve

Yield Change	Bond A		Bond B		Bond C	
	Accumulated value	Yield (%)	Accumulated value	Yield (%)	Accumulated value	Yield (%)
5	88.374	−23.250	73.589	−52.821	73.823	−52.353
4.5	90.029	−19.940	75.759	−48.480	75.764	−48.471
4	91.722	−16.554	78.042	−43.915	77.809	−44.381
3.5	93.455	−13.088	80.444	−39.110	79.966	−40.067
3	95.229	−9.541	82.974	−34.051	82.242	−35.514
2.5	97.044	−5.911	85.638	−28.722	84.648	−30.703
2	98.901	−2.196	88.447	−23.105	87.191	−25.616
1.5	100.803	1.606	91.409	−17.181	89.883	−20.233
1	102.75	5.5	94.534	−10.931	92.735	−14.529
0.5	104.742	9.485	97.833	−4.333	95.758	−8.482
0	106.782	13.565	101.317	2.635	98.967	−2.064
−0.5	108.871	17.743	105	10	102.376	4.752
−1	111.010	22.020	108.892	17.785	106	12
−1.5	113.200	26.401	113.010	26.021	109.856	19.713
−2	115.443	30.887	117.369	34.738	113.964	27.928
−2.5	117.740	35.481	121.983	43.967	118.343	36.687
−3	120.093	40.186	126.872	53.744	123.017	46.034
−3.5	122.502	45.005	132.053	64.107	128.008	56.016
−4	124.971	49.942	137.547	75.095	133.344	66.688
−4.5	—	—	143.376	86.753	139.054	78.108
−5	—	—	149.563	99.127	145.168	90.337

interest rates fall, and conversely if they rise. As before, this behavior is due to the differences in generalized durations with respect to the short-term interest rate. It is also verified that the additional yield of the portfolio increases with interest rates.

Hence, the analysis of these three types of changes suggests that the generalized duration measures inform appropriately about the future behavior of a portfolio when there are unexpected changes in the yield curve. This fact has important practical consequences for the management of fixed income securities: given a certain portfolio, it is possible to build a second one with the same sensitivity to changes in the spread and in the long-term rate. The relative behavior of both portfolios does not depend on the type (nor the size) of the future change in the yield curve. If we expect an increase (decrease) in

Table 4.8 Relative behavior of two portfolios with respect to an increase in the slope of the yield curve

Yield Change	Portfolio 1		Portfolio 2		Difference (%)
	Accumulated value	Yield (%)	Accumulated value	Yield (%)	
5	82.476	−35.047	78.661	−42.677	−7.630
4.5	84.312	−31.375	80.802	−38.395	−7.019
4	86.215	−27.569	83.030	−33.938	−6.369
3.5	88.188	−23.622	85.349	−29.301	−5.678
3	90.236	−19.526	87.762	−24.474	−4.947
2.5	92.363	−15.273	90.275	−19.448	−4.174
2	94.572	−10.855	92.892	−14.215	−3.359
1.5	96.868	−6.263	95.617	−8.764	−2.500
1	99.256	−1.487	98.457	−3.085	−1.598
0.5	101.741	3.483	101.416	2.832	−0.651
0	104.329	8.659	104.5	9	0.340
−0.5	107.026	14.052	107.714	15.429	1.377
−1	109.837	19.675	111.066	22.133	2.458
−1.5	112.771	25.542	114.562	29.125	3.583
−2	115.833	31.667	118.209	36.419	4.752
−2.5	119.032	38.065	122.014	44.029	5.963
−3	122.377	44.755	125.985	51.971	7.216
−3.5	125.876	51.753	130.130	60.261	8.507
−4	129.539	59.079	134.457	68.915	9.835
−4.5	—	—	138.976	77.953	—
−5	—	—	143.696	87.392	—

interest rates, we must choose the portfolio where the generalized duration with respect to the short-term interest rates is lower (higher).

4.6 CONCLUSION

Interest rate risk is associated to changes in the yield curve. We can distinguish two types of risk: market risk and yield curve one. The conventional duration is the classic solution to manage the first type of risk but it is not so clear how to manage the second type of risk.

This chapter has presented and applied new measures of generalized duration and convexity to manage this type of risk. These measures are based on a continuous-time model for interest rates. This model assumes that default free discount bond prices are determined by the time to maturity and two factors, the long-term interest rate and the spread.

The generalized duration is useful to compute hedging ratios. We have also shown a numerical example that illustrates how these new measures can mitigate the limitations of the conventional duration. Analysing different situations, it has been checked that the generalized durations do provide adequate information on the future behavior of a bond portfolio with respect to unexpected changes in the yield curve.

Hence, these measures can be a useful tool for managing fixed income portfolios. The relevant characteristics to determine the future behavior of these portfolios are (a) the generalized duration with respect to the short-term interest rate and (b) the expectations on the movements (increase or decrease) in interest rates. Thus, if two portfolios have the same generalized durations with respect to the spread and to the long-term rate, the best portfolio is the one with lower (higher) generalized duration with respect to the short-term interest rate if interest rates rise (fall).

NOTES

1. This fact is illustrated in Nelson and Schaefer (1983) and Smithson and Smith (1995).
2. Jones (1991), Litterman and Scheinkman (1991), and Knez, Litterman and Scheikman (1994) show empirical evidence of these movements.
3. See for instance, Ingersoll, Skelton and Weil (1978), Cox, Ingersoll and Ross (1979), and D'Ecclesia and Zenios (1994), among others.
4. These limitations are because conventional duration does not provide adequate information about the future performance of a bond portfolio when the yield curve changes in a non-parallel way. That is, the relative behavior of two portfolios with the same modified duration, measured by the difference in yields, depends on the size and the type of change in yields.
5. This assumption has been empirically shown in papers as Ayres and Barry (1980), Schaefer (1980) and Nelson and Schaefer (1983), among others.
6. Many other types of interest rates derivatives were priced by solving the valuation equation with the appropriate terminal condition. See Moreno (2003) for details.
7. This value is an indicative measure of the change in the call price to changes in this factor.
8. See Fabozzi (1993), chapter 15.

REFERENCES

Ayres, H.R. and Barry, J.Y. (1980) "A Theory of the U.S. Treasury Market Equilibrium", *Management Science*, 26(6): 539–69.

Bierwag, G.O., Kaufman, G.G. and Toevs, A. (1983) *Innovations in Bond Portfolio Management: Duration Analysis and Immunization* (Greenwich, CT: JAI Press).

Chen, L. (1996) *Interest Rate Dynamics, Derivatives Pricing, and Risk Management* (Berlin: Springer-Verlag).

Cox, J.C., Ingersoll, J.E. and Ross, S.A. (1979) "Duration and the Measurement of Basis Risk", *Journal of Business*, 52(1): 51–61.

D'Ecclesia, R.L. and Zenios, S.A. (1994) "Risk Factor Analysis and Portfolio Immunization in the Italian Bond Market", *Journal of Fixed Income*, 4(2): 51–8.

Elton, E.J., Gruber, M.J. and Michaely, R. (1990) "The Structure of Spot Rates and Immunization", *Journal of Finance*, 45(2): 629–42.

Fabozzi, F.J. (1993) *Fixed Income Mathematics* (Chicago: Probus Publishing Company).

Ho, T.S.Y. (1992) "Key Rate Durations: Measures of Interest Rate Risks", *Journal of Fixed Income*, 2(2): 29–44.

Ingersoll, J.E., Skelton, J. and Weil, R.L. (1978) "Duration Forty Years Later", *Journal of Financial and Quantitative Analysis*, 13(4): 627–50.

Jeffrey, A. (2000) "Duration, Convexity, and Higher Order Hedging (Revisited)", Working Paper no. 00–22, Yale School of Management.

Jones, F.J. (1991) "Yield Curve Strategies", *Journal of Fixed Income*, 1(2): 33–41.

Klaffky, T.E., Ma, Y.Y. and Nozari, A. (1992) "Managing Yield Curve Exposure: Introducing Reshaping Durations", *Journal of Fixed Income*, 2(3): 39–45.

Knez, P.J., Litterman, R. and Scheinkman, J. (1994) "Explorations Into Factor Explaining Money Market Returns", *Journal of Finance*, 49(5): 1861–82.

Litterman, R. and Scheinkman, J. (1991) "Common Factors Affecting Bond Returns", *Journal of Fixed Income*, 1(1): 54–61.

Moreno, M. (2003) "A Two-Mean Reverting-Factor Model of the Term Structure of Interest Rates", *Journal of Futures Markets*, 23(11): 1075–105.

Munk, C. (1999) "Stochastic Duration and Fast Coupon Bond Option Pricing in Multi-Factor Models", *Review of Derivatives Research*, 3(2): 157–81.

Nelson, J. and Schaefer, S.M. (1983) "The Dynamics of the Term Structure and Alternative Portfolio Immunization Strategies", in G.O. Bierwag, G.G. Kaufman and A. Toevs (1983) *Innovations in Bond Portfolio Management: Duration Analysis and Immunization* (Greenwich, CT: JAI Press).

Reitano, R.R. (1992) "Non-Parallel Yield Curve Shifts and Immunization", *Journal of Portfolio Management*, 18(3): 36–43.

Reitano, R.R. (1996) "Non-Parallel Yield Curve Shifts and Stochastic Immunization", *Journal of Portfolio Management*, 22(2): 71–8.

Schaefer, S. (1980) "Discussion", *Journal of Finance*, 35(2): 417–19.

Smithson, C.W. and Smith, Jr C.W. (1995) *Managing Financial Risk* (New York: Irwin).

Wu, X. (2000) "A New Stochastic Duration Measure Based on the Vasicek and CIR Term Structure Theories", *Journal of Business Finance and Accounting*, 27(7 & 8): 911–32.

CHAPTER 5

An Essay on Stochastic Volatility and the Yield Curve

*Raymond Théoret, Pierre Rostan and, Abdeljalil El-Moussadek**

5.1 INTRODUCTION

In this chapter we consider the issue of forecasting the stochastic volatility and the yield curve. These two concepts are very important in financial engineering, especially in risk management. Forecasting stochastic volatility is indeed an essential ingredient in VaR computations, and for immunizing bond portfolios a prediction of the yield curve is a *sine qua non*.

Volatility has many avatars. Financial theory has evolved from the concept of historical volatility to the concept of stochastic volatility. Between these concepts, the apparition of conditional volatility, introduced by Engle (1982) to forecast the volatility of inflation, among others, was perhaps an accident in financial theory. Anyway, there is a relation between conditional volatility and stochastic volatility which is confusing, even in financial theory. Nelson (1990) was the first to show that ARCH models converge weakly in distribution to continuous stochastic volatility diffusion

* The authors want to thank Mr François-Éric Racicot, PhD, Professor of Finance in the Department of Administrative Sciences at the University of Quebec (Outaouais) for his help and numerous suggestions.

processes, so that the parameters of stochastic volatility models may be estimated by ARCH processes. Consequently, we must postulate a distribution to relate conditional volatility to the stochastic one. But this may suggest that stochastic volatility is the continuous counterpart of conditional volatility, which is not the case. Conditional volatility is an observed variable while stochastic volatility is not: it is latent. This last one must be filtered.

Campbell *et al.* (1997) mentioned in their book the work of Nelson (1990) on the link between conditional volatility and stochastic volatility, but research on this subject was not developed. It revived recently through an article written by Fornari and Mele (2005) who apply the generalized error distribution to show how CEV-ARCH[1] models are approximations of volatility diffusion models in the sense that these models are Euler–Maruyama discrete time approximations of diffusion processes.

After reviewing the relation between conditional volatility and stochastic volatility, which is fundamental in risk management, we transpose these concepts to the forecasting of the term structure of interest rates.[2] Fong and Vasicek (1992) hereafter F&V, proposed a two-factor model with a mean-reverting process and a structure that makes the short-rate variance depend on the level of interest rates with a suitable restriction that the short-rate could not become negative. This model is infrequently used in practice by financial analysts because of the problem of hidden stochastic volatility which is the black box for this kind of model.

In this chapter we use the F&V model to forecast the Canadian interest-rate term structure and we apply the Extended Kalman Filter (EKF) as a tool to compute the unobserved stochastic volatility. We also introduce Bollinger bands, a well-known tool used in technical analysis, as a reduction variance technique to improve the Monte Carlo simulation performance.[3] This is a brand-new approach in the sense that we propose this variance reduction technique based on Bollinger bands to restrain the movements of volatility in a Monte Carlo simulation and, consequently, to improve its performance. Incidentally, this method has been never applied before to volatility forecasting.

The remaining of the chapter is organized as follows. In section 5.2 we discuss the concepts of stochastic volatility as opposed to conditional volatility. In section 5.3 we discuss the importance of forecasting the yield curve. In section 5.4 we present the most popular interest-rate term structure models and we also provide some details about the F&V model (1992) and about the intuition behind the EKF. In section 5.5 we explain the EKF scheme and the implementation of our specific model. The data and calibration are described in section 5.6. In section 5.7 we detail the approach used for the simulation, and empirical results are discussed in section 5.8. Finally, some interesting conclusions are offered in section 5.9.

5.2 VARIATIONS ON STOCHASTIC VOLATILITY AND CONDITIONAL VOLATILITY[4]

It is usual to model stochastic volatility, like conditional volatility, by using a product process. Let P be the price of a financial instrument and let us assume the following differential equation for the logarithm of P:

$$d(\log(P)) = \frac{dP}{P} = \mu dt + \sigma(t)\, dz_{1t} \tag{5.1}$$

where μ is the expected yield of the financial instrument; σ, the volatility of the yield; dt, a small time increment; and dz, a standard Wiener process.

Its discrete time approximation is the following product process:

$$x_t = \mu + \sigma_t U_t \tag{5.2}$$

with $x_t = \Delta\log(P_t)$ and U_t, a standard variable so that $E(U_t) = 0$ and $V(U_t) = 1$.

The conditional variance of x_t is equal to:

$$V(x_t | \sigma_t) = V(\mu + \sigma_t U_t) = \sigma_t^2$$

σ_t is consequently the conditional standard deviation of x_t, or the conditional volatility of x_t.

According to Mills (1999), a lognormal distribution is appropriate for this conditional volatility, so:

$$h_t = \log(\sigma_t^2) = \gamma_0 + \gamma_1 h_{t-1} + \xi_t \tag{5.3}$$

with $\xi_t \sim N(0, \sigma_\xi^2)$. Equation (5.2) may be rewritten as:

$$x_t = \mu + U_t e^{\frac{h_t}{2}} \tag{5.4}$$

Mills (1999) assumes that μ is nil in equation (5.4) because daily and intra-daily stocks and currencies returns have a mean equal to 0. To linearize equation (5.4), we square x_t and we take logarithms:

$$x_t^2 = U_t^2 e^{h_t}$$
$$\log(x_t^2) = \log(U_t^2) + h_t \tag{5.5}$$

We assume that $U_t \sim N(0,1)$, and we consequently know the distribution of $\log(U_t^2)$. It is a logarithmic χ^2 distribution, with mean -1.27 and variance equal to $0.5\pi^2$, or approximately 4.93. Let us note that the distribution of $\log(U_t^2)$ is very similar to the distribution of the payoffs of a short put position. Consequently, this distribution is very appropriate for taking into account left tail risk, a kind of risk associated with rare events like stock markets collapses.[5] We will use this distribution in the empirical section of this paper.

To take into account these results, let us add and subtract $\mathrm{E}\log(U_t^2)$ in equation (5.5). We have:

$$\log(x_t^2) = \mathrm{E}\log(U_t^2) + h_t + [\log(U_t^2) - \mathrm{E}\log(U_t^2)] \tag{5.6}$$

We can rewrite (5.6) to estimate it as:

$$\log(x_t^2) = 1.27 + h_t + \varsigma_t \tag{5.7}$$

with $\varsigma_t = [\log(U_t^2)] - \mathrm{E}\log(U_t^2)]$.

In equation (5.7), ς is an innovation with a logarithmic χ^2 distribution. Its expectation is:

$$E(\varsigma_t) = E[\log(U_t^2) - E\log(U_t^2)] = E\log(U_t^2) - E\log(U_t^2) = 0$$

and its variance:

$$V(\varsigma_t) = E(\varsigma_t^2) = E[\log(U_t^2) - E\log(U_t^2)]^2 = 0.5\pi^2 \approx 4.93$$

In equation (5.7), h_t is the stochastic volatility expressed in logarithmic form, i.e. $\sigma_t = \sqrt{e^{\frac{h_t}{2}}}$. It is unobserved and must be filtered. To do so, we use the extended Kalman filter in this paper. We are now in a position to compare the concept of stochastic volatility with the concept of conditional volatility of the ARCH models. In a GARCH (1,1) model, the conditional volatility ν_t may be expressed as:

$$y_t = c + \varepsilon_t \nu_t$$
$$\nu_t^2 = \beta_0 + \beta_1 \nu_{t-1}^2 + \beta_2 \varepsilon_{t-1}^2 \tag{5.8}$$

with y_t a price variable and c, a constant. If we examine equation (5.8), we note that the conditional volatility is observed at time t because it is conditional on observations made at time $(t-1)$. But according to equation (5.3), stochastic volatility is unobserved at time t because its equation contains an innovation, contrarily to conditional volatility whose equation has no innovation. This is the major difference between stochastic volatility and conditional volatility.

Taylor (1994) reported that there is a mixing variable M_t which relates stochastic volatility to conditional volatility. This variable is distributed like an inverse gamma. We can write:

$$\sigma_t = (M_t \nu_t^2)^{\frac{1}{2}} \tag{5.9}$$

with σ_t, the stochastic volatility and ν_t, the conditional volatility. It is consequently a random variable which links stochastic volatility to conditional volatility. This random variable M_t represents the uncertainty associated to the news revealed at time t. That is why σ_t is unobserved, contrarily to ν_t which is observed at time t.

Perhaps the view initiated by Nelson (1990) that ARCH models are approximations of diffusion models created confusion in the financial literature about the relation between stochastic and conditional volatility. Let us see how this approximation works. Assume the following GARCH (1,1) process:

$$\nu_{n+1}^2 = w + \beta\nu_n^2 + \alpha\varepsilon_n^2 \tag{5.10}$$

with $\varepsilon = \mu\nu$, $\mu \sim N(0,1)$. We can write (5.10) as:

$$\nu_{n+1}^2 - \nu_n^2 = w - (1 - \alpha E(\mu^2) - \beta)\,\nu_n^2 + \alpha\nu_n^2(\mu^2 - E(\mu^2)) \tag{5.11}$$

This equation converges weakly in distribution to:

$$d\nu(\tau)^2 = (\omega - \varphi\nu(\tau)^2)\,d\tau + \psi\nu(\tau)^2\,dz(\tau) \tag{5.12}$$

with z_t a Wiener process. The convergence between the parameters of the discrete equation (5.10) to the parameters of the continuous equation (5.12) is the following:

$$\begin{aligned} &\lim h^{-1}w_h = \omega \\ &\lim h^{-1}(1 - \alpha_h - \beta_h) = \varphi \\ &\lim h^{-\frac{1}{2}}\sqrt{2}\alpha_h = \psi \end{aligned} \tag{5.13}$$

with h a very small time increment expressed in fraction of year. The $\sqrt{2}$ term in the equation of ψ is explained by the fact that $\xi = \mu^2 - E(u^2)$ is a chi-square variable with one degree of freedom and a variance of 2. The sequence ξ_n is consequently a sequence of chi-square variables and is the discrete time approximation of the Brownian increments dz.[6]

Fornari and Mele (2005) have generalized Nelson's (1990) model to the class of CEV-ARCH models. They show how volatility diffusions may be approximated by these models by assuming a distribution for the innovation which encompasses the normal distribution: the generalized error distribution (ged). Let us assume the following CEV-ARCH model for the interest rate r in its continuous form:

$$\begin{aligned} dr(\tau) &= (\iota - \theta r\tau)\,d\tau + \nu(\tau)\sqrt{r(\tau)}\,dz_1 \\ d\nu(\tau)^\delta &= (\omega - \varphi\nu(\tau)^\delta)d\tau + \psi\nu(\tau)^{\delta\eta}\left(\rho\,dz_1 + \sqrt{1-\rho^2}\,dz_2\right) \end{aligned} \tag{5.14}$$

with dz_1 and dz_2, two Wiener processes with a correlation of ρ, the correlation being performed by a Cholesky decomposition. Equation (5.14) is the continuous representation of the CEV-ARCH model. There are three additional parameters to estimate in comparison with standard ARCH models: δ, η and

ρ; η is the coefficient of elasticity. The GARCH(1,1) model is a special case of equation (5.14) for which $\delta = 2$, $\eta = 1$ and $\rho = 0$.

Let us consider the following discrete time approximation of equation (5.14):

$$\nu_{n+1}^{\delta} = w + \alpha \nu_n^{\delta\eta} \, |\mu_n|^{\delta\eta} + \beta \nu_n^{\delta} + \alpha E(|\mu|^{\delta\eta})(\nu_n^{\delta} - \nu_n^{\eta\delta}) \tag{5.15}$$

This process converges weakly in distribution to:

$$d\sigma(\tau)^{\delta} = (\omega - \varphi\sigma(\tau)^{\delta})\, d\tau + \psi\sigma(\tau)^{\delta\eta}\, dz_2 \tag{5.16}$$

Fornari and Mele (2005) assume that μ obeys to the ged distribution. Let us write:

$$n_{\delta,\upsilon} = \frac{2^{\frac{\delta}{\upsilon}-1}\nabla_{\upsilon}^{\delta}\Gamma\left(\frac{\delta+1}{\upsilon}\right)}{\Gamma(\upsilon^{-1})}$$

with $\nabla_{\upsilon}^{\delta} = \frac{\Gamma(\upsilon^{-1})}{2^{\frac{\delta}{\upsilon}}\Gamma(3\upsilon^{-1})}$ and Γ, the gamma function. When the distribution is assumed normal, the coefficient of the gamma function υ is equal to 2 and consequently: $\nabla_2^2 = \frac{\Gamma\left(\frac{1}{2}\right)}{2\Gamma\left(\frac{3}{2}\right)} = 1$. At the limit, we have the following relations between the coefficients of equations (5.15) and (5.16):

$$\varphi = 1 - n_{\delta,\upsilon}\left[(1-\gamma)^{\delta} + (1+\gamma)^{\delta}\right]\alpha - \beta$$

$$\psi = \sqrt{\left(m_{\delta,\upsilon} - n_{\delta,\upsilon}^2\right)\left((1-\gamma)^{2\delta} + (1+\gamma)^{2\delta}\right) - 2n_{\delta,\upsilon}^2\,(1-\gamma)^{\delta}\,(1+\gamma)^{\delta}}$$

with $m_{\delta,\upsilon} = \frac{2^{\frac{2\delta}{\upsilon}-1}\nabla_{\upsilon}^{2\delta}\Gamma(\frac{2\delta+1}{\upsilon})}{\Gamma(\upsilon^{-1})}$

According to equations (5.15) and (5.16), the CEV-ARCH model converges weakly in distribution to a continuous diffusion model. But this link is only an approximation. We must assume a distribution to prove this convergence and the choice of this distribution has a great impact on the relation between the coefficients of an ARCH model and the corresponding diffusion model. Anyway, for simulating stochastic volatility via the Monte Carlo method, we must have estimates of the parameters of the diffusion process governing stochastic volatility; and, as we saw, ARCH models are a way to compute them.

It is relevant to repeat that stochastic volatility is forwards-looking while conditional volatility as estimated by an ARCH model is backwards-looking. Stochastic volatility is based on the uncertainty of the news revealed to markets at time t as evidenced by equation (5.9): it anticipates this information. On its side, conditional volatility as computed by an ARCH model is based on observed information incorporated in market prices. Conditional volatility consequently has an advantage on stochastic volatility in the sense that the first is observed and the second is latent. But stochastic volatility is susceptible to deliver more information to the risk manager about the market pulse. Consequently, it must be estimated and compared with more traditional measures of volatility as we do in this essay for the case of interest rates.

5.3 INTEREST RATE TERM STRUCTURE FORECASTING

Forecasting the term structure is of great interest because it is considered as a leading indicator of economic activity. Some findings suggest that the spread between long-term and short-term interest rates has proven to be an excellent predictor of changes in economic activity. As a general rule, when long-term interest rates have been much above short-term rates, strong increases in output have followed within about one year; however, whenever the yield curve has been inverted for any extended period of time, a recession has followed.[7] Day and Lange (1997) have shown that the slope of the nominal term structure from 1- to 5-year maturities is a reasonably good predictor of future changes in inflation over these horizons.

5.4 INTEREST RATE TERM STRUCTURE MODELS

The recent literature has produced major advances in theoretical models of the term structure. Term structure models include no-arbitrage and equilibrium models. The no-arbitrage tradition focuses on perfectly fitting the term structure at a point in time to ensure that no arbitrage possibilities exist, which is essential for pricing derivatives. The equilibrium tradition focuses on modeling the dynamics of the instantaneous rate, typically using affine models that provide clear economic intuitions connecting the term structure with economic fundamentals. Equilibrium and arbitrage models have similar structures. The difference comes from the nature of the input used to calibrate the model parameters. The equilibrium models explicitly specify the market price of risk; the model parameters, assumed to be time-invariant, are estimated statistically from historical data. These models are often used by economists to understand the relationship between the shape of the term structure and its forecast for future economic conditions. Traders, however, would rather use arbitrage models because they are calibrated to match the model price of the underlying security with its market price but also because they circumvent the difficult estimation of the market price of risk. Another classification of term structure models can be made according to the number of factors involved. One should make a distinction between one-factor and multifactor models. One-factor models are popular because of their simplicity. Empirical evidence on principal component analysis has shown that almost 90 percent of the variation in the changes of the yield

Table 5.1 Interest-rate term structure models

Equilibrium models	**No-arbitrage models**
Vasicek (1977)*	Ho-Lee
Cox-Ingersoll Ross (CIR 85)*	Hull-White (90)
Brennan-Schwartz (79)**	Black-Derman-Toy (BDT-90)
Fong-Vasicek (92)**	Heath-Jarrow Morton (HJM 92)
Longstaff-Schwartz (92)**	

*One-factor models; **Two factor models.

curve is attributable to the variation in the first factor which is considered to be the level of the interest rate.[8] Because the first factor relates to the interest rate level, any point on the yield curve may be used as a proxy for it. For most one-factor models, the factor is generally taken to be the instantaneous short rate, $r(t)$. On the other hand, the multifactor models postulate that the evolution of the interest-rate term structure is driven by the dynamics of several factors and therefore, the yields are functions of these factors. These factors can be represented by macroeconomics shocks or be related to the level, slope and curvature of the yield curve itself. Table 5.1 outlines the most popular interest-rate term structure models.

Interest rate forecasting is crucial for bond portfolio management and for predicting the future changes in economic activity. The arbitrage-free term structure literature has little to say about dynamics or forecasting, as it is concerned primarily with fitting the term structure at a point in time. The affine equilibrium term structure literature is concerned with dynamics driven by the short rate, and so is potentially linked to forecasting.

Since the main aim of this chapter is to forecast the Canadian interest-rate term structure, we choose a two-factor model that belongs to equilibrium models.

5.4.1 The Fong–Vasicek model (1992)

Empirical studies have revealed that the volatility of the changes in the short rate is time-varying and stochastic. To explicitly model the stochastic changes in the interest rate volatility and their effect on bond prices and option values, Fong and Vasicek (1992) proposed a two-factor extension of the Vasicek model in which the Ornstein–Uhlenbeck process is modified to include a stochastic variance that follows a square-root process:

$$dr_t = k(\mu - r_t)\,dt + \sqrt{v_t}\,dW_t$$
$$dv_t = \lambda(\nu - v_t)\,dt + \tau\sqrt{v_t}\,dW_s$$

with $E(dW_t, dW_s) = \rho$ and where W_t and W_s are two correlated Brownian motions under the risk-neutral distribution. We are positioning ourselves in a risk-neutral world where investors require no compensation for risk and the expected return of securities is the risk-free interest rate.

As we can see, this model allows for a stationary mean reverting process whose variance is again a stationary stochastic process. Here μ is the unconditional expectation of the short rate process and k controls the degree of persistence in interest rates in the sense that it measures the speed with which the interest rate returns to its mean. In order to interpret the other parameters, let us observe that the second equation of the model is just a square root process for variance v_t. Now we can interpret parameter v as the unconditional average variance. The parameter λ accounts for the degree of persistence in the variance. Finally, the parameter τ is the unconditional infinitesimal variance of the unobserved variance process. The hidden volatility is an obvious weakness of the model and makes it hard to use. In this case using the technique of filtering is very natural to infer the values of the unobserved volatility process.

5.4.2 The extended Kalman filter (EKF)

To use the F&V model we have to deal with the unobserved volatility process. To estimate it, we apply the Kalman filter.[9] This filter is a widely used methodology which recursively calculates optimal estimates of unobservable state variables, given all the information available up to some moment in time. Estimates are improved as new data arrive. The application of Kalman filtering methods in the estimation of term structure models using cross-sectional/time series data has been investigated by Pennacchi (1991), Lund (1994, 1997), Chen and Scott (1995), Duan and Simonato (1995), Geyer and Pichler (1996), Ball and Torous (1996), Jegadeesh, and Nowman (1999), Babbs, De Jong and Santa-Clara (1999), De Jong (2000), Dewachter and Maes (2001) and Sørensen (2002). James and Webber (2000, chap. 18) gives and extensive survey of the EKF method and its use for estimating term structure models while Hamilton (1994, chap. 13) does a rigorous presentation of the Kalman filter and its extended version from which we borrowed.

The use of the state space formulation of term structure models and the application of the Kalman filter have the advantage to allow the underlying state variables to be handled correctly as unobservable variables compared to using a short-term rate historical volatility as a proxy.

5.5 METHODOLOGY

In this section, we provide a brief overview of the EKF method followed by its application to the F&V model.

5.5.1 The extended Kalman filter (EKF)

The Kalman filter uses data observed in the market to infer values for unobserved state variables. The idea is to express a dynamic system in a particular form called the *state-space representation*. A state-space model is characterized by a measurement equation and a transition one. Once this has been made, a three-step iteration process can begin. There is one iteration for each observation date t, and one iteration includes three steps, as is shown in Figure 5.1.

During the first step called the prediction phase, the values of non-observable variables in $(t-1)$ are used to compute their expected value in t, conditionally to the information available in $(t-1)$. The predictions rely on the transition equation. The predicted values $\tilde{\alpha}_{t/t-1}$ are then introduced in the measurement equation to determine the measure $\tilde{y}_t$. In this equation, the errors have zero mean and are not serially nor temporarily correlated. They represent every kind of disturbances likely to lead to errors in the data. The second step or innovation phase allows for the computation of the innovation ν_t. Lastly, non-observable variables values, which where computed in the prediction phase, are updated conditionally to the information given by ν_t. Once this calculation has been made, $\tilde{\alpha}_t$ is used to begin a new iteration.

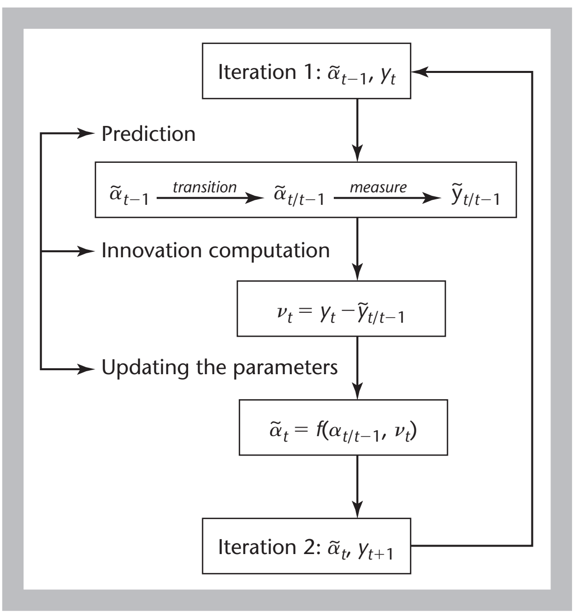


Figure 5.1 The three steps of an iteration

5.5.2 Algorithm

A standard setup of the Kalman filter is applicable to the linear state-space model of the form:

$$y_n = Z_n\alpha_n + d_n + \varepsilon_n$$
$$\alpha_n = T\alpha_{n-1} + c_n + R_n\eta_n$$

with $\text{var}(\varepsilon_n) = H_n$ and $\text{var}(\eta_n) = Q_n$. The first equation is the measurement equation and the second equation is the transition equation. (ε_n) and (η_n) are independent normal random variables with zero mean.

The Kalman filter for this approximate state-space model is then given by:

$$\begin{aligned} a_{n/n-1} &= T_n(a_{n-1}), \\ P_{n/n-1} &= \hat{T}_n P_{n-1}\hat{T}_n' + \hat{R}_n Q_n \hat{R}_n', \\ F_n &= \hat{Z}_n P_{n/n-1}\hat{Z}_n' + H_n, \\ a_n &= a_{n/n-1} + P_{n/n-1}\hat{Z}_n' F_n^{-1}(y_n - Z_n(a_{n/n-1})), \\ P_n &= P_{n/n-1} - P_{n/n-1}\hat{Z}_n' F_n^{-1}\hat{Z}_n P_{n/n-1} \end{aligned}$$

To apply this algorithm, one should proceed to a discretization and a linearization of the F&V short rate model which is discussed in next section.

5.5.3 Applying the Kalman filter to the F&V model

Recall that Fong and Vasicek model the short rate and its stochastic variance with the following equations:

$$dr_t = k(\mu - r_t)\,dt + \sqrt{v_t}\,dW_t$$
$$dv_t = \lambda(\nu - v_t)\,dt + \tau\sqrt{v_t}\,dW_s$$

As we can see, the model respects the Kalman filter state-space form. One can consider the first and the second equation as the measurement equation and the transition equation respectively. But the model is still in its continuous and non linear form. Before applying directly the algorithm of extended Kalman filter, we try to put these two equations in their discrete and linear form.[10]

Discretization

An application of Ito formula to the first equation of the F&V model yields:

$$de^{kt}(r_t - \mu) = e^{kt}\sqrt{v_t}\,dW_t$$

After integrating by parts, we obtain the discrete time specification of the F&V model:

$$\begin{aligned} r_{t+h} &= \mu + e^{-kh}(r_t - \mu) + \varepsilon_t(h)_t \\ \upsilon_{t+h} &= \nu + e^{-\lambda h}(\upsilon_t - v) + \eta_t(h)_t \end{aligned} \quad t = 0, h, 2h, \ldots,$$

Where h denotes the sampling interval expressed in year (for example, on quarterly frequency $h = 3/12$), and where $\varepsilon_t(h)$ and $\eta_t(h)$ are innovations.

Linearization

To get the linear form of the F&V model, we apply the procedure of linearization, proposed by the EKF and explained above, to the discrete form of F&V model. We consider the observation y_n to be $\ln(R_n^2/h)$:

$$y_n = \ln V_n + \ln \varepsilon_n^2$$

Clearly $\ln \varepsilon_n^2$ is not Gaussian, but has the distribution of $\ln \chi_1^2$. To use EKF, we replace this by a normal random variable with mean −1.270363 and variance 4.934802, the mean and variance, respectively, of a $\ln \chi_1^2$ random variable as explained in section 5.2. We then apply the EKF methodology with:

$$Z_n(x) = \ln x - 1.270363; \quad H_n = 4.934802;$$

$$T_n(x) = e^{-\lambda h}x + (1 - e^{-\lambda h})v; \quad R_n(x) = \tau e^{-\lambda h}\sqrt{h}\sqrt{x}; \quad Q_n = 1$$

Using the usual Taylor expansion to perform the linearization, we finally get the discrete and linear form of the F&V state-space model:

$$y_n = \frac{1}{a_{n/n-1}}\alpha_n + \ln(a_{n/n-1}) - 2.2703 + \varepsilon_n$$

$$\alpha_n = e^{-\lambda k}\alpha_{n-1} + (1 - e^{\lambda h})\nu + \tau e^{-\lambda h}\sqrt{h}\sqrt{a_{n-1}}\eta_n$$

The last step, before applying directly the Kalman filter to infer the values of unobserved volatilities, and proceed to the forecasting of the interest rate term structure, is to estimate the values of the parameters of the model. The next section gives a data description and explains the methodology adopted for calibration.

5.6 DATA AND CALIBRATION OF THE FONG AND VASICEK MODEL[11]

5.6.1 Data

We use Treasury bill yields provided by the Bank of Canada for the 1-, 3-, 6-month, 1-year maturity and the Canadian government yield curve

Table 5.2 Example of ML parameter estimates obtained for the 10 Dec. 2002

Parameter	F&V (1992)
k	0.176272
μ	0.055796
λ	3.860698
ν	0.00135
τ	0.003576
ρ	0.611125816

provided by Bloomberg for the 2-, 5-, 10-, 20-, 30-year maturities. We consider each yield of the government bond term structure as the "short interest rate r" of the Fong and Vasicek model as well as its corresponding variance v. The span of data goes from 23 October 2002 to 23 October 2003 (250 days).

5.6.2 Calibration of the model

The inputs are: (1) the daily Canadian government yield curve obtained from the Bank of Canada and from Bloomberg; and (2) the daily Canadian government yield variances term structure computed from GARCH(1,1) model[12] applied to historical data (400 past daily observations of the interest rate) adjusted for the interest rate using the Campbell, Lo and Mc Kinlay methodology (1997). The yield and variance curves have been smoothed by 3rd degree polynomial functions to generate 3,000 data for each curve. The outputs of the calibration, k, μ, λ, ν and τ from the F&V model, and the correlation ρ[13] between the two factors are obtained by Full Information Maximum Likelihood Marquardt (ML) (see Table 5.2) or Three-Stage Least-Squares 3SLS method when the ML did not converge. The 3SLS is an appropriate technique when right-hand-side variables are correlated with the error terms, and there is both heteroskedasticity, and contemporaneous correlation in the residuals.

5.7 SIMULATION

5.7.1 Evolved approach

The simulation approach adopted in this chapter is based on the Monte Carlo simulation of every yield of the Canadian yield curve. We simulate a 1,000 trajectories for each yield; r_0 and v_0, the initial values of the simulation, are respectively the yield observed at day 1 and its annualized variance obtained from EKF.

5.7.2 Improvement of the Monte Carlo simulation

To improve the performance of Monte Carlo simulation, we use in conjunction two variance reduction techniques: the classical antithetic variable and Bollinger bands,[14] a technique borrowed to the technical analysis. Bollinger bands become narrower during less volatile periods and wider during more volatile periods. Variance reduction with Bollinger bands is obtained by forcing the simulated rate to remain inside predetermined upper and lower bands during the simulation.

Remembering that Bollinger bands are bands usually drawn at ±2 standard deviations off the value of the 20-day moving average of the times series under study, have the standard deviation used to compute Bollinger bands is the conditional standard deviation obtained from the extended Kalman filter for each simulated yield. Moreover, instead of using the 20-day moving average of the yield that we are simulating as the central value of the bands, we use the value of the expected 3-month CDOR (Canadian Dollar Offer Rate) in 20 days[15] *minus* the 20-day historical average spread of the simulated yield over the 3-month CDOR spot. Our assumptions are that (1) the spreads between the CDOR rate and the other yields of the term structure over 20 days remain constant; and (2) the implied future CDOR rates obtained from the BAX futures contract prices are a good proxy of what will be the level of the CDOR rates in 20 days.

More precisely, we test the performance of the Monte Carlo simulation with Bollinger bands drawn at ±1 or ±2 standard deviations off the spread.

Each yield that has been computed with the Monte Carlo simulation is the forecasted yield in 20 days. Repeating the methodology for each component of the curve, we forecast the Canadian Government yield curve in 20 days.

5.8 EMPIRICAL RESULTS

The extended Kalman filter allows us to obtain an estimation of the stochastic volatility for each maturity and for each day in our sample. In other words, the output of the filter is a vector of 250 variance term structures (one term structure per day). Figure 5.2 illustrates the variance term structure for the first day of our sample (October, 23 2002).

The variance term structure provided by the extended Kalman filter has the same shape of what we can empirically observe. We observe that the variance of the long-term interest rate is lower than the variance of the short-term interest rate which gives a downward-sloping curve.

We generate 1,000 trajectories for each yield. The number of time steps x is computed with the following equation: $\frac{30 \text{ years}}{3{,}000} = \frac{\frac{20}{250}}{x}$ given $x = 8$ time

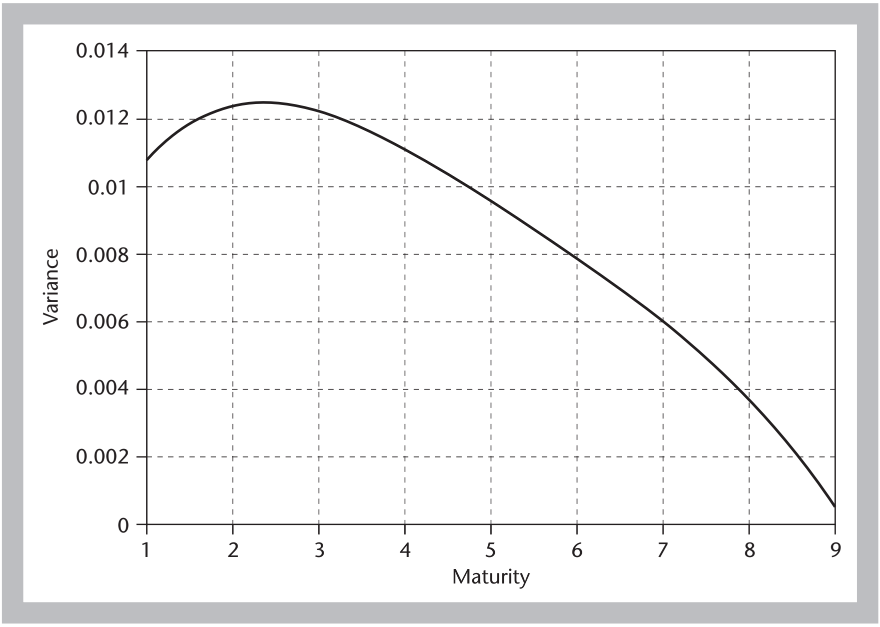


Figure 5.2 Variance term structure obtained from EKF for 23 October 2002

steps for 20 days of simulation, $dt = 20/250 = 0.08$, 30 years, the maximum maturity of the yield curve and 3,000 the number of observations used to calibrate the F&V model.

5.8.1 Quality of fit

The simulation procedure yields 250 interest rate term structures. Figure 5.3 shows the forecasted term structures obtained by various methods of simulation on 23 October 2002.

The EKF method gives the best fit to the observed interest rate term structure followed closely by the evolved method with ± 1 sigma.[16] In addition, we observe that the quality of forecast decreases as the Bollinger bands become larger. In the following section, we measure more precisely the simulation performance by computing the error of estimation.

5.8.2 Root Mean Square Error (RMSE)

On one hand, we obtain from the simulation 250 forecasted interest rate term structures. On the other hand, we observe 250 realized interest rate term structures, we are thus able to measure the performance of the simulation

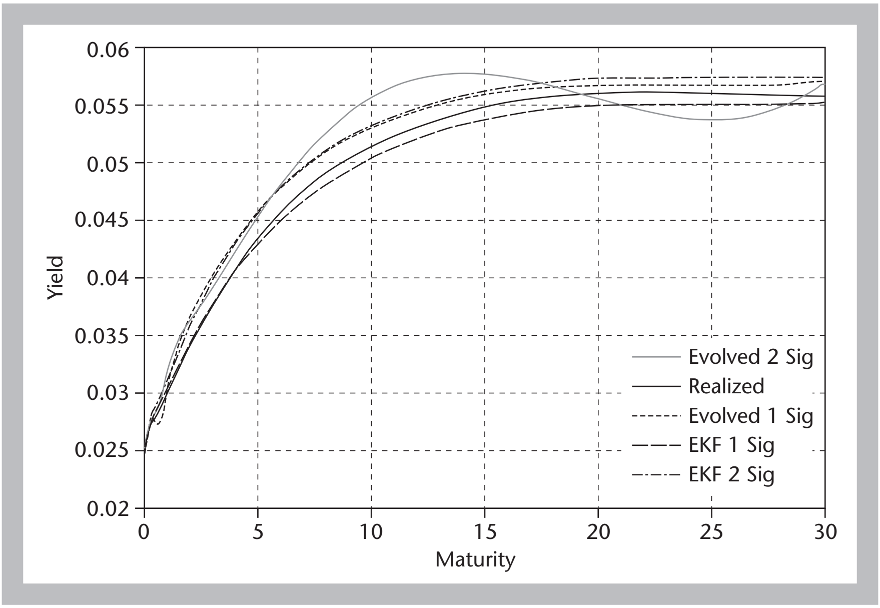


Figure 5.3 Interest-rate term structures forecasted versus realized on 23 October 2002

by computing the RMSE:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\text{Forecasted yield} - \text{Realized yield})^2}$$

Figure 5.4 shows the term structures of the RMSE obtained from different methods. We observe that the EKF approach with Bollinger bands and antithetic variable with ±1 sigma performs best. Followed by the same approach except that the volatility has been estimated by GARCH (1,1). The evolved approach with ±2 sigma whatever the method of volatility estimation (EKF or GARCH) performs better than the naïve approach.[17]

In addition, we observe that the RMSE term structures are downward-sloping. One of the possible reasons is that this method probably does not use all the information about the factor values contained in the cross-sectional dimension.

In order to measure the exact contribution of introducing the Bollinger bands to the F&V and EKF models, we perform the *F*-ratio test on the RMSE.

H_0 is rejected in both tests since $F_{stat} > F_{(241,240,.05)}$. We conclude that the differences in RMSE are significant for the two levels of Sigma (One-sigma and Two-sigma) used in the Bollinger bands technique compare to F&V coupled to EKF *without* Bollinger bands. This result suggests that associating

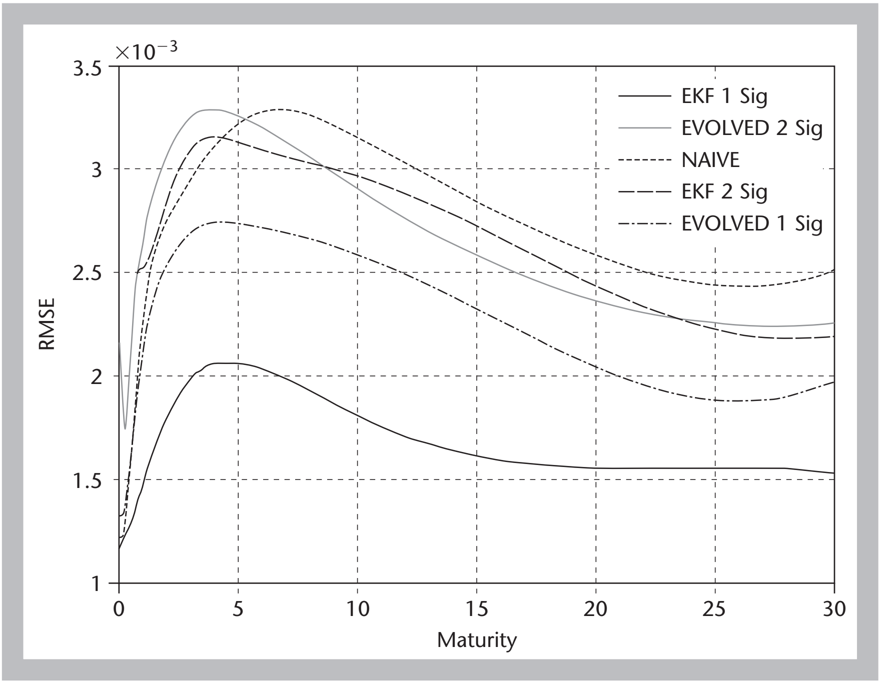


Figure 5.4 RMSE for different maturities of the forecasted interest-rate term structures versus realized on the whole sample (252 days)

Bollinger bands to the F&V model and EKF increases the performance of the Monte Carlo simulation in term of reducing the estimation error. Moreover, the RMSE decreases as we make the interval of Bollinger bands narrower (from ±2 sigma to ±1 sigma).

5.9 CONCLUSION

We have proposed a method of forecasting the interest-rate term structure. This method is based on applying the EKF to the F&V model (1992). We found that the estimation of the unobservable component approach by EKF improved significantly the 20-day forecast of the yield curve.

Furthermore, we observed a drastic improvement of the RMSE by using the Extended Kalman Filter instead of the GARCH(1,1) method when the two methods are separately applied to the F&V model. We conclude of the superiority of the EKF method over the GARCH(1,1) method to estimate the volatility.

In addition, the test of equality of mean applied to the RMSEs provided by the F&V model coupled to EKF and provided by the addition of the Bollinger bands technique suggests that the Bollinger bands technique significantly improves the Monte Carlo simulation when it is applied to the F&V model.

Table 5.3 F&V + EKF are the RMSE computed with the plain F&V model coupled to EKF

	Root Mean Square Error		
Time to maturity in years	**F&V + EKF Boll 2-Sigma**	**F&V + EKF Boll 1-Sigma**	**F&V + EKF**
0,0833	0.0021976	0.0012302	0.0017157
0,25	0.0022187	0.0014267	0.0017856
0,5	0.0023247	0.0017646	0.0023786
1	0.00244	0.0019703	0.0041471
2	0.0025822	0.0020256	0.0063822
5	0.0026933	0.0016508	0.0091518
10	0.0029729	0.001552	0.0066138
20	0.0025283	0.0015518	0.0040684
30	0.0033794	0.0015318	0.0038321
Total RMSE	**0.0233371**	**0.0147038**	**0.0400753**
F-ratio test (F_{stat})	***1.717235646***	***2.725506332***	$F_{(241,240,.05)} \approx$ ***1.24***

F&V + EKF Boll 2-Sig and **F&V + EKF Boll 1-Sig** are the RMSE obtained with F&V + EKF model improved by the reduction technique which is the Bollinger bands with two levels of sigma. We perform two different F-ratio tests.
Test 1: H_0: $RMSE_{0\sigma} = RMSE_{1\sigma}$ versus H_1: $RMSE_{0\sigma} \neq RMSE_{1\sigma}$
Test 2: H_0: $RMSE_{0\sigma} = RMSE_{2\sigma}$ versus H_1: $RMSE_{0\sigma} \neq RMSE_{2\sigma}$
With: $\alpha = .05$, $df_1 = 241$, $df_2 = 240$.

However, we have seen that one of the fundamental hypotheses of the EKF is that the errors should be Gaussian which is not the case in our model. As indicated by De Jong (2000), the EKF in this situation leads to inconsistent estimation of parameters, though without high bias.

Therefore, we can suggest for future research that other filtering techniques suitable for nonlinear models with non-Gaussian errors, like ged innovations, are necessary. Nevertheless, the use of Kitagawa method (1987) removes the inconsistency problem (De Jong, 2000) that comes with the use of Kalman filter. This technique has been used by Danilov and Mandal (2000) to estimate stochastic volatility in two-factor short rate models.

Another way of research would be to compare our method to other methods of forecasting the interest-rate term structure such as Diebold and Li (2003).

NOTES

1. For example, constant elasticity of variance (CEV) ARCH models.
2. Previous studies on this subject are: Litterman and Scheinkman (1991), Chen and Scott (1993), Dai and Singleton (2000), and De Jong (2000).

3. This technique of variance reduction was first introduced by Théoret and Rostan (2002a, 2002b).
4. In writing this section we have used Mills (1999), Taylor (1994) and Fornari and Mele (2005). For a recent review on stochastic volatility see Andersen, T.G., Bollerslev, T., Christoffersen, P.F. and Diebold, F. (2006). See also Racicot and Théoret (2005).
5. Incidentally, many hedge funds have a return distribution which is similar to the distribution of the payoffs of a short put position.
6. This intuition linking the Brownian motion increment to its discrete counterpart is due to Fornari and Mele (2005).
7. An explanation of this result is that the term spread reflects both current monetary conditions, which affect short-term interest rates, and expected real returns on investment and expectations of inflation, which are the main determinants of long-term rates. For more details see Clinton (1995).
8. See Champman and Pearson (2001) for a detailed discussion.
9. To estimate unobserved state variables and nonlinearities, we can also use the Markov–Chain Monte Carlo. See Eraker (2001).
10. For more details on linearization and discretization of interest rate models, Jarrow (1996) is a very good reference from which we have borrowed. See also James and Webber (2000) and Gouriéroux and Monfort (1996).
11. The empirical work was performed on EViews and Matlab softwares.
12. The variance obtained from GARCH(1,1) is used only for calibration purposes. For forecasting purposes, we used the variance provided by the extended Kalman filter.
13. We impose the correlation between the two random variables during the simulation by applying the Cholesky decomposition.
14. This method is detailed in Théoret and Rostan (2002a.).
15. The expected 3-month CDOR (Canadian Dollar Offer Rate) in 20 days is obtained from the BAX futures price traded on the Montreal Exchange (MX), using a linear interpolation of the BAX futures price. Our assumption is that the CDOR rate will vary linearly overtime. In Canada, the 3-month CDOR rate is the 3-month bankers' acceptance rate. It is used as the floating leg rate to price plain-vanilla swap contracts. It represents the main benchmark of the Canadian money market.
16. The results of the EKF method have been compared to the results obtained from the evolved approach. In the latter, the simulation is performed in the same conditions as the EKF approach except for using GARCH(1,1) as a volatility estimation method instead of EKF.
17. The naïve approach consists on computing the spreads between the 3-month CDOR over the yields composing the term structure. These spreads are assumed to be constant in the next 20 days. Only the reference 3-month CDOR will be simulated overtime to obtain the forecasted interest-rate term structure.

REFERENCES

Andersen, T.G., Bollerslev, T., Christoffersen, P.F. and Diebold, F.X. (2006) "Volatility Forecasting", Working paper to be published in *Handbook of Economic Forecasting* (Amsterdam: North-Holland).

Black, F. (1976) "The Pricing of Commodity Contracts", *Journal of Financial Economics*, 3(1–2): 167–79.

Black, F., Derman, E. and Toy, W. (1990) "A One-Factor Model of Interest Rates and its Application to Treasury Bond Options", *Financial Analysts Journal*, 46(1): 33–9.

Bollerslev, T. (1986) "Generalized Autoregressive Conditional Heteroscedasticity", *Journal of Econometrics*, 31(3): 307–27.

Campbell, J.Y., Lo, A.W. and MacKinlay, A.C. (1997) *The Econometrics of Financial Markets* (Princeton, NJ: Princeton University Press).

Carverhill, A. and Clewlow, L. (1994) "On the Simulation of Contingent Claims", *Journal of Derivatives*, 2(1): 66–74.

Chen, R. and Scott, L. (1993) "Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rates", *Journal of Fixed Income*, 4(1): 14–31.

Clewlow, L. and Strickland, C. (1997) "Monte Carlo Valuation of Interest Rate Derivatives Under Stochastic Volatility", *Journal of Fixed Income*, 7(1): 35–45.

Clinton, K. (1995) "The Term Structure of Interest Rates as a Leading Indicator of Economic Activity: A Technical Note", *Bank of Canada Review*, Winter 1994–5.

Dai, Q. and Singleton, K.J. (2000) "Specification Analysis of Affine Term Structure Models", *Journal of Finance*, 55(5): 1943–78.

Danilov, D. and Mandal, K. (2002) *Cross Sectional Efficient Estimation of Stochastic Volatility Short Rate Models*, ISSN 0169-2690, Memorandum no. 1614.

Day, J. and Lange, R. (1997) "The Structure of Interest Rates in Canada: Information Content about Medium-term Inflation", Working Paper, Bank of Canada, 97–100.

De Jong, F. (2000) "Time Series and Cross-Section Information in Affine Term-Structure Models", *Journal of Business & Economic Statistics*, 18(3): 300–14.

Diebold, F. and Li, C. (2003) "Forecasting the Term Structure of Government Bond Yields", Working Paper 10048, National Bureau of Economic Research, http://www.ssc.upenn.edu/~diebold

Engle, R.F. (1982) "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation", *Econometrica*, 50(4): 987–1008.

Eraker, B. (2001) "MCMC Analysis of Diffusion Models with Application to Finance", *Journal of Business and Economic Statistics*, 19(2): 177–91.

Fong, H.G. and Vasicek, O.A. (1992) "Interest Rate Volatility as a Stochastic Factor", Working Paper, Gifford Fong Associates.

Fornari, F. and Mele, A. (2005) "Approximating Volatility Diffusions with CEV-ARCH Models", *Journal of Economic Dynamics and Control*, 30(6): 931–66.

Gouriéroux, C. and Monfort, A. (1996) *Simulation Based Econometric Methods* (Oxford, UK: Oxford University Press).

Guay, A. and Ghysels, E. (1998) *Structural Change Tests for Simulated Method of Moments*, Cirano 98s-19.

Hamilton, J.D. (1994) *Time Series Analysis* (Princeton, NJ: Princeton University Press).

Heath, D., Jarrow, R.A. and Morton, A. (1989) "Contingent Claim Valuation with a Random Evolution of Interest Rates", *Review of Futures Markets*, 9(1): 54–76.

Heath, D., Jarrow, R.A. and Morton, A. (1992) "Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation", *Econometrica*, 60(1): 77–106.

Ho, T.S.Y. and Lee, S.B. (1986) "Term Structure Movements and Pricing Interest Rate Contingent Claims", *Journal of Finance*, 41(5): 1011–28.

Hull, J. and White, A. (1990) "Valuing Derivative Securities Using the Explicit Finite Differences Method", *Journal of Financial Quantitative Analysis*, 25(1): 87–100.

James, J. and Webber, N. (2000) *Interest Rate Modelling* (New York, NY: John Wiley & Sons).

Jamshidian, F. (1989) "An Exact Bond Option Formula", *Journal of Finance*, 44(1): 205–9.

Jarrow, R. (1996) *Modelling Fixed Income Securities and Interest Rate Options* (New York, NY: Mc Graw Hill).

Lautier, D. (2002) "The Kalman Filter in Finance: An Application to Term Structure Models of Commodity Prices and a Comparison between the Simple and the Extended Filters". *Centre de la Recherche sur la Gestion*, Working Paper, Université Paris Dauphine, 2002–11.

Litterman, R. and Scheinkman, J. (1991) "Common Factors Affecting Bond Returns", *Journal of Fixed Income*, 1(1): 54–61.

Longstaff, F. and Schwartz, E. (1992) "Interest Rate Volatility and the Term Structure: A Two Factor General Equilibrium", *Journal of Finance*, 47(4): 1259–82.

McFadden, D. (1989) "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration", *Econometrica*, 57(5): 995–1026.

Mills, T.C. (1999) *The Econometric Modelling of Financial Time Series* (Cambridge, UK: Cambridge University Press).

Nelson, D.B. (1990) "ARCH Models as Diffusion Approximation", *Journal of Econometrics*, 45(1): 7–38.

Racicot, F.-É. and Théoret, R. (2005) "Quelques Applications du Filtre de Kalman en Finance: Estimation et Prévision de la Volatilité Stochastique et du Rapport Cours-Bénéfices", University of Quebec at Montreal Working Paper.

Rebonato, R. (1998) *Interest-Rate Option Models* (New York, NY: John Wiley & Sons).

Selby, M.J.P. and Strickland, C. (1995) "Computing the Fong and Vasicek Pure Discount Bond Pricing Formula", *Journal of Fixed Income*, 5(1): 78–84.

Taylor, S.J. (1994) "Modeling Stochastic Volatility: A Review and Comparative Study", *Mathematical Finance*, 2(2): 183–204.

Théoret, R. and Rostan, P. (2002a) *Improving the Monte Carlo Simulation in the Fong & Vasicek Framework Using Bollinger Bands*. Centre de Recherche en Gestion, Document de travail, http://www.esg.uqam.ca/esg/crg/papers/2002/23-2002.pdf

Théoret, R. and Rostan, P. (2002b) *Empirical Foundations of Interest Rate Option Models: The Case of the Montreal Exchange*. Centre de Recherche en Gestion, Document de travail 15 – 2002, http://www.esg.uqam.ca/esg/crg/papers/2002/25-2002.pdf

Théoret, R. and Rostan, P. (2004) "Les Bandes de Bollinger Comme Technique de Réduction de la Variance des Prix d'Options sur Obligations Obtenus par la Simulation de Monte Carlo", *L'Actualité économique, Revue d'analyse économique*.

Théoret, R. and Rostan, P. (2004) "De la Rehabilitation du Modèle de Black: Tests Empiriques de Modèles d'Options sur Taux d'Intérêt", Luxembourg Economic Papers, No. 18, pp. 69–84.

Théoret, R., Rostan, P. and El-Moussadek, A. (2004)"Forecasting the Canadian Interest-Rate Term Structure",. Working Paper, UQAM, http://www.esg.uqam.ca/document/

Vasicek, O. (1977) "An Equilibrium Characterization of the Term Structure", *Journal of Financial Economics*, 5(2): 177–88.

CHAPTER 6

Idiosyncratic Risk, Systematic Risk and Stochastic Volatility: An Implementation of Merton's Credit Risk Valuation

*Hayette Gatfaoui**

6.1 INTRODUCTION

Originally Sharpe (1963) stated the dependence of stock returns vis-à-vis systematic (for example, market or undiversifiable) risk and idiosyncratic (for example, specific or diversifiable) risk. Indeed, systematic risk is common to any risky asset in the financial market whereas idiosyncratic risk is peculiar to the asset under consideration. Therefore, credit risky assets, such as corporate bonds or debt, should satisfy such a dependence feature. Many

* I would like to thank participants at the International AFFI conference (Cergy, France, June 2004) and Deloitte Risk Management Conference (Antwerp, Belgium, May 2005). I also thank Professors J.-P. Chateau and J. Wu (ESC Rouen) for their remarks. The usual disclaimer applies.

authors have investigated this assumption to test whether credit risk is of systematic or idiosyncratic nature. We focus on the most recent findings (see Gatfaoui, 2003, for a brief survey).

In 1989, Fama and French showed the influence of systematic risk on default risk premia of corporate bonds. Specifically, as systematic risk is correlated with macro-economy, its influence on credit risk is studied through business-cycle indicators (see Wilson, 1998; Nickell, Perraudin and Varotto, 2000; Gatfaoui and Radacal, 2001; Bangia, Diebold, Kronimus, Schlagen and Schuermann, 2002). Further, Spahr, Schwebach and Sunderman (2002) study speculative grade debt along with the (Fama and French, 1989) definition of bonds' risk premia. Estimating historical default losses with Altman's actuarial approach, the authors find that the speculative bond market prices both default and systematic risk with efficiency. Studying contemporaneous and first order correlations between frequency and severity of annual defaults, they show that default and systematic risk are coincident risks. In the same way, Koopman and Lucas (2005) resort to a multivariate unobserved component approach to describe jointly credit spreads and business failure rates with macro-economic behavior. They find empirical evidence of a correlation between credit risk and macro-economy. Moreover, Elton, Gruber, Agrawal and Mann (2001) show the existence of a systematic risk premium in corporate spot spreads (for example, the difference between corporate and Treasury yields). Further, Delianedis and Geske (2001) study the components of corporate credit spreads in the lens of a structural model. First, they find that default risk represents only a small portion of credit spreads. Then, they conclude that both default and recovery risk fail to explain fully credit risk and credit spreads whereas taxes, jumps in firm value, liquidity and market risk factors explain mainly such variables. More precisely, Collin-Dufresne, Goldstein and Martin (2001) find that a common latent factor in corporate bonds drives mostly credit spreads' changes. Analogously, Aramov, Jostova and Philipov (2004) find that systematic factors drive two thirds of credit spread changes whereas the other third is driven by firm-level fundamentals. Differently, Campbell and Taksler (2003) study the effect of equity volatility on corporate bond yields. They find that idiosyncratic volatility is as much important as credit ratings in explaining cross-sectional variation in yields. On average, idiosyncratic risk and ratings explain two-thirds of such variations. Along the same lines, Malkiel and Xu (2002) conclude that idiosyncratic volatility explains cross-sectional expected asset returns more than the CAPM beta coefficient or size measures do. Linear as well as non-linear influences of the beta coefficient on returns are mitigated. Later in 2003, they show that idiosyncratic risk affects market returns. Whereas, Goyal, and Santa-Clara (2003) study the average stock risk in addition to market risk. They estimate the average stock risk (for example, cross-sectional average stock variances) with the methodology of Campbell, Lettau, Malkiel and Xu (2001). First, they find that average stock risk is mainly driven by idiosyncratic risk (Campbell and Taksler, 2003). Though market variance has no predictive power for market return, a significant positive relationship prevails between average stock variance and market return. Finally, Stein, Kocagil, Bohn and Akhavein (2003) analyze default risk in the lens of idiosyncratic and systematic risk. Given their results, idiosyncratic information is mostly important for predicting middle market defaults.

Documented research has shed light on the typology and components of credit risk. Given the state of the art, credit risk has to be envisioned along with two dimensions, namely systematic and idiosyncratic risk. Such a typology is used by Gatfaoui (2003) to price risky debt in a Merton (1974) framework where diffusion parameters are constant. However, under its constant parameter assumption, Merton's model leads to implied spreads, which are far below observed credit spreads. Indeed, Eom, Helwege, and Huang (2004) show that adding stochastic interest rates correlated with firm value in Merton's model fails to offset the implied credit spreads' prediction problem. To solve this problem, Hull, Nelken, and White (2003) study the implications of Merton's model regarding implied at-the-money volatility and volatility skews. Their findings are supported by empirical data. First, implied volatility is sufficient to predict credit spreads. Second, there is a positive relationship between credit spreads and implied volatility, and between volatility skews and both implied credit spreads and implied volatility.

Third, implied volatility plays a major role in explaining credit spreads. Finally, as historical volatility leads to implied credit spreads, which underestimate their observed counterparts, the implied volatility approach exhibits a superior performance in predicting credit spreads over time. Such findings are consistent with Black and Scholes (1973) option pricing-type models. Specifically, such models exhibit a volatility smile (for example, implied volatility is a U-shaped function of the option's moneyness), which is determined by stochastic volatility, maturity and systematic risk among others (see Äijö, 2003; Backus, Foresi, Li and Wu, 1997; Duque and Lopes, 2003 for details, and Psychoyios, Skiadopoulos and Alexakis, 2003, for a survey about stylized facts of volatility as well as stochastic volatility models). Moreover, Eberlein, Kallsen and Kristen (2002/2003) study different representations of asset returns' volatility. They consider successively a constant parameter, a non-parametric model, a GARCH model, an autoregressive exponential model, a composite model,[1] and finally a stochastic volatility diffusion model generating an implied volatility. Their classification according to Basel rules[2] shows that first implied volatility models, and second GARCH-type models perform much better than other volatility representations in terms of Value-at-Risk forecasts (for example, frequency of excessive losses that determines required capital reserves).

In the light of such results, we extend the work of Gatfaoui (2003) to price risky debt in a Merton framework with stochastic volatility. For this purpose, our chapter is organized as follows: section 6.2 states the basis for the stochastic functional-based credit pricing model, then we underline the link with stochastic volatility models and introduce our pricing methodology. After specifying our stochastic functionals, we formalize our stochastic volatility model in section 6.3. Then, section 6.4 undertakes a simulation study to assess the impact of stochastic volatility both on risky debt valuation and credit spreads, and finally section 6.5 draws some concluding remarks.

6.2 THE GENERAL MODEL

We introduce the dynamic pricing of a firm's risky debt while valuing its total assets (for example, firm assets value). The mathematical background as well as pricing methodology is introduced along with the setting proposed by Gatfaoui (2003).

6.2.1 Basic setting

Consider a probability space (Ω, F, P) with a natural filtration $F_t = \sigma(w_s, 0 \leq s \leq t)$ where $w'_t = (W_t^X, W_t^I)$. (W_t^X) and (W_t^I) are two independent P-Brownian motions and represent the public information set at current time t. Let $F = (F_t)_{t \in [0,T]}$ be the P-augmentation of F_t with $T < \infty$. We set the assumptions prevailing in the Black and Scholes (1973) and Merton (1974) worlds except that diffusion parameters are rather stochastic than constant (for example, incompleteness of financial market). Briefly, there is no arbitrage opportunity and the spot risk free interest rate r is constant.

Consider a firm whose assets value at current time t is V_t, which is an F_t-adapted process. This firm is supposed to issue two kinds of financial assets, namely a risky debt represented by a discount bond maturing at time T with terminal value B (for example, promised payment to debtholders), and no-dividend-paying equity. The firm's potential default can only occur at time T. Let X_t and I_t be the systematic and idiosyncratic risk factors respectively, describing any financial asset in the market, and therefore firm value. Namely, X_t represents that part of firm value, which depends purely on market conditions, and I_t represents that part of firm value, which results from firm-specific patterns. These two risk factors are F_t-adapted processes whose dynamics are:

$$\frac{dX_t}{X_t} = \mu_X(t, X_t)dt + \sigma_X(t, X_t)dW_t^X \qquad (6.1)$$

$$\frac{dI_t}{I_t} = \mu_I(t, I_t)dt + \sigma_I(t, I_t)dW_t^I \tag{6.2}$$

where functionals $\mu_X(t, X_t)$, $\sigma_X(t, X_t)$, $\mu_I(t, I_t)$ and $\sigma_I(t, I_t)$ are continuous F_t-measurable functions on $[0,T] \times R$. To ensure strong solutions to previous SDEs, we assume that these functionals are also bounded (Karatzas and Shreve, 1991). For this purpose, we set whatever $t \in [0,T]$ and X_t, $I_t \in R$: $\mu_X^l < \mu_X(t, X_t) < \mu_X^u$, $\sigma_X^l < \sigma_X(t, X_t) < \sigma_X^u$, $\mu_I^l < \mu_I(t, I_t) < \mu_I^u$, $\sigma_I^l < \sigma_I(t, I_t) < \sigma_I^u$ with $\mu_X^l, \mu_X^u, \sigma_X^l > 0$, $\sigma_X^u, \mu_I^l, \mu_I^u$, $\sigma_I^l > 0, \sigma_I^u$ constant values. As introduced in Gatfaoui (2003), the dependence of firm assets value *vis-à-vis* the two risk factors is as follows:

$$V_t = X_t^\beta I_t \tag{6.3}$$

where β is the beta of firm assets value (for example, a constant estimate over our time horizon) as defined by the CAPM. Recall that X represents market conditions as well as business cycle, and I represents firm-specific features such as default and liquidity risk. Moreover, beta parameter is commonly thought as a systematic risk measure. As in Gatfaoui (2003), observing simultaneously systematic risk factor X and idiosyncratic risk factor I is equivalent to observe simultaneously firm assets value V and its specific risk factor I. Moreover, applying generalized Ito's lemma leads to the next expression for firm value under original probability P:[3]

$$\frac{dV_t}{V_t} = \mu_V(t, V_t, I_t)dt + \left[\beta\sigma_X(t, X_t)dW_t^X + \sigma_I(t, I_t)dW_t^I\right] \tag{6.4}$$

with[4] $\mu_V(t, V_t, I_t) = \beta\mu_X\left(t, \left(\frac{V_t}{I_t}\right)^{\frac{1}{\beta}}\right) + \mu_I(t, I_t) + \frac{1}{2}\beta(\beta - 1)\sigma_X^2\left(t, \left(\frac{V_t}{I_t}\right)^{\frac{1}{\beta}}\right)$.

Setting $d\langle V, I\rangle_t = \rho(t, V_t, I_t)dt$, the instantaneous (stochastic) correlation between firm value and its idiosyncratic risk factor is then:

$$\rho(t, V_t, I_t) = \frac{\sigma_I(t, I_t)}{\sigma_V(t, V_t, I_t)} \tag{6.5}$$

where $\sigma_V(t, V_t, I_t) = \sqrt{\beta^2\sigma_X^2\left(t, \left(\frac{V_t}{I_t}\right)^{\frac{1}{\beta}}\right) + \sigma_I^2(t, I_t)}$ is the global volatility of the instantaneous return of a firm's assets value. This global stochastic volatility depends on the beta parameter, and the respective volatilities of the two risk factors affecting firm value. Since our diffusions' functionals are bounded, global volatility is therefore bounded as a continuous function of these functionals. Our specification follows the results of Campbell, Lettau, Malkiel and Xu (2001) who show that the global volatility of any financial asset has both a systematic component (systematic volatility) and an idiosyncratic component (a specific volatility). Specifically, unsystematic volatility allows for accounting for security- as well as event-specific factors (for example, financial and corporate events), which are known to impact asset returns' volatility (Hilliard and Savickas, 2002). However, though idiosyncratic risk plays an increasing role, given market history, global volatility remains driven by its systematic component (market volatility). Recall that global volatility is the global risk level's target of the firm in accordance with its shareholders' interests (financial and dividend policies). Moreover, volatility is also considered as a proxy of liquidity risk. Indeed, Karpoff (1987), Lamoureux and Lastrapes (1990) and Schwert (1989) show that volatility is correlated with trading volume.

Moreover, considering expression (6.5) of the correlation coefficient and the firm's value dynamic (6.4), the diffusion of the firm's assets value takes a new form in the historical universe:

$$\frac{dV_t}{V_t} = \mu_V(t, V_t, I_t)dt + \sigma_V(t, V_t, I_t)\left[\varsigma_\beta\sqrt{1-\rho^2(t, V_t, I_t)}dW_t^X + \rho(t, V_t, I_t)\, dW_t^I\right] \tag{6.6}$$

where $\zeta_\beta = \text{sign}(\beta)$ represents the sign of beta (for example, $\zeta_\beta = 1$ if $\beta \geq 0$ and $\zeta_\beta = -1$ if $\beta < 0$). Therefore, describing the firm value's dynamic with relations (6.1), (6.2) and (6.3) is equivalent to characterizing the firm's assets value with relations (6.4) or, equivalently, (6.6) and (6.2). As this specification introduces two risk factors whilst we only observe firm assets value, we therefore lie in an incomplete market. Such a setting appears to be equivalent to a stochastic volatility framework provided that global volatility $\sigma_V(t, V_t, I_t)$ is non-zero whatever $(t, V_t, I_t) \in [0, T] \times R^2$.

6.2.2 Stochastic volatility and Merton's pricing

Indeed, relations (6.6) and (6.2) are similar to the state-diffusion and stochastic volatility model of Hofmann, Platen and Schweizer (1992). In this case, we have more risk factors (systematic and idiosyncratic risk) than existing or, equivalently, primary assets (firm value). Consequently, we are unable to give a unique price to any contingent claim on firm assets value. At best, we can define bounds for such a price (Frey and Sin, 1999) or minimize the uncertainty while computing a price. We address these two points therein. First, to shed light on the stochastic volatility analogy, we assume that firm value's global volatility $\sigma_V(t, V_t, I_t)$ is a $C^{1,2}([0, T] \times R^2)$ function (continuous, once derivable relative to time, and twice derivable relative to its two last arguments). It is sufficient to assume that $\sigma_X(t, X_t)$ and $\sigma_I(t, I_t)$ are two $C^{1,2}([0, T] \times R^2)$ functions. Consider the firm value's global variance $R(t, V_t, I_t) = {\sigma_V}^2(t, V_t, I_t)$, and let $R_x(t, V_t, I_t) = \partial R(t, V_t, I_t)/\partial x$, $R_{xx}(t, V_t, I_t) = \partial^2 R(t, V_t, I_t)/\partial x^2$ and $R_{xy}(t, V_t, I_t) = \partial^2 R(t, V_t, I_t)/\partial x \partial y$ for

$x, y \in \{t, V_t, I_t\}$. Applying multivariate Ito's lemma to the global variance of firm value's instantaneous return gives $dR(t, V_t, I_t) = Trend\ dt + Vol_1\ dW_t^X + Vol_2\ dW_t^I$:

$$Trend = R_t(t, V_t, I_t) + R_V(t, V_t, I_t)\, V_t \mu_V(t, V_t, I_t) + R_I(t, V_t, I_t)\, \mu_I(t, I_t)\, I_t$$

where

$$+ \frac{R_{VV}(t, V_t, I_t)}{2} \sigma_V^2(t, V_t, I_t)\, V_t^2 + \frac{R_{II}(t, V_t, I_t)}{2} \sigma_I^2(t, I_t)\, I_t^2$$
$$+ R_{VI}(t, V_t, I_t)\, \sigma_V(t, V_t, I_t)\, V_t \rho(t, V_t, I_t)\, \sigma_I(t, I_t)\, I_t$$

$$Vol_1 = R_V(t, V_t, I_t)\sigma_V(t, V_t, I_t) V_t \varsigma_\beta \sqrt{1 - \rho^2(t, V_t, I_t)}$$
$$Vol_2 = R_V(t, V_t, I_t)\sigma_V(t, V_t, I_t) V_t \rho(t, V_t, I_t) + R_I(t, V_t, I_t)\sigma_I(t, I_t)\, I_t$$

Hence, the stochastic volatility framework becomes obvious. Indeed, the dynamics of firm value and its global variance depend on two stochastic parts which are correlated.[5] This setting has important implications for Merton-type pricing models.

Following Merton (1974), the firm assets value is the sum of equity value $E(V, \tau)$ and debt value $D(V, \tau)$ such that $V_t = E(V_t, \tau) + D(V_t, \tau)$ with $\tau = (T - t)$ time to maturity, and following conditions: $E(0, \tau) = 0$, $E(V_t, \tau) = V_t - D(V_t, \tau) \geq 0$, $E(V_T, 0) = \max(0, V_T - B) = (V_T - B)^+$. The option nature of a firm's balance sheet leads to consider equity as a European call on firm value, with a strike equal to the promised payment (to debtholders) at firm debt's maturity. Hence, valuing risky debt requires pricing a European call. However, as we lie in an incomplete market, there exists an infinity of equivalent martingale measures allowing to price this European call (Mele and Fornari, 2000). On the other hand, taking the risk-free asset as a numeraire, the discount price of firm value becomes a semi-martingale under historical probability P. Nevertheless, among the set of equivalent martingale measures compatible with V, there exists a unique equivalent martingale measure $\hat{P}$, which minimizes the surrounding uncertainty or, equivalently, the relative entropy measure (Delbaen and Schachermayer, 1996; Föllmer and Schweizer, 1991; Gouriéroux, Laurent and Pham, 1998; Musiela and Rutkowski, 1998). Similar to a Hull and White (1987) setting, fluctuations in stochastic volatility generate a risk, which is not compensated. This feature explains the existence of $\hat{P}$, which is called the minimal equivalent martingale measure. Probability measure $\hat{P}$ is uniquely defined by its Girsanov density (Karatzas and Shreve, 1991) and Karatzas (1996) as:

$$\hat{L}(t) = \left.\frac{d\hat{P}}{dP}\right|_{F_t} = \exp\left\{-\int_0^t \alpha_1(s, V_s, I_s) dW_s^X - \int_0^t \alpha_2(s, V_s, I_s) dW_s^I\right\}$$
$$\times \exp\left\{-\frac{1}{2}\int_0^t (s, V_s, I_s) ds\right\}$$

where $\alpha_1(t, V_t, I_t) = \alpha(t, V_t, I_t)\varsigma_\beta\sqrt{1-\rho^2(t, V_t, I_t)}$; $\alpha_2(t, V_t, I_t) = \alpha(t, V_t, I_t)$ $\rho(t, V_t, I_t)$; $\alpha(t, V_t, I_t) = \frac{\mu_V(t, V_t, I_t) - r}{\sigma_V(t, V_t, I_t)}$.

$\alpha(t, V_t, I_t)$ is the global market risk premium due to the global (aggregate) risk borne by firm value whereas $\alpha_1(t, V_t, I_t)$ and $\alpha_2(t, V_t, I_t)$ are the market risk premia related respectively to the systematic and idiosyncratic risk factors affecting firm value. Consequently, the dynamic of $\ln(V)$ (for example, firm value's dynamic) under the minimal martingale measure $\hat{P}$ writes after applying generalized Ito's lemma:

$$d\ln(V_t) = \left[r - \frac{\sigma_V^2(t, V_t, I_t)}{2}\right]dt + \sigma_V(t, V_t, I_t)\,\varsigma_\beta\sqrt{1-\rho^2(t, V_t, I_t)}d\hat{W}_t^X + \sigma_V(t, V_t, I_t)\,\rho(t, V_t, I_t)d\hat{W}_t^I \tag{6.7}$$

where $d\hat{W}_t^X = \alpha_1(t, V_t, I_t)dt + dW_t^X$ and $d\hat{W}_t^I = \alpha_2(t, V_t, I_t)dt + dW_t^I$ are two independent F_t-adapted $\hat{P}$-Brownian motions. Under the incomplete-market assumption, the no-arbitrage principle and minimal martingale measure allow us to price the European call on firm value V (see Hofmann, Platen and Schweizer (1992) for option pricing in an incomplete market, and El Karoui, Jeanblanc and Shreve (1998) for properties of the Black and Scholes formula). Indeed, the European call's current value (firm's equity) is the expected discount value of its terminal payoff:

$$E(V_t, \tau) = E^{\hat{P}}\left\lfloor e^{-r\tau}(V_T - B)^+ \middle| F_t \right\rfloor \tag{6.8}$$

Using a Monte Carlo method (Jäckel, 2002) allows us to compute this expectation, and finally estimate debt value since we have $D(V_t, \tau) = V_t - E(V_t, \tau) = V_t - E^{\hat{P}}\lfloor e^{-r\tau}(V_T - B)^+ | F_t \rfloor$.

The stochastic volatility framework has nice properties since it adds flexibility to asset pricing, and then can improve Merton's debt valuation. However, the computational cost may be high since we need to simulate two Brownian motion paths. Nevertheless, such a setting may be extremely simple in some cases and highly useful for debt pricing. We focus on some useful and optimal simplification for Merton's debt pricing.

6.3 A STOCHASTIC VOLATILITY MODEL

In this section, we price risky debt under our previous stochastic volatility framework. We start from a general case, and then concentrate on a particular case while specifying our functionals. Our simplified framework allows for a tractable and easy computation of a firm's debt.

6.3.1 Model specification

Targeting a convenient degree of simplification, we make two major assumptions. First, we assume that the volatility functional of the idiosyncratic risk factor depends only on I. Second, we assume that the drift and volatility functionals of the systematic risk factor are deterministic functions of time. Our assumptions are motivated by the empirical features exhibited by equity volatility. To account for realistic features of equity volatility (see Psychoyios, Skiadopoulos and Alexakis, 2003, and Phoa, 2003), the stochastic global volatility of firm value has to be a stationary and mean reverting process (to encompass some shock effects on volatility). We explain therein how our assumptions fit the empirical characteristics.

Let constant real numbers ω and δ depend on prevailing financial and economic conditions, and assume that whatever $t \in]0, T]$ for $\mu_X(t)$ and $\sigma_X(t)$, and whatever $t \in [0,T]$ else:

$$\mu_X(t) = \omega t^\delta \quad \sigma_X(t) = \gamma t^\alpha \quad \mu_I(I_t) = \lambda \left(\frac{\varepsilon}{I_t} - 1 \right) \quad \sigma_I(I_t) = \Omega \sqrt{I_t}$$

where $\gamma > 0, \alpha < 0, \lambda > 0, \in > 0$ and, $\Omega > 0$ are constant parameters such that:

$$\frac{dX_t}{X_t} = \omega t^\delta dt + \gamma t^\alpha \, dW_t^X \qquad dI_t = \lambda(\varepsilon - I_t)dt + \Omega I_t \sqrt{I_t} dW_t^I$$

We can also assume that debt is issued at time $t_0 > 0$ and matures at $T = t_0 + \tau$ where τ is the initial lifetime of debt. Moreover, we exclude the case $t = 0$ for $\mu_X(t)$ since δ can take negative values. We also assume that $\mu_X(0) = \mu_0$ and $\sigma_X(0) = \sigma_0$ where $\mu_0 \in R$ and $\sigma_0 > 0$ are bounded constant values. Hence, the variance of instantaneous return of firm value writes $\sigma_V^2(t, I_t) = \beta^2 \sigma_X^2(t) + \sigma_I^2(I_t) = \beta^2 \gamma^2 t^{2\alpha} + \Omega^2 I_t = R(t, I_t)$ whatever $t \in]0, T]$, with $R(0, I_0) = \sigma_V^2(0, I_0) = \beta^2 \sigma_0^2 + \Omega^2 I_0$ being bounded. Our specification is consistent with Andersen, Bollerslev, Diebold and Ebens (2001) who study model-free measures of volatility and correlation of daily stock prices. The authors analyse time-varying features of stock returns (see, Bekaert and Wu, 2000; Bollerslev and Mikkelsen, 1999; Campbell, Lettau, Malkiel and Xu, 2001; Christensen and Prabhala, 1998) and they find two main results. First, variances exhibit a systematic common component in their evolution. Second, an asymmetric relationship prevails between returns and volatility. We then obtain:

$$R_V(t, I_t) = R_{VV}(t, I_t) = R_{II}(t, I_t) = R_{IV}(t, I_t) = 0 \quad R_I(t, I_t) = \Omega^2$$
$$R_t(t, I_t) = 2\alpha \beta^2 \gamma^2 t^{2\alpha - 1}$$

In this case, the variance satisfies the following SDE in historical universe:

$$dR(t, I_t) = \left[2\alpha \beta^2 \gamma^2 t^{2\alpha - 1} + \Omega^2 \lambda (\varepsilon - I_t) \right] dt + \Omega^3 I_t \sqrt{I_t} \, dW_t^I \qquad (6.9)$$

or, equivalently, under the minimal martingale measure:

$$dR(t,I_t) = \left[2\alpha\beta^2\gamma^2 t^{2\alpha-1} + \Omega^2\lambda(\varepsilon - I_t) - \Omega^3 I_t\sqrt{I_t}\alpha_2(t,I_t)\right]dt + \Omega^3 I_t\sqrt{I_t}\,d\hat{W}_t^I$$

with $\alpha_2(t,I_t) = \frac{\mu_V(t,I_t) - r}{\sigma_V(t,I_t)}\rho(t,I_t)$; $\rho(t,I_t) = \frac{\sigma_I(t,I_t)}{\sigma_V(t,I_t)}$; $\mu_V(t,I_t) = \beta\mu_X(t) + \mu_I(t,I_t) + \frac{1}{2}\beta(\beta-1)\,\sigma_X^2(t)$.

In the original universe, such a diffusion process behaves almost like a mean reverting square-root process except that the random shocks affecting its trend are higher in magnitude.[6] Moreover, when time t tends towards infinity, the stochastic variance of a firm's value tends towards $R(t,I_t) = \Omega^2 I_t$ such that global volatility reads $\sqrt{R(t,I_t)} = \Omega\sqrt{I_t}$. In the same way, the previous diffusion asymptotically takes the new form:

$$dR(t,I_t) = \lambda\left[\Omega^2\varepsilon - R(t,I_t)\right]dt + R(t,\,I_t)\sqrt{R(t,\,I_t)}\,dW_t^I$$

If variance is asymptotically zero, then its diffusion becomes $dR(t,I_t) = \lambda\Omega^2\varepsilon\,dt > 0$. Therefore, when t tends towards infinity and variance is zero, the zero threshold becomes a reflecting barrier for firm value's variance. Indeed, empirical features of equity volatility exhibit stationarity and mean reversion patterns. In the asymptotic case, $\Omega^2\varepsilon$ is the long-run mean of firm value's stochastic variance, and λ is the velocity of mean reversion. Our specification implies that the randomness in global variance comes only from idiosyncratic risk factor. Indeed, stochastic volatility comes from the non-observability of idiosyncratic risk factor. Namely, stochastic volatility is due to the intrinsic risk of firm value because such a risk is non-tradable. This setting gives some nice properties to firm's debt pricing. Under this framework, the next section undertakes some simulations of I_t and $R(t,I_t)$ for given parameter values, and varying β and λ.

6.3.2 Implication for debt pricing

The randomness of firm value's variance depends only on idiosyncratic risk factor (relation (6.9)). Moreover, correlation coefficient now expresses $\rho(t,I_t) = \frac{\Omega\sqrt{I_t}}{\sqrt{\beta^2\gamma^2 t^{2\alpha} + \Omega^2 I_t}}$, and the firm value's dynamic under the minimal martingale measure $\hat{P}$ then reads on time subset $[t,T]$:

$$\ln\left(\frac{V_T}{V_t}\right) = \left(r - \frac{\bar{\sigma}_V^2}{2}\right)\tau + \int_t^T \sigma_V(s,\,I_s)\varsigma_\beta\sqrt{1-\rho^2(s,I_s)}d\hat{W}_s^X + \int_t^T \sigma_V(s,\,I_s)\,\rho(s,\,I_s)d\hat{W}_s^I$$

where $\bar{\sigma}_V^2 = \frac{1}{\tau}\int_t^T \sigma_V^2(s,\,I_s)ds$ is the firm value's average variance over the time to maturity of debt (remaining life of the European call). Consider

$G_t = F_t \cup \{I_s, t \leq s \leq T\}$ and compute the two first moments of probability distribution of $\ln\left(\frac{V_T}{V_t}\right)$ conditional on G_t: $E^{\hat{P}}\left[\ln\left(\frac{V_T}{V_t}\right)\middle| G_t\right] = \left(r - \frac{\bar{\sigma}_V^2}{2}\right)\tau$, and $Var^{\hat{P}}\left[\ln\left(\frac{V_T}{V_t}\right)\middle| G_t\right] = \bar{\sigma}_V^2\,\tau$.[7] Hence, conditional on G_t, the firm value's natural logarithm $\ln\left(\frac{V_T}{V_t}\right)$ follows a normal law (firm value follows a lognormal law) with volatility parameter $\sqrt{\bar{\sigma}_V^2}$.

On the other hand, recall expression (6.8) of firm's equity or, equivalently, the European call on firm value under the minimal martingale measure. Applying the iterated expectations theorem, we get $E(V_t, \tau) = E^{\hat{P}}\left[E^{\hat{P}}\left[e^{-r\tau}(V_T - B)^+\middle| G_t\right]\middle| F_t\right]$. However, given the law of $\ln\left(\frac{V_T}{V_t}\right)$ conditional on G_t, equity value then reads $E(V_t, \tau) = E^{\hat{P}}\left[C_{BS}\left(\tau, r, V_t, B, \sqrt{\bar{\sigma}_V^2}\right)\middle| F_t\right]$ where $C_{BS}\left(\tau, r, V_t, B, \sqrt{\bar{\sigma}_V^2}\right)$ is the Black and Scholes (1973) price employed with an average time-dependent volatility. Consequently, equity value is the average Black and Scholes European call price over each possible volatility path. Our deterministic systematic risk volatility assumption leads then to an optimal Monte Carlo European call pricing. Indeed, we only need to generate one Brownian motion, namely the randomness affecting stochastic volatility (for example, idiosyncratic risk's Brownian motion). This setting allows a simple computation of debt value since:

$$D(V_t, \tau) = V_t - E^{\hat{P}}\left[C_{BS}\left(\tau, r, V_t, B, \sqrt{\bar{\sigma}_V^2}\right)\middle| F_t\right] \tag{6.10}$$

Given current information set, we can price firm's debt such that uncertainty is minimized. Moreover, as functional diffusion parameters are bounded on $[0, T]$, stochastic volatility is also bounded whatever $t \in [0, T]$ since $\sigma_V^l < \sigma_V(t, I_t)\sqrt{1 - \rho^2(s, I_s)} < \sigma_V^u$, with $\sigma_V^l = \beta^2\sigma_X^{l2} + \sigma_I^{l2} > 0$ and $\sigma_V^u = \beta^2\sigma_X^{u2} + \sigma_I^{u2}$. Therefore, the Black & Scholes call price is bounded by (Frey and Sin, 1999):

$$C_{BS}(\tau, r, V_t, B, \sigma_V^l) < C_{BS}\left(\tau, r, V_t, B, \sqrt{\bar{\sigma}_V^2}\right) < C_{BS}(\tau, r, V_t, B, \sigma_V^u) \tag{6.11}$$

which implies that both firm equity and debt values are bounded. We are then able to price corporate debt under the minimal martingale measure. We can also establish debt bounds depending on the magnitude of variations in firm value's volatility. Under our assumptions, systematic risk drives volatility's trend whereas idiosyncratic risk affects this trend through shocks. Consequently, the magnitude of variations in firm value's global volatility is driven by the impact of both systematic and idiosyncratic risk factors.

6.4 SIMULATION STUDY

We use Monte Carlo accelerators to examine behaviors of debt pricing as well as its related credit spread in a stochastic volatility setting. We study our pricing framework as a function of a systematic risk measure β and velocity λ. For statistical investigation purposes, the range of values we consider can be larger than the realistic range of values that describes the real world.

6.4.1 Volatility and debt

We simulate stochastic volatility and its impact on a firm's debt and equity. Then, we plot the paths obtained for these random variables or display their average values (arithmetic means of simulated data). Assuming that the initial debt's time to maturity is $\tau = T - t = 10$ years, we set $\alpha = -\frac{1}{4}$, $\varepsilon = 0.5$, $\gamma = 1$, $I_t = 3.5$, $\Omega = \sqrt{\frac{0.2}{\varepsilon}}$, and $R(t, I_t) = \beta^2 t^{-\frac{1}{2}} + 0.4\, I_t$. Specifically, we assume that debt is issued at time $t > 0$ and matures at $T = t + \tau$. We also assume that the starting value of the remaining life of firm's debt is $\tau = \tau_0 = 10$ years. Daily values of global volatility $\sqrt{R(t, I_t)}$ (equation (6.9)) are computed for different values of beta and lambda parameters (for example, $\beta = 0$, 0.5, 1, 1.5 and $\lambda = 0.2$, 1, 5) with t running from T-t to T (see Figures 6.1 to 6.3). Recall that $V_t = I_t$ when $\beta = 0$. Moreover, only β^2 intervenes in our global volatility.

The higher lambda is, the more stable are the evolutions and convergence to long-run means of stochastic volatility $\sqrt{R(t, I_t)}$, idiosyncratic factor I_t

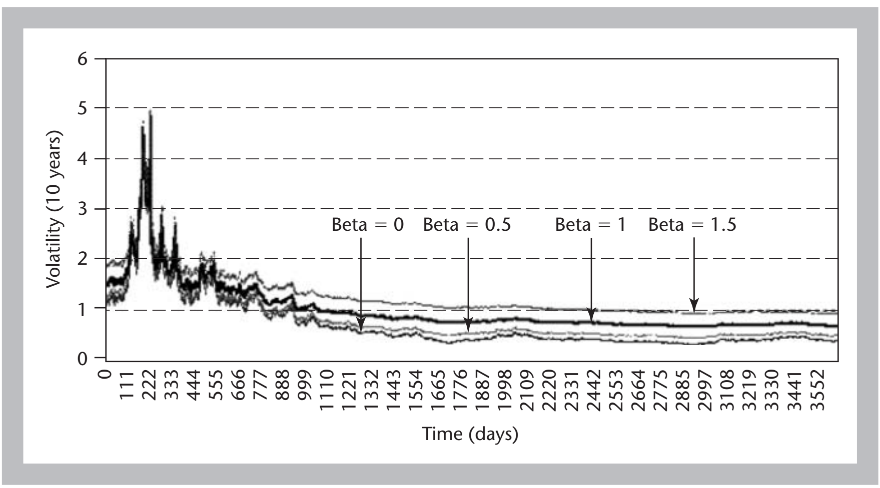


Figure 6.1 Simulated volatility when $\lambda = 0.2$

and correlation coefficient $\rho(t, I_t)$. On the other hand, the greater beta is, the higher firm value's volatility is for fixed λ. An increase in firm's global volatility can result from short-term capital movements as well as short-term investments among others. Moreover, Savickas (2003) focuses on event-induced variance increases while underlining the stochastic feature of asset returns' volatility. Specifically, the occurrence of given events in the market often engenders increases in volatility of asset returns. In contrast, the higher beta is, the lower the correlation coefficient becomes (Table 6.1). Notice that the idiosyncratic risk factor I_t depends only on λ and not on β.

Since I is independent of β, its average simulated values are 2.10, 1.50 and 0.60 when λ is 0.2, 1 and 5 respectively. First, the average level of idiosyncratic

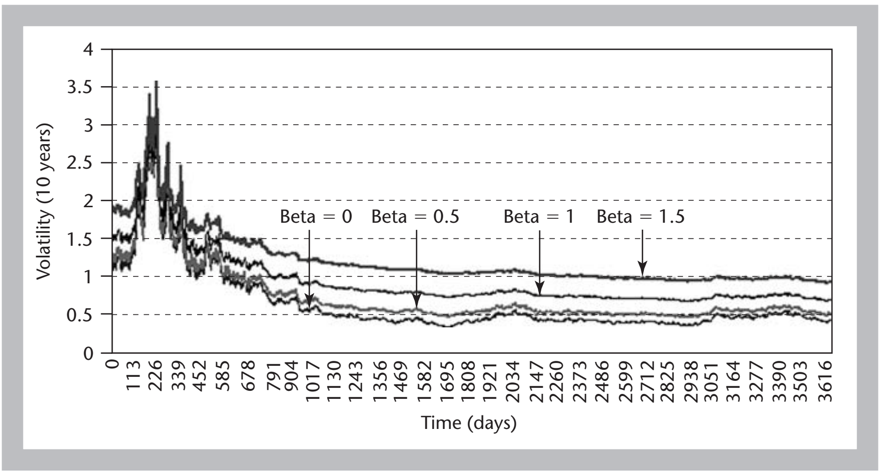


Figure 6.2 Simulated volatility when $\lambda = 1$

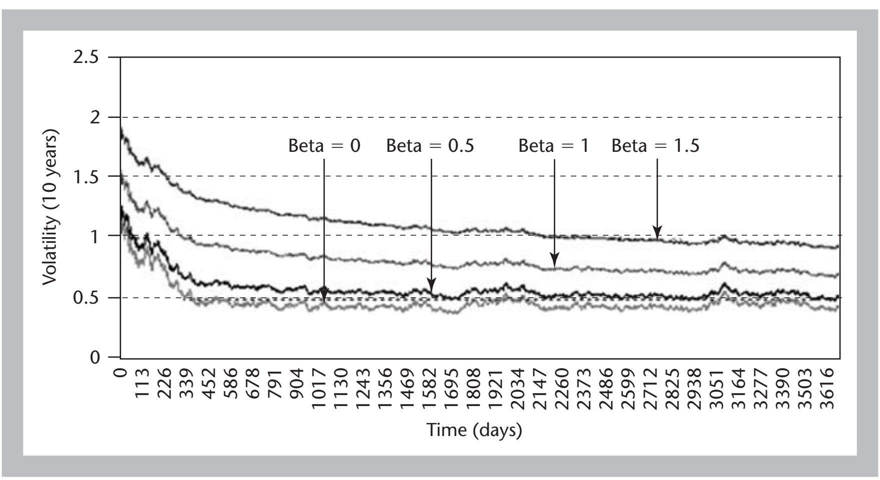


Figure 6.3 Simulated volatility when $\lambda = 5$

Table 6.1 Average values of daily simulated variables on $[t, T]$

Variable	$\lambda \mp \beta$	0	0.5	1	1.5
$\sigma_V(\%)$	0.2	71.69	80.53	101.19	127.24
ρ	0.2	1.00	0.85	0.65	0.51
$\sigma_V(\%)$	1	64.10	73.36	94.87	121.76
ρ	1	1.00	0.84	0.63	0.49
$\sigma_V(\%)$	5	47.79	58.57	82.68	111.76
ρ	5	1.00	0.81	0.58	0.43

risk factor is a decreasing function of λ. Second, the average correlation coefficient $\rho(.)$ and global volatility $\sigma_V(.)$ are respectively decreasing and increasing functions of the absolute value of β. Such a behavior is trivial given that $\rho(.)$ is the correlation between firm value and its idiosyncratic risk factor. Third, average correlation coefficient $\rho(.)$ and global volatility $\sigma_V(.)$ are both decreasing functions of λ.

We further set $I_t = 0.1$ and assume $\omega = 0$, which implies that $\mu_X(t) = 0$ whatever t. The diffusion of the idiosyncratic factor under the minimal martingale measure then writes as:

$$dI_t = \left[\lambda(\varepsilon - I_t) - \frac{\Omega^2 I_t^2}{\sigma_V^2(t, I_t)}(\mu_V(t, I_t) - r)\right] dt + \Omega I_t \sqrt{I_t}\, d\hat{W}_t^I$$

with $\sigma_V^2(t, V_t, I_t) = R(t, V_t, I_t)$, and $\mu_V(t, I_t) = \lambda(\frac{0.5}{I_t} - 1) + \frac{1}{2}\beta(\beta - 1)t^{-\frac{1}{2}}$. Then, the related average stochastic variance $\bar{\sigma}_V^2$ conditional on G_t reads $\bar{\sigma}_V^2 = \frac{1}{\tau}\int_t^T \sigma_V^2(s, I_s)ds = \frac{2\beta^2}{\tau}\left(\sqrt{T} - \sqrt{t}\right) + \frac{\Omega^2}{\tau}\int_t^T I_s ds$ when $\rho(s, I_s)$ is zero.

From formula (6.10), debt computation requires the estimation of the call's price, which we realize with Monte Carlo simulation and antithetic variables-based accelerators for variance reduction principle (Jäckel, 2002; Ripley, 1987; Rubinstein, 1981). Let $nsim$ be the number of simulations and $C_{BS,k}(.)$ be the call's price of the k-th simulation. Then, the estimated equity value conditional on current information is the arithmetic mean of simulated variables, namely:

$$E(V_t, \tau) = E^{\hat{P}}\left[C_{BS}\left(\tau, r, V_t, B, \sqrt{\bar{\sigma}_V^2}\right) \middle| F_t\right]$$
$$= \frac{1}{nsim}\sum_{k=1}^{nsim} \frac{C_{BS,k}\left(\tau, r, V_t, B, \sqrt{\bar{\sigma}_V^2}\right) + C_{BS,k}\left(\tau, r, V_t, B, -\sqrt{\bar{\sigma}_V^2}\right)}{2}$$

We realize monthly simulations with $r = 8\%$, $B = 13$, $V_t = 52$ where V and B are expressed in billions of dollars. Our examination then uses varying

Table 6.2 Average monthly simulated values of firm's debt

λ \ β	−1.5	−1	−0.5	0	0.5	1	1.5
0.2	2.25	4.93	8.02	8.94	7.65	4.24	1.79
1	3.48	4.54	7.47	8.94	6.86	3.75	1.82
5	3.70	4.34	7.08	8.94	6.72	3.65	1.88

moneyness (ratio V_t/B), volatility and time to maturity insofar as we focus on the combined effect of these determinants. After $nsim = 1000$ simulations, we compute average values of firm's debt for various levels of β and λ; as in Table 6.2.

Whatever λ, the firm's average debt is a concave function of β with a maximal value at $\beta^* = 0$. When $|\beta| < 1.5$, debt is a decreasing function of λ. In contrast, when $|\beta| > 1.5$, the reverse behavior takes place. Moreover, average debt's value is constant whatever λ when $\beta = 0$ since debt value is independent of λ under the minimal martingale measure. Debt value depends on equity, and equity does not depend on λ under the minimal martingale measure. Indeed, equity is a function of both firm value and idiosyncratic risk factor (see European call's expression). However, neither firm value nor idiosyncratic factor depends on λ when β is zero. Finally, realistic values of beta in Table 6.2 (for example, $\beta \in [0, 1]$) lead to interesting conclusions. First, an increase in beta generates a decrease in debt value. Namely, increasing systematic risk's impact allows for reducing debt level. Such a pattern represents the kind of phenomenon that takes place in a good side of business cycle. Indeed, a growth business cycle trend will improve the financial market's trend. Hence, increasing the correlation between firm value and market will decrease the firm's credit risk and debt level (good spillover effect, and benefits from the good side of the business cycle). Second, increasing idiosyncratic risk control (high λ) reduces the firm value's global risk, and then decreases the debt level for a given systematic risk level. Hence, when idiosyncratic risk is managed and stabilized around a convenient level, the firm can concentrate on the systematic risk side that impacts its business profile and profits. Systematic risk becomes the most important risk dimension in this case. Conversely, when idiosyncratic risk cannot be conveniently managed (low λ), such a risk can highly disturb the firm's financial and economic equilibrium since market conditions can strongly magnify its consequences. Specifically, the firm can undergo hard times before any mean reversion in idiosyncratic risk occurs. The two dimensions of a firm's business risk have then to be jointly managed. Indeed, the firm exhibits an increased level of structural financial risk when idiosyncratic risk is uncontrolled. Namely, the firm's accounting and financial patterns usually behave quite badly in such a situation (see Table 6.3).

Table 6.3 Average monthly simulated values of firm's equity

λ \ β	−1.5	−1	−0.5	0	0.5	1	1.5
0.2	50.07	61.55	70.32	71.52	56.83	59.55	129.46
1	51.19	60.47	67.87	71.52	52.76	58.64	135.57
5	54.15	62.77	70.97	71.52	57.51	65.90	154.30

Table 6.4 Average monthly simulated values of path dependent stochastic volatility in percent

λ \ β	−1.5	−1	−0.5	0	0.5	1	1.5
0.2	85.32	68.82	45.45	27.22	50.20	71.45	97.46
1	84.53	73.00	51.80	27.22	58.28	75.32	100.10
5	82.96	74.39	54.19	27.22	56.56	74.97	100.26

Generally, equity is a non-monotonous function of β and λ parameters. Equity increases for growing $\beta \leq 0$, decreases in $\beta = 0.5$, and goes on increasing for growing $\beta \in]0.5, 1.5]$. Specifically, equity is a convex function of λ for $|\beta| < 1.5$ with a minimum at $\lambda^* = 1$. Finally, it becomes an increasing function of λ for $|\beta| = 1.5$. Hence, results show global volatility's impact on both firm's equity and debt. Incidentally, existing literature has also shown that firm's global volatility impacts capital structure. Indeed, Leland and Toft (1996) shows that the longer debt's maturity is, the more sensitive are firm and debt values to firm value's global volatility. Moreover, the sensitivity of debt relative to an increase in global volatility reacts in the opposite way to equity's sensitivity relative to the same increase in global volatility.

Briefly, we also display in Table 6.4 the conditional expected value of average monthly stochastic volatility

$$\bar{\sigma}_V^e = E^{\hat{P}}\left[\sqrt{\frac{1}{\tau}\int_t^T \sigma_V^2(s, I_s)ds}\,\middle|\, G_t\right] = E^{\hat{P}}\left[\sqrt{\bar{\sigma}_V^2}\,\middle|\, G_t\right]$$

over the remaining time to maturity of debt, and under the minimal martingale measure. The stochastic integral composing firm value's variance is computed using the finite difference method, and average stochastic volatility is computed always using Monte Carlo simulation methodology.

Whatever λ, the average stochastic volatility is a convex function of β with a minimal value at $\beta^* = 0$. When $-1.5 < \beta < 0$ and $\beta = 1.5$, the average stochastic volatility is an increasing function of λ. In contrast, when $\beta = -1.5$, the average stochastic volatility decreases as a function of λ. On the other

Table 6.5 Average monthly simulated values of credit spreads in basis points

λ \ β	−1.5	−1	−0.5	0	0.5	1	1.5
0.2	4265.33	1921.20	345.89	2.86	507.29	2486.03	5285.82
1	2945.94	2194.89	566.89	2.86	862.23	2908.93	5156.47
5	2748.30	2348.05	743.01	2.86	938.75	2982.52	5030.70

hand, when $\beta = 0.5$ or $\beta = 1$, the average stochastic volatility is a concave function of λ. Moreover, the average stochastic volatility is constant whatever λ when $\beta = 0$. Such a behavior of stochastic volatility has some impact on the term structure of corporate credit spreads.

6.4.2 Credit spread

We extend our study while assessing the impact of stochastic volatility on credit spreads. In particular, we focus on the term structure of credit spreads.

Let $y(\tau)$ be the yield-to-maturity of firm's risky debt (for example, the default risky yield). Such a yield is linked with the current value of firm's debt such that $D(V_t, \tau) = e^{-y(\tau)\tau}B$, which implies that $y(\tau) = \frac{-1}{\tau}\ln\left(\frac{D(V_t,\tau)}{B}\right)$. Hence, related credit spreads (for example, yield spreads against government bonds) write $S(\tau) = y(\tau) - r = \frac{-1}{\tau}\ln\left(\frac{D(V_t,\tau)}{B}\right) - r$. For the sake of simplicity, we assume a flat risk-free term structure here. Then, simulated debt values allow to compute monthly related credit spreads with varying moneyness and time to maturity. Results are first displayed in Table 6.5. Second, part of these results is summarized in Figure 6.4, which plots credit spreads when lambda is 5.

Whatever λ, credit spreads are a convex function of β with a minimum at $\beta^* = 0$. Moreover, credit spread behaviors relative to λ are mitigated. For $|\beta| = 1.5$, credit spreads are decreasing functions of λ whereas the reverse is true for $|\beta| = 0.5$ or 1. Finally, credit spreads are constant when $\beta = 0$ due to the independence of debt relative to λ under the minimal martingale measure.

First, the higher the absolute value of beta is, the wider the related credit spread becomes for a given level of lambda. Second, the credit spread's level related to a given negative value of beta lies slightly under the credit spread's level related to the corresponding positive value of beta. Finally, credit spreads are convex decreasing functions of time to maturity. Equivalently, credit spreads are convex increasing functions of debt's maturity. Indeed, when time to maturity τ decreases from $\tau_0 = 10$ years to 0, maturity

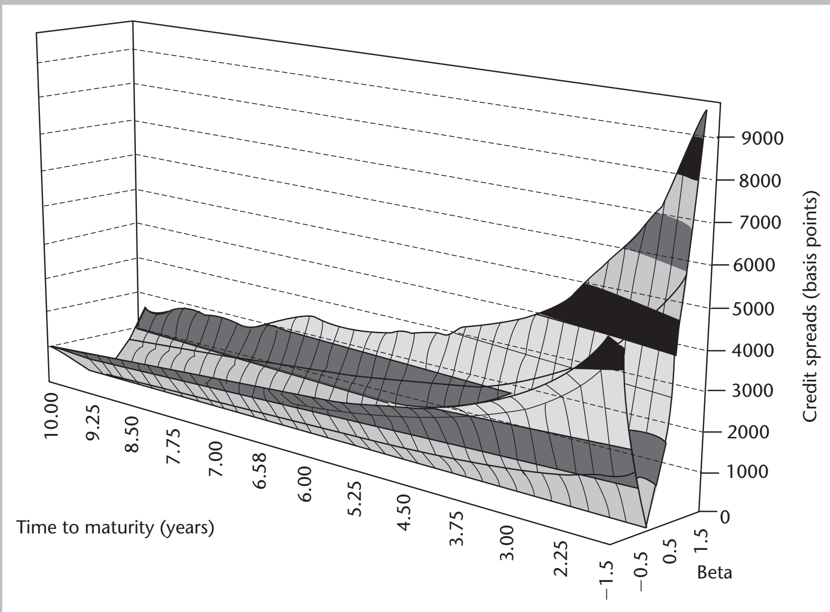


Figure 6.4 Credit spread when $\lambda = 5$

T increases from current time t to $(t + \tau_0)$. Moreover, current time t is equal to $(T - \tau_0)$ and T when τ is respectively τ_0 and 0. Such a behavior is consistent with the work of Collin-Dufresne and Goldstein (2001) and Gemmill (2002). Indeed, Gemmill (2002) exhibits an upward-sloping credit spread term structure, which is consistent with Merton-type profiles provided that firms' leverages exhibit a drift over time. Collin-Dufresne and Goldstein (2001) find a convex decreasing shape (relative to time to maturity) for credit spreads of firms with stationary leverages. Moreover, the convexity pattern describes investment grade bonds whereas a concavity pattern describes speculative grade bonds (junk bonds). Analogously to Collin-Dufresne and Goldstein (2001), we consider the quasi-debt leverage ratio $d_t = \frac{Be^{-r\tau}}{V_t}$ of (Merton, 1974). Our log-leverage ratio $d\ell_t = \ln(d_t)$ follows a stationary normal process such that $d\ell_t \sim N\left(\frac{\sigma_V^2(t, V_t, I_t)}{2}dt, \sigma_V^2(t, V_t, I_t)dt\right)$ conditional on F_t. The starting value of our quasi-debt leverage is 11.23% under our assumptions. Thus, we consider investment grade-type debt in a high volatility framework where firm's global volatility stabilizes after the first five years following debt's issue. Notice that the first two moments describing d_t's lognormal law depend on firm value's global variance $\sigma_V^2(t, V_t, I_t)$, which follows an asymptotically mean reverting process. However, unlike Collin-Dufresne and Goldstein (2001), our quasi-debt leverage is not mean-reverting, which allows to take

Table 6.6 Average bounds of simulated aggregate volatility in percent

λ \ β	0	0.5	1	1.5
0.2	17.84	41.13	64.68	90.34
	34.04	65.74	103.07	151.35
1	17.84	45.10	66.92	92.17
	34.04	84.04	104.06	151.45
5	17.84	46.56	67.91	92.39
	34.04	74.97	103.63	151.33

into account part of speculative grade corporate debt. Under convenient assumptions about starting values and stochastic variables, a relevant choice of stochastic functionals will describe some specific rating grades among given speculative grade rating classes of corporate debt.

In our bounded volatility/bounded diffusion parameters setting, we can establish bounds for credit spreads. Hence, we propose an alternative approach to the one of Chen and Huang (2002). The authors give analytical bounds to credit spread term structure in order to solve the problem of negative implied default probabilities. Such a problem arises when calibrating credit models to empirical data, and comes from the no-arbitrage principle's violation. In the same way, we establish bounds for implied credit spreads. We first give the average aggregate volatility's bounds we get under our framework. Briefly, we compute related bounds σ_V^l and σ_V^u for our *nsim* simulations of V_t, as well as the arithmetic mean of all obtained σ_V^l and σ_V^u in Table 6.6.

Whatever λ, average volatility's bounds are constant when $\beta = 0$. When beta is non-zero, average aggregate volatility's upper bounds are concave functions of λ with a maximum value at $\lambda^* = 1$. Differently, average aggregate volatility's lower bounds increase strictly with λ. Moreover, our volatility's bounds are increasing functions of beta parameter. Consequently, formulae (6.10) and (6.11) allow to bound debt $D^l(\tau) < D(V_t, \tau) < D^u(\tau)$ with $D^l(\tau) = V_t - E^{\hat{P}}[C_{BS}(\tau, r, V_t, B, \sigma_V^u)| F_t]$ and $D^u(\tau) = V_t - E^{\hat{P}}[C_{BS}(\tau, r, V_t, B, \sigma_V^l)|F_t]$. Thus, the risky yield-to-maturity becomes bounded as $\frac{1}{\tau}\ln\left(\frac{B}{D^u(\tau)}\right) < y(\tau) < \frac{1}{\tau}\ln\left(\frac{B}{D^l(\tau)}\right)$. Hence, related credit spread is bounded since $S^l(\tau) < S(\tau) < S^u(\tau)$ where $S^l(\tau) = \frac{1}{\tau}\ln\left(\frac{B}{D^u(\tau)}\right) - r$, and $S^u(\tau) = \frac{1}{\tau}\ln\left(\frac{B}{D^l(\tau)}\right) - r$ (see Table 6.7).

Average credit spread bounds behave like the average monthly credit spreads reported in Table 6.5. Therefore, we can give an interval for possible

Table 6.7 Average bounds of simulated credit spreads in basis points

λ \ β	−1.5	−1	−0.5	0	0.5	1	1.5
0.2	3481.72	1349.20	159.84	0.06	205.49	1493.86	3779.83
	5220.18	2316.99	373.59	20.93	614.73	2521.02	5686.73
1	2583.16	1479.75	268.96	0.06	305.11	2263.50	3549.02
	3636.47	2394.72	665.56	20.93	1173.51	3260.99	5517.20
5	2537.73	1630.65	377.01	0.06	473.15	2422.85	3114.64
	3231.33	2545.06	781.39	20.93	1215.32	3362.97	5178.89

variation of credit spread at each time between debt's issue date and maturity. Evolutions of credit spread bounds over time can be viewed as extreme scenarios describing credit spread's evolution (best and worst possible situations).

6.5 CONCLUSION

We focused on the credit risk valuation of Gatfaoui (2003) whose modeling proposes to value corporate debt in a Merton framework, and accounts for systematic and idiosyncratic risk. Specifically, the option nature of debt allows the author to price corporate debt through a call on firm value consistently with the constant parameter-based dynamics of systematic and idiosyncratic risk factors. Our work addressed the extension of such a setting in two key points.

First, we considered the stochastic parameter-based dynamics of the two previous risk factors. Under regularity conditions, this setting is equivalent to a stochastic volatility option pricing (stochastic credit pricing) model. Namely, we consider two risk sources affecting firm value, whereas we only observe firm value. Consequently, we lie in an incomplete market where incompleteness is due to the unobservable idiosyncratic part of firm value (incompleteness engenders a stochastic volatility for firm value). Hence, the no-arbitrage principle and minimal martingale measure allow to price firm's equity and therefore debt. Such a valuation becomes possible in the historical universe as well as the minimal martingale measure's universe under a bounded volatility assumption. The equivalent minimal martingale measure is useful insofar as it reduces global risk to its minimal component, namely intrinsic (idiosyncratic) risk.

Second, we illustrated such a framework while specifying the stochastic parameters of the diffusions under consideration. We undertook corporate debt pricing in a bounded volatility case. Under our functional assumptions, we obtain an asymptotically mean reverting stochastic volatility process relative to time. Moreover, an interesting implication of our model is the stochastic correlation coefficient prevailing between firm value and its idiosyncratic risk factor. We study this dependence feature as well as the stochastic volatility process through simulations. Specifically, accelerator-based Monte Carlo simulations are undertaken to study the behaviors of equity, debt and credit spreads as functions of our model's parameters. In the same way, we also simulated the path-dependent average stochastic volatility of our pricing framework. The advantage of such a setting is the flexibility given by parameters since we are able to account for many risk scenarios and various market-linked firms. Moreover, the boundedness of firm value's stochastic volatility implies the boundedness of related equity, debt and credit spreads. Such bounds can be viewed as extreme scenarios (worst/minimal potential losses due to increased/reduced global risk where the level of firm's global risk depends on systematic and idiosyncratic risk factors). In particular, our stochastic setting can allow for a more accurate computation of historical conditional default probabilities. As default probabilities allows for assessing creditworthiness of counterparts, the possible boundedness of such probabilities given likely scenarios has some non-negligible importance and significance.

Our paper presents then some non-negligible advantages. First, volatility is fundamental for asset valuation, risk management and portfolio diversification (Eberlein, Kallsen and Kristen, 2002/2003). Stochastic volatility models are useful tools to account for fundamental time-varying volatility (latent volatility component) of financial assets (Hwang and Satchell, 2000). Moreover, volatility is commonly thought as a liquidity indicator (Kerpoff, 1987; Lamoureux and Lastrapes, 1990; Schwert, 1989). Hence, incorporating a stochastic volatility in credit risk modeling implicitly accounts for some liquidity effects describing credit risky assets (Collin-Dufresne, Goldstein and Martin, 2001; Delianedis and Geske, 2001). Incidentally, Ericsson and Renault (2003) show that credit spreads encompass a liquidity premium, which is an increasing function of firm value, leverage and aggregate volatility. Therefore, stochastic volatility will help accounting for a widening of credit spreads due to an increase in the liquidity premium they encompass (Cunningham, Dixon and Hayes, 2001, regarding sovereign bonds). Finally, the stochastic aggregate volatility we obtain is the result of our stochastic functionals' combination. Thus, the flexibility offered by possible specifications of such functionals allows considering investment grade debt and part of speculative grade debt.

On the other hand, our credit pricing model is equivalent to a stochastic volatility Merton-type pricing model, which is valuable. Indeed, Kealhofer and Kurbat (2001) show that Merton's approach outperforms both Moody's credit ratings and well-known accounting ratios in predicting default. The Merton-type approach contains any information embedded in such ratings and ratios. In the same way, Phoa (2003) underlines the coherency of Merton-type structural models with observed risky debt market data. This author points out the usefulness of such models as risk management tools. In contrast, Eom, Helwege and Huang (2004) find that new simplified structural models misestimate credit risk. Therefore, our stochastic volatility framework can solve this problem by better fitting to empirical behavior of credit spreads, and then reconciling all points of view. To investigate this issue, future research should estimate the result of our model using risky debt data, and then test its performance.

Our work's significance and future implementation are of major importance for a sound assessment of credit risk. First, credit spreads and default rates are key determinants for both pricing and hedging of credit instruments along with dynamic credit portfolio management. Second, as idiosyncratic risk is diversifiable, systematic risk is more important at a portfolio level (Jarrow, Lando and Yu, 2005; Frey and McNeil, 2001; Lucas, Klaassen, Spreij and Straetmans, 2001; Giesecke and Weber, 2004). However, Goetzmann and Kumar (2001) show the existence of many under-diversified portfolios. Such portfolios are usually naively diversified and bear an important idiosyncratic risk. Consequently, credit portfolio management has to integrate idiosyncratic and systematic risk trade-off. Such a consideration is all the more important at an individual firm viewpoint.

NOTES

1. Volatility is represented by a combination of both an autoregressive process and an additional ARCH process.
2. Such rules require computing the frequency of occurrence of excessive losses (for example, observed losses that lie above the loss forecasts computed from Value-at-Risk models). The lower this frequency, the better the model performs.
3. Sharpe (1963) proposes a two-factor model where only one factor plays a role on an average basis. Analogously to the asset pricing theory, we propose a two-factor model where idiosyncratic risk is explicitly taken into account. Moreover, our general framework reduces to Sharpe's (1963) setting when $E[d\ln I_t]=(1-\beta)r$. Notice also that $E[d\ln I_t \mid F_t]=\lfloor \mu_I(t,V)-\frac{\sigma_I^2(t,V)}{2}\rfloor\, dt$.
4. Such a characterization is only valid when β is non-zero (for example, $1/\beta$ is defined). When β is zero, drift $\mu_V(t,V_t,I_t)$ reduces to $\mu_V(t,I_t)=\mu_I(t,I_t)$, and global volatility $\sigma_V(t,V_t,I_t)$ reduces to $\sigma_V(t,I_t)=\sigma_I(t,I_t)$ since $V_t=I_t$.
5. In general, stochastic variance and firm value are non-perfectly correlated. First, assume that the volatility of the systematic risk factor X is at best a deterministic function of time, the firm value's global variance is then independent of X. In this case, we have a non-perfect correlation between firm value and its variance $Corr(d\sigma_V^2(t,V_t,I_t),\ dV_t)=\rho(t,V_t,I_t)$ since $R_V(t,V_t,I_t)=0$. Second, assume that the volatility of idiosyncratic risk factor I is at best a deterministic function of time, then the firm value's global variance is independent of I. In this case, we have a perfect correlation between firm value and its variance $Corr(d\sigma_V^2(t,V_t,I_t),\ dV_t)=1$ since $R_I(t,V_t,I_t)=0$.

6. When $\mu_I(I_t) = \frac{\lambda\varepsilon}{\Omega} - \left(\frac{\Omega}{2} - \lambda\right)\ln(I_t)$, $\sigma_I(I_t) = \sqrt{\Omega \ln(I_t)}$, and $I_t > 1$; $R(t, I_t) = \beta^2\gamma^2 t^{2\alpha} + \Omega \ln(I_t)$. Then, $R(t, I_t) = \Omega \ln(I_t)$ when t is infinity, and variance follows a mean reverting square-root process $dR(t, I_t) = \lambda[\varepsilon - R(t, I_t)]dt + \Omega\sqrt{R(t, I_t)}\, dW_t^I$. However, we avoid logarithmic specifications, which require values of random variables to be above unity.
7. We assumed some specific constraints such that:

$$\int_t^T \sigma_V(s, I_s)\rho(s, I_s) d\hat{W}_s^I = 0$$

conditional on G_t. Such constraints are compatible with our following framework. Simulations were undertaken with respect to such constraints insofar as we selected the simulated paths satisfying this criterion while computing the firm value (that is, call pricing).

REFERENCES

Äijö, J. (2003) "Determinants of the Volatility Smile: A Study on the German Individual Stock Options", Working Paper, Department of Accounting and Finance, University of Vaasa, Finland.

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Ebens, H. (2001) "The Distribution of Stock Return Volatility", *Journal of Financial Economics*, 61(1): 43–76.

Aramov, D., Jostova, G. and Philipov, A. (2004) "Credit Risk Changes: Common Factors and Firm-Level Fundamentals", Working Paper, Department of Finance, Robert H. Smith School of Business, University of Maryland.

Backus, D., Foresi, S., Li, K. and Wu, L. (1997) "Accounting for Biases in Black-Scholes", Working Paper, Stern School of Business, New York University, NY.

Bangia, A., Diebold, F.X., Kronimus, A., Schlagen, C. and Schuermann, T. (2002) "Ratings Migration and the Business Cycle, with Application to Credit Portfolio Stress Testing", *Journal of Banking and Finance*, 26(2/3): 445–74.

Bekaert, G. and Wu, G. (2000) "Asymmetric Volatility and Risk in Equity Markets", *Review of Financial Studies*, 13(1): 1–42.

Black, F. and Scholes, M. (1973) "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy*, 81(3): 637–53.

Bollerslev, T. and Mikkelsen, H.O. (1999) "Long-Term Equity Anticipation Securities and Stock Market Volatility Dynamics", *Journal of Econometrics*, 92(1): 75–99.

Campbell, J.Y. and Taksler, G.B. (2003) "Equity Volatility and Corporate Bond Yields", *Journal of Finance*, 58(6): 2321–50.

Campbell, J.Y., Lettau, M., Malkiel, B.G. and Xu, Y. (2001) "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk", *Journal of Finance*, 56(1): 1–43.

Chen, R.-R. and Huang, J.-Z. (2002) "Credit Spread Bounds and Their Implications for Credit Risk Modeling", Working Paper, Finance Department, Rutgers University, New Brunswick.

Christensen, B.J. and Prabhal, N.R. (1998) "The Relation between Implied and Realized Volatility", *Journal of Financial Economics*, 50(2): 125–50.

Collin-Dufresne, P.P. and Goldstein, R.S. (2001) "Do Credit Spreads Reflect Stationary Leverage Ratios?" *Journal of Finance*, 56(5): 1929–57.

Collin-Dufresne, P.P., Goldstein, R.S. and Martin, J.S. (2001) "The Determinant of Credit Spread Changes", *Journal of Finance*, 56(6): 2177–207.

Cunningham, A., Dixon, L. and Hayes, S. (2001) "Analyzing Yield Spreads on Emerging Market Sovereign Bonds", *Financial Stability Review*, 11(December): 175–86.

Delbaen, F. and Schachermayer, W. (1996) “The Variance-Optimal Martingale Measure for Continuous processes”, *Bernoulli*, 2(1): 81–105.

Delianedis, G. and Geske, R. (2001) “The Components of Corporate Credit Spreads: Default, Recovery, Tax, Jumps, Liquidity and Market Factors”, Working Paper No. 22-01, University of California, Los Angeles.

Duque, J. and Lopes, P.P.T. (2003) “Maturity and Volatility Effects on Smiles or Dying Smiling?”, *Portuguese Economic Journal*, 2(3): 173–93.

Eberlein, E., Kallsen, J. and Kristen, J. (Winter 2002/2003) “Risk Management Based on Stochastic Volatility”, *Journal of Risk*, 5(2): 19–44.

El Karoui, N., Jeanblanc, M. and Shreve, S.E. (1998) “Robustness of the Black Scholes Formula”, *Mathematical Finance*, 8(2): 93–126.

Elton, E.J., Gruber, M.J., Agrawal, D. and Mann, C. (2001) “Explaining the Rate Spread on Corporate Bonds”, *Journal of Finance*, 56(1): 247–77.

Eom, J.H., Helwege, J. and Huang, J.-Z. (2004) “Structural Models of Corporate Bond Pricing: An Empirical Analysis”, *Review of Financial Studies*, 17(2): 499–544.

Ericsson, J. and Renault, O. (2003) “Liquidity and Credit Risk”, Working Paper No. 895, EFA Annual Conference, Glasgow.

Fama, E.F. and French, K.R. (1989) “Business Conditions and Expected Returns on Stocks and Bonds”, *Journal of Financial Economics*, 25(1): 23–49.

Föllmer, H. and Schweizer, M. (1991) “Hedging of Contingent Claims Under Incomplete Information”, in M. Davis and R. Elliott (eds), *Applied Stochastic Analysis*, pp. 389–414 (New York: Gordon & Breach).

Frey, R. and McNeil, A.J. (2001) “Modelling Dependent Defaults”, Technical Report, Federal Institute of Technology, ETH Zurich, *Journal of Banking and Finance*.

Frey, R. and Sin, C.A. (1999) “Bounds on European Option Prices Under Stochastic Volatility”, *Mathematical Finance*, 9(2): 97–116.

Gatfaoui, H. (2003) “Risk Disaggregation and Credit Risk Valuation in a Merton Framework”, *Journal of Risk Finance*, 4(3): 27–42.

Gatfaoui, H. and Radacal, F. (2002) “Le Cycle d’Affaires et le Risque de Défaut”, *Revue Finance, Special issue: Financial Mathematics*, 23(December): 45–75.

Gemmill, G. (2002) “Testing Merton’s Model for Credit Spreads on Zero-Coupon Bonds”, Working Paper, Faculty of Finance, City University Business School, London.

Giesecke, K. and Weber, S. (2004) “Cyclical Correlations, Credit Contagion, and Portfolio Losses”, *Journal of Banking & Finance*, 28(12): 3009–36.

Goetzmann, W.N. and Kumar, A. (2001) “Equity Portfolio Diversification”. NBER Working Paper No. 8686.

Gouriéroux, C., Laurent, J.P.P. and Pham, H. (1998) “Mean-Variance Hedging and Numeraire”, *Mathematical Finance*, 8(3): 179–200.

Goyal, A. and Santa-Clara, P.P. (2003) “Idiosyncratic Risk Matters!”, *Journal of Finance*, 58(3): 975–1007.

Hilliard, J.E. and Savickas, R. (2002) “On the Statistical Significance of Event Effects on Unsystematic Volatility”, *Journal of Financial Research*, 25(4): 447–62.

Hofmann, N., Platen, E. and Schweizer, M. (1992) “Option Pricing Under Incompleteness and Stochastic Volatility”, *Mathematical Finance*, 2(3): 153–87.

Hull, J., Nelken, I. and White, A. (2003) “Merton’s Model, Credit Risk and Volatility Skews”, Working Paper, University of Toronto.

Hull, J. and White, J. (1987) “The Pricing of Options on Assets with Stochastic Volatilities”, *Journal of Finance*, 42(2): 281–300.

Hwang, S. and Satchell, S.E. (2000) “Market Risk and the Concept of Fundamental Volatility: Measuring Volatility Across Asset and Derivative Markets and Testing for the Impact of Derivatives Markets on Financial Markets”, *Journal of Banking and Finance*, 24(5): 759–85.

Jäckel, P.P. (2002) *Monte Carlo Methods in Finance* (New York, NY: John Wiley & Sons).

Jarrow, R.A., Lando, D. and Yu, F. (2005) "Default Risk and Diversification: Theory and Applications", *Mathematical Finance*, 15(1): 1–26.
Karatzas, I. (1996) "Lectures On the Mathematics of Finance", CRM Monograph Series, 8, American Mathematical Society, Providence (Rhode Island).
Karatzas, I. and Shreve, S.E. (1991) *Brownian Motion and Stochastic Calculus*, 2nd edn (New York, NY: Springer-Verlag).
Karpoff, J.M. (1987) "The Relation between Price Changes and Trading Volume: A Survey", *Journal of Financial and Quantitative Analysis*, 22(1): 109–26.
Kealhofer, S. and Kurbat, M. (2001) "The Default Prediction Power of the Merton Approach Relative to Debt Ratings and Accounting Variables", Working Paper, KMV Corpp., San Francisco, California.
Koopman, S.J. and Lucas, A. (2005) "Business and Default Cycles for Credit Risk", *Journal of Applied Econometrics*, 20(2): 311–23.
Lamoureux, C.G. and Lastrapes, D.W. (1990) "Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects", *Journal of Finance*, 45(1): 221–29.
Leland, H.E. and Toft, K.B. (1996) "Optimal Capital Structure, Endogenous Bankruptcy, and the Term Structure of Credit Spreads", *Journal of Finance*, 51(3): 987–1019.
Lucas, A., Klaassen, P.P., Spreij, P.P. and Straetmans, S. (2001) "An Analytic Approach to Credit Risk of Large Corporate Bond and Loan Portfolios", *Journal of Banking and Finance*, 25(9): 1635–64.
Malkiel, B.G. and Xu, Y. (2002) "Idiosyncratic Risk and Security Returns", Working Paper, Department of Economics, Princeton University, Princeton, NJ.
Malkiel, B.G. and Xu, Y. (2003) "Investigating the Behavior of Idiosyncratic Volatility", *Journal of Business*, 76(4): 613–44.
Mele, A. and Fornari, F. (2000) *Stochastic Volatility in Financial Markets: Crossing the Bridge to Continuous Time* (New York, NY: Kluwer Academic Publishers).
Merton, R.C. (1974) "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates", *Journal of Finance*, 29(2): 449–70.
Musiela, M. and Rutkowski, M. (1998) *Martingale Methods in Financial Modelling*, 2nd edn (Berlin Heidelberg: Springer-Verlag).
Nickell, P.P., Perraudin, W. and Varotto, S. (2000) "Stability of Rating Transitions", *Journal of Banking and Finance*, 24(1/2): 203–27.
Phoa, W. (2003) "Implications of Merton Models for Corporate Bond Investors", Report, Capital Strategy Research, Los Angeles.
Psychoyios, D., Skiadopoulos, G. and Alexakis, P.P. (2003) "A Review of Stochastic Volatility Processes: Properties and Implications", *Journal of Risk Finance*, 4(3): 43–59.
Ripley, B.D. (1987) *Stochastic Simulation* (New York, NY: John Wiley & Sons).
Rubinstein, R.Y. (1981) *Simulation and the Monte Carlo Method* (New York, NY: John Wiley & Sons).
Savickas, R. (2003) "Event-Induced Volatility and Tests for Abnormal Performance", *Journal of Financial Research*, 26(2): 165–78.
Schwert, W.G. (1989) "Why Does Stock Market Volatility Change Over Time?", *Journal of Finance*, 44(4): 1115–53.
Sharpe, W.F. (1963) "A Simplified Model for Portfolio Analysis", *Management Science*, 9(January): 499–510.
Spahr, R.W., Schwebach, R.G. and Sunderman, M.A. (2002) "Original-Issue Systematic and Default Risk Pricing Efficiency of Speculative-Grade Bonds", *Journal of Risk and Insurance*, 69(4): 489–516.
Stein, R.M., Kocagil, A.E., Bohn, J. and Akhavein, J. (2003) "Systematic and Idiosyncratic Risk in Middle-Market Default Prediction: A Study of the Performance of the RiskCalc^{TM} and PFM^{TD} Models", *Special Comment, Moody's Investor Service*.
Wilson, T.C. (1998) "Portfolio Credit Risk", *FRBNY Economic Policy Review*, October: 71–82.

CHAPTER 7

A Comparative Analysis of Dependence Levels in Intensity-Based and Merton-Style Credit Risk Models

Jean-David Fermanian and Mohammed Sbai

7.1 INTRODUCTION

In finance, especially for credit portfolio modeling, basket credit derivatives (CDOs, n-th to default) pricing and hedging, the building of an accurate measure of the dependence between the underlying default events is becoming a key-challenge (see Crouhy, Galai and Mark, 2002; Koyluoglu and Hickman, 1998, for a review of the current credit risk portfolio models). This new frontier has induced a huge amount of literature for several years: Nyfeler (2000), Frey and McNeil (2001), Schönbucher and Schubert (2001), Das, Geng and Kapadia (2002), Elizalde (2003), Turnbull (2003), Yu (2003), among others.

There are mainly two usual approaches to simulate dependent default events (Schlögl, 2002, for example): in the structural framework (Merton, 1974) a firm is falling into default when its asset value falls below its debt level. In its multidimensional version, the default process of all the underlying obligors is directly deduced from the joint process of asset values. Most of the time, the increments of the asset process are assumed Gaussian. Thus, a correlation matrix allows a full description of the dependence between the default events.

In the intensity-based (or reduced-form) approach (Jarrow, Lando and Turnbull, 1997; Duffie and Singleton, 1999), we focus directly on the joint law of defaults, conditionally on some factors, without trying to explain the firm behaviors. Sometimes, such models seek to exhibit some observable variables for explaining the defaults, or consider defaults simpler as exogenous processes. They are trying to answer the following questions: "How and when do rating transitions happen", or "how do the spread curves behave", rather than "why".

Such a distinction may appear to be a bit artificial. As every duration model, Merton-style models can be rewritten in terms of intensities.[1] Moreover, when dealing with portfolios, the dependence structures obtained by both approaches are induced most of the time by some extra-random factors. Thus, most of the models that are built in practice can be considered as factor-models (Schönbucher, 2001). Nonetheless, we keep the distinction between structural and intensity models because it is now a type of common language in the credit risk arena.

The aim of this chapter is to exhibit simple intensity models that induce a sufficient amount of dependence. To be more specific, we would like that some dependence indicators cover a large scope of values. We prove the intensity-based approach is as flexible as the Merton-style one, in terms of dependence between obligors. It is just necessary to adopt the right point of view, and to specify conveniently such intensity-based models.

In sections 7.2 and 7.3, we detail both frameworks, and compare the respective loss distributions. Subsequently, some dependence indicators are provided and compared in section 7.4. In section 7.5, we extend the previous basic intensity-based model towards two directions : correlated frailty models and α-stable distributions.

7.2 MERTON-STYLE MODELS

In such approaches, a value A_i is associated with any firm i. An obligor is defaulting when its asset value falls below a barrier, generally representing its debt. Given these barrier levels and the dynamic of the asset values, we are able to draw the loss distribution for a whole portfolio. Thus, we consider a portfolio of k obligors and we set a fixed time horizon T, typically $T=1$ year. The default probability for firm $i=1,\ldots,k$ is:

$$p_i = P(A_i < D_i)$$

In this model, the correlation between default events is related to the correlation between assets values. Here, the latter correlation coefficient is equal to

$$corr_{ij} = corr(A_i,\ A_j)$$

Even if there exist many alternative models for setting the dynamic of the asset value, we will consider in this paper the usual simple one factor model:

$$A_i = \rho V + \sqrt{1-\rho^2}\varepsilon_i, \tag{7.1}$$

where V follows a standard normally distributed random variable. It may be seen as an overall macro-economic factor that influences all the firm values. ρ is a constant between −1 and 1. We will consider positive ρ only because it is the case most of the time in practice.[2] ε_i is a standard normally distributed random variable, specific to the obligor i. As usual, we assume that all the ε_i are mutually independent and independent from V.

Therefore, the firm's value is also normally distributed and

$$corr_{ij} = corr(A_i,\ A_j) = \rho^2. \tag{7.2}$$

In order to simulate the portfolio loss distribution, we follow these successive steps:

1 For any firm i, we get its mean historical default probability p_i at the horizon T, as given by the rating agencies (here Standard & Poor's).

2 We calculate the barrier $l_i = \Phi^{-1}(p_i)$ where Φ is the cumulated distribution function of a N(0, 1) (see (7.1)).

3 We generate some random variables V for the whole portfolio and ε_i for every firm. Both are N(0, 1). Then, we compare $\rho V + \sqrt{1-\rho^2}\varepsilon_i$ with l_i and record if a i's default is triggered or not.

4 We finally cumulate the losses and repeat the same procedure many times in order to get the loss distribution.

The calibration will be done on ρ. Below is an example of what we get with the following parameters:

- $\rho = \sqrt{0.2}$ (the choice promoted by Basel 2).
- A time-horizon $T = 1$ year.
- One year default probabilities given by Standard & Poor's in Table 7.1:
- A portfolio of 100 firms:[3]
  - 10 firms rated AAA
  - 20 firms rated AA

Table 7.1 Average default rates over 1981–2002

Rating	CCC	B	BB	BBB	A	AA	AAA
PD (%) (1 year)	27.87	6.20	1.38	0.37	0.05	0.01	0.00

Source: Standard & Poor's.

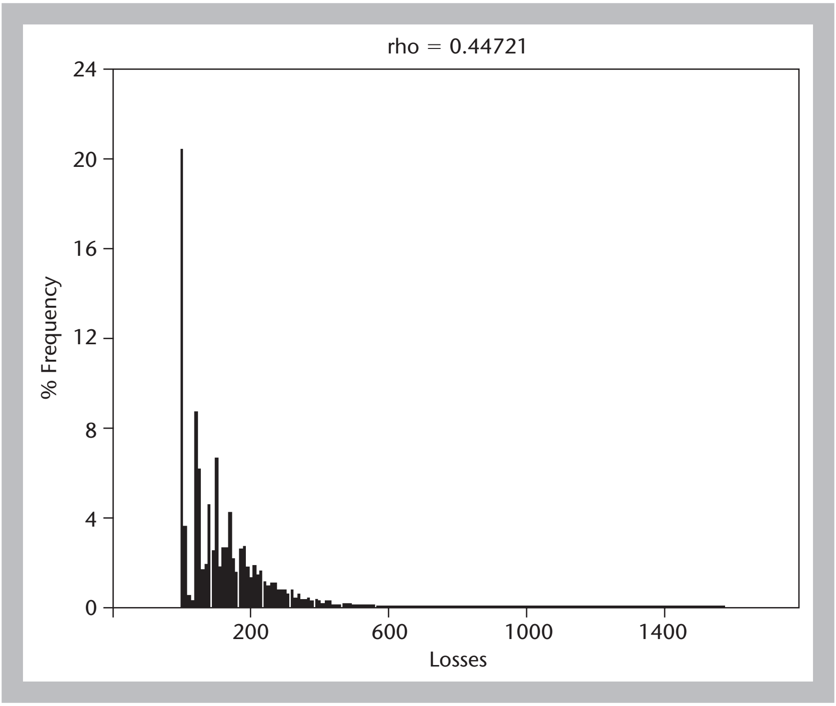


Figure 7.1 Histogram of losses in the Merton model

– 20 firms rated A

– 20 firms rated BBB

– 15 firms rated BB

– 10 firms rated B

– 5 firms rated CCC

- Constant exposure levels drawn randomly between 0 and 100.[4] Once they have been simulated, these exposure levels will be kept constant during the whole study. Their maturities are assumed infinite: when a default event is simulated, it always induces a non zero loss (whose value is the previous level associated with the defaulted counterparty). With such choices, we obtain Figure 7.1.

Table 7.2 One-year default events correlations between firms as a function of their ratings (%), with $\rho = \sqrt{0.2}$

	AAA	AA	A	BBB	BB	B	CCC
AAA	0.27	0.27	0.32	0.58	0.80	1.04	1.09
AA	0.27	0.27	0.32	0.58	0.80	1.04	1.09
A	0.32	0.32	0.38	0.69	0.96	1.27	1.35
BBB	0.58	0.58	0.69	1.33	1.94	2.70	3.06
BB	0.80	0.80	0.96	1.94	2.90	4.20	5.02
B	1.04	1.04	1.27	2.70	4.20	6.42	8.23
CCC	1.09	1.09	1.35	3.06	5.02	8.23	11.65

For $\rho = \sqrt{0.2}$, we also calculate the linear correlation between the default events for couples of firms that belong to pre-specified rating classes. The results are gathered in Table 7.2. In the Appendix we explain how we calculate such correlations. As empirically measured previously, the correlation levels we get among speculative grade firms are higher than those obtained with investment firms. They cover a range between 0.7 percent up to 11.6 percent, which is coherent with the empirical literature (de Servigny and Renault, 2002).

7.3 INTENSITY-BASED MODELS

Such models are based on a direct evaluation of the intensity processes themselves. We are reminded that the default intensity is the instantaneous arrival rate of default:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(\tau \in [t, t + \Delta t] | \tau > t)$$

denoting by τ the default time. Let f be the probability density function of τ and S its survival function. For every time t, we have obviously:

$$\lambda(t) = \frac{f(t)}{S(t)}$$

Just as the density f, the functions λ and S determine the law of τ, because

$$S(t) = \exp\left(-\int_0^t \lambda(s) ds\right)$$

The model we consider now belongs to the well-known frailty models family (Clayton and Cuzick, 1985). It has been used extensively in Survival Analysis (Hougaard, 2000). Frailty models are extensions of the Cox model

(Cox, 1972), where the (conditionally on the covariates) default intensities are multiplied by some unobservable random effects. Thus, in the basic version of frailty models, we set for every time t and every firm i:

$$\lambda_i(t, X_i, Z) = Z\lambda_0(t)\exp(\beta^T X_i), \qquad (7.3)$$

where β is a vector-valued parameter of interest. X_i is the vector of observable covariates of the firm i. They may be firm specific and/or systemic (macro-economic indices). λ_0 is the deterministic baseline hazard function. Z is a frailty, an unobservable gamma distributed random variable. We assume it is the same for every obligor.

The random variable Z can be interpreted as a synthetic macro-economic factor that has not been included into the observable covariates X_i. For the sake of simplicity, we assume that λ_0 is a constant function and that β is equal to 0 (no observable covariates). Thus, the dependence is driven by Z only. Moreover, the Z realizations are assumed constant. This constancy is clearly a strong assumption, but it is realistic when we restrict ourselves to a one or two year horizon. This is indeed the case in this section. Then:

$$\lambda_i(t) = \lambda_i = Z\lambda_{0,i} \quad \text{where } Z \text{ is following a gamma law } G(\alpha, \theta) \qquad (7.4)$$

This implies that the expectation of Z is α/θ and that its variance is α/θ^2. The default probabilities are taken from the same source as in the Merton model. We consider one year as the time unit, say T is expressed in years. Thus, λ_i can be identified with the yearly default intensity. We get the random default probability at time T as:

$$p_i(T|\lambda_i) = P(\tau \leq T|\lambda_i) = 1 - \exp(-\lambda_i T) \qquad (7.5)$$

When we take the expectation with respects to Z, we have:

$$E(1 - \exp(-T\lambda_i)) = 1 - \left(\frac{\theta}{\theta + T\lambda_{0,i}}\right)^{\alpha} = \overline{p_i}(T) \qquad (7.6)$$

This provides a first condition on the parameters (α, θ) and $\lambda_{0,i}$ since we know the mean historical probabilities $\overline{p}_i(T)$. In order to make the baseline hazard function $\lambda_{0,i}$ identifiable, we normalize the frailty variable : $E(Z) = 1$, i.e $\alpha = \theta$. In this case, Var(Z) $= 1/\alpha$. Now, the key parameter is α.

We consider the same portfolio as in the Merton model and we follow the following steps to get the loss distribution: for every time T

1 we invoke $\overline{p}_i(T)$, the mean default probability (see(7.6)) to deduce $\lambda_{0,i}$;

2 we simulate Z and deduce $\lambda_{0,i}$ for each obligor i (see (7.4));

3 we draw a uniform random variable and we compare it to $\overline{p}_i(T|\lambda_i)$ to see if a default is triggered or not; see (7.5); and finally,

4 we cumulate the losses and we repeat the same procedure many times.

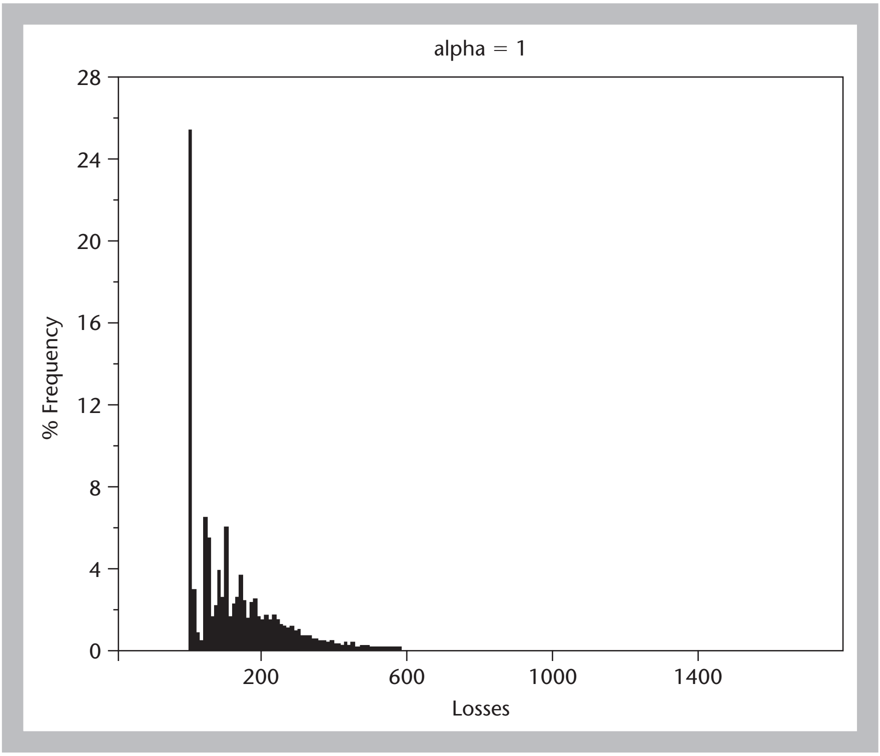


Figure 7.2 Histogram of the losses in an intensity-based model

Table 7.3 One-year default events correlations between firms with different ratings (%), with $\alpha = 1$

	AAA	AA	A	BBB	BB	B	CCC
AAA	0.03	0.03	0.04	0.10	0.20	0.42	0.78
AA	0.03	0.03	0.04	0.10	0.20	0.42	0.78
A	0.04	0.04	0.05	0.13	0.26	0.54	1.00
BBB	0.10	0.10	0.13	0.37	0.71	1.46	2.72
BB	0.2	0.2	0.26	0.71	1.36	2.81	5.25
B	0.42	0.42	0.54	1.46	2.81	5.84	11.00
CCC	0.78	0.78	1.00	2.72	5.25	11.00	21.79

For example, for $\alpha = 1$ and $T = 1$, we get the histogram of the losses in Figure 7.2. Such empirical distribution looks like the one obtained with the Merton-style model (graph 1), especially in the right tail.

Again, for $\alpha = 1$ we calculate the default events correlations between firms with different ratings: Table 7.3. We get levels that are comparable with those obtained in Table 7.2, especially for speculative grade firms. Nonetheless, the differences by rating classes seem to be even stronger in the intensity framework. In other words, it is not easy to get significant correlation levels for couple of investment grade firms.

The same tabulars have been calculated with larger time horizons $T = 5$ and $T = 20$ years. See Appendix B. The conclusions are broadly the same, in terms of comparison between Merton-style and intensity-style models. Nonetheless, it is difficult to draw any general conclusions by focusing on some particular values for ρ and α.

7.4 COMPARISONS BETWEEN SOME DEPENDENCE INDICATORS

For several years, there has been a debate in the financial literature and among practitioners to compare the advantages and the drawbacks of both the previous approaches. Some authors[5] have come to the conclusion that realistic dependence levels between obligors cannot be easily obtained with intensity models. Notably, Schönbucher (2003) argues that, under some hypotheses, the strongest possible default correlation in an intensity-based model is of the same order of magnitude as the default probabilities. We briefly detail his technical argument.

Consider two firms A and B. For a fixed time horizon T, let

- p_A and p_B be the two individual default probabilities of A and B;
- λ_A and λ_B their random default intensities. For every realization ω, the functions $\lambda_A(\omega)$ and $\lambda_B(\omega)$ are assumed constant between 0 and T for the sake of simplicity;
- p_{AB} their joint default probability;
- ρ_{AB} the correlation coefficient between both default events.

By simple calculations, we obtain:

$$\begin{aligned}
p_{AB} &= \mathbb{E}(1_{\{A\}}1_{\{B\}}) \\
&= \mathbb{E}(\mathbb{E}(1_{\{A\}}1_{\{B\}}|\lambda)) \\
&= \mathbb{E}\left(1 - \exp\left(-\int_0^T \lambda_A(s)ds\right)\right)\left(1 - \exp\left(-\int_0^T \lambda_B(s)ds)\right)\right) \\
&= 1 - (1 - p_A) - (1 - p_B) + \mathbb{E}\left(\exp\left(-\int_0^T \lambda_A(s) + \lambda_B(s)ds\right)\right) \\
&= p_A + p_B + \mathbb{E}\left(\exp\left(-\int_0^T \lambda_A(s) + \lambda_B(s)ds\right)\right) - 1
\end{aligned}$$

If both intensities are perfectly correlated: $\lambda_A = \lambda_B = \lambda$, then:

$$p_{AB} = 2p + \mathbb{E}\left(\exp\left(-2\int_0^T \lambda(s)ds\right)\right) - 1, \quad \text{where } p = p_A = p_B$$

The correlation between the two default events is then:

$$\rho_{AB} \stackrel{def}{=} \frac{p_{AB} - p_A p_B}{\sqrt{p_A(1-p_A)p_B(1-p_B)}} \tag{7.7}$$

$$= \frac{2p + \mathbb{E}(\exp(-2\int_0^T \lambda ds)) - 1 - p^2}{p(1-p)}$$

$$= \frac{\mathbb{E}(\exp(-2\int_0^T \lambda ds)) - (1-p)^2}{p(1-p)}$$

$$= \frac{\mathbb{V}ar(\exp(-\int_0^T \lambda ds))}{p(1-p)} \tag{7.8}$$

If we assume that the variance of the survival probability is at most of order p^2, then the correlation is of order p. Nonetheless, we argue that this is far from being satisfied usually.

To justify his assumption, Schönbucher (2003) suggested a normally distributed integrated intensity, for which we assume that the integrated hazard function between 0 and T is following a normal law N (μ, σ^2).

Note that such an assumption does not generate a "true" intensity process because some values of the integrated intensity may be negative. Nonetheless, forgetting such a detail, we get:

$$\mathbb{E}\left(\exp\left(-\int_0^T \lambda(s)ds\right)\right) = 1 - p = \exp\left(-\mu + \frac{1}{2}\sigma^2\right)$$

$$\mathbb{E}\left(\exp\left(-2\int_0^T \lambda(s)ds\right)\right) = \exp(-2\mu + 2\sigma^2)$$

and we deduce:

$$\rho = (e^{-2\mu+2\sigma^2} - e^{-2\mu+\sigma^2})/(p-p^2) \approx (1-p)(e^{\sigma^2} - 1)/p \tag{7.9}$$

If $\sigma \approx \lambda T$, we get that ρ and p are of the same order with this normal intensities specification. Clearly, it is a very crude approximation. A more careful approximation provides:

$$\exp(\sigma^2) - 1 \approx \sigma^2 \approx 2(\mu - p),$$

because $1 - p = \exp(-\mu + \sigma^2/2) \approx 1 - \mu + \sigma^2/2$. Thus, we get $\rho \approx 2(\mu - p)/p$, but we have no ideas (a priori) concerning the size of the latter ratio. To conclude, it seems that no strong argument has been done to conclude that the correlation levels ρ induced by intensity-based models are most of the time insufficient in practice.

In our previous setting, it would be more realistic to assume the random intensities follow the usual log-normal assumption:

$$\lambda = \lambda_0 \exp(-\sigma^2/2 + \sigma\varepsilon), \quad \varepsilon \sim \mathrm{N}(0,1)$$

In this case, the intensities are positive and they can be dealt as usual market factors in pricing formulas. Thus, we can evaluate the variance of the survival probability in equation (7.8). Remind that, if a random variable X follows a lognormal law, say $X = \exp(Z)$ with Z following a N(0,1), then:

$$\mathbb{E}(\exp(-tX)) = \sum_{p=0}^{\infty} \frac{(-t)^p}{p!} \exp\left(p\mu + \frac{p^2\sigma^2}{2}\right)$$

Here, λ is assumed constant between 0 and T. Thus:

$$\mathbb{E}\left(\exp\left(-t\int_0^T \lambda\right)\right) = \sum_{p=0}^{\infty} \frac{(-t)^p}{p!} (\lambda_0 T)^p \exp\left(-\frac{p\sigma^2}{2} + \frac{p^2\sigma^2}{2}\right).$$

By a limited expansion, we get:

$$\begin{aligned} Var\left(\exp\left(-\int_0^T \lambda\right)\right) &= E\left[\exp\left(-2\int_0^T \lambda\right)\right] - E\left[\exp\left(-\int_0^T \lambda\right)\right]^2 \\ &\approx (\lambda_0 T)^2 \left(\exp(\sigma^2 - 1)\right) \end{aligned}$$

Thus, from equation (7.8), the correlation level between the two default times of the obligors A and B is approximately:

$$\rho_{AB} \approx p(\exp(\sigma^2) - 1)$$

Note that the coefficient σ has not the same meaning as in (7.9). Moreover, $Var(\lambda) = \lambda_0^2\,(\exp(\sigma^2) - 1)$. It is reasonable to assume that the standard deviation of the variations of λ is two or three times λ_0 (see Figure 7.3).

Thus, $\exp(\sigma^2) - 1$ is easily 4, 9 or more. For instance, if the default rate of the obligors is 1 percent between 0 and T, then the correlation level can reasonably be of the order 5 percent or 10 percent. Higher correlation levels can even be reached when assuming more volatility for the random intensities. In our current framework,[6] we can remind the following useful rule-of-thumb: when the standard deviation of the changes in random intensities is q times the mean level of these intensities, then the correlation levels are of order q^2 times the mean probability of default.

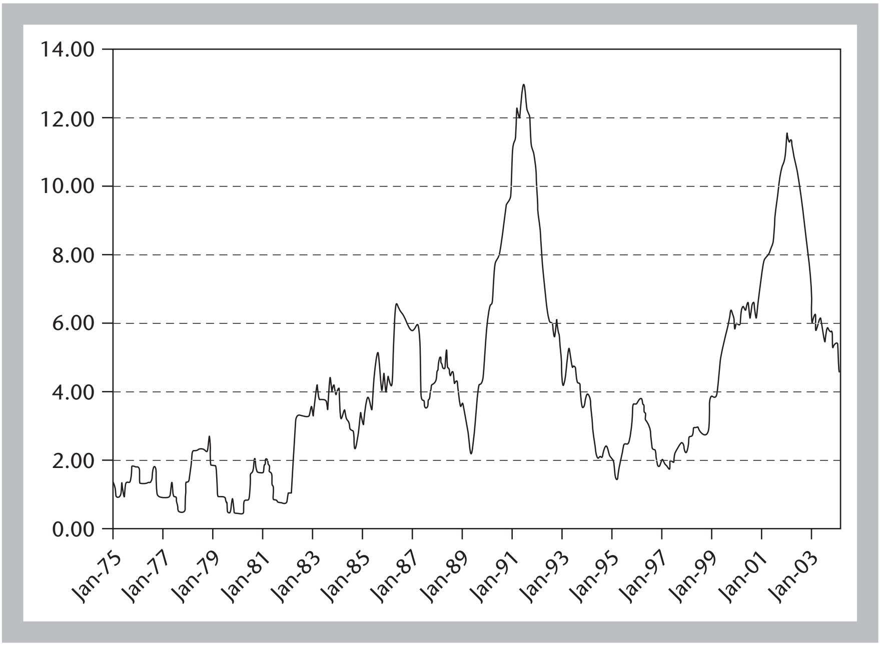


Figure 7.3 Monthly default rates, US bonds speculative grade, trailing 12 months, in percents
Source: Moody's

Table 7.4 Features of the loss distribution for different ρ values (Merton model), $T = 1$ year

ρ	**0.01**	**0.1**	**0.3**	**0.4**	**0.6**	**0.7**	**0.9**	**0.95**
Quantile of order 99%	320	328	413	473	679	876	1163	1297
E(losses | losses > q99%)	356	367	482	571	858	1266	1615	1945
Skewness	0.64	0.70	1.15	1.50	2.47	4.13	4.33	5.47
K kurtosis	3.19	3.27	4.80	6.25	13.49	35.67	29.82	49.76
Average correlation (%)	10^{-4}	0.04	0.46	0.94	3.31	6.06	20.82	28.93

We led many simulations for different values of the parameters ρ and α. Tables 7.4 and 7.5 summarize the results we obtained. We took the same default probabilities and the same exposure in the two cases in order to have the same mean distribution.

We note that the dependence indicators between default events take some values of the same order of magnitude in the two cases. Empirically, default event correlations are varying from 0 percent to 30 percent for the Merton model, and from 0 percent to 20 percent for the reduced-form model. For some "reasonable" ρ and α levels ($\rho = 0.4$ and $\alpha = 2$, for instance), the

Table 7.5 Features of the loss distribution for different α values (intensity-based model), $T = 1$ year

$\mathrm{Var}(Z) = 1/\alpha$	0.01	0.1	0.5	2	5	10	50	100
Quantile of order 99%	331	350	414	592	783	946	1278	1401
E(losses | losses > q99%)	368	392	477	685	912	1112	1638	1803
Skewness	0.69	0.79	1.09	1.60	2.03	2.46	3.73	4.06
k kurtosis	3.36	3.54	4.22	5.74	7.50	9.87	19.84	23.50
Average correlation (%)	0.01	0.08	0.39	1.37	2.80	4.46	10.78	14.56

sizes of the dependence indicators are the same. These levels are consistent with those obtained by de Servigny and Renault (2002): the latter authors report intra industries empirical correlation levels between one-year default events less than 10 percent, with typical levels around 2–3 percent for the speculative grade firms.

We note that the values of α considered in Table 7.5 are not unrealistic: they correspond to a standard deviation of the frailty variable Z varying from 0.1 to 10. Historically, important variations of default rates from one year to another have been met: see Figure 7.3. For instance, the mean default rate for US speculative grade bonds was more than 12 percent at the mid-year 1991, and fell below 2 percent in 1995.[7]

We have calculated the same indicators for $T = 10$ years: see Appendix B. The Merton model seems to generate relatively more dependence in this case, especially under some extreme conditions (small or large ρ).

7.5 EXTENSIONS OF THE BASIC INTENSITY-BASED MODEL

7.5.1 A multi-factor model

The main idea here is to introduce an additional idiosyncratic unobservable random variable that summarizes the effect of an unobservable micro-economic factor.[8] We keep the same notations as in the first intensity model. We choose the correlated frailty model framework (Yashin and Iachine, 1995) whose asymptotic theory has been studied in Parner (1998). Such models allow taking into account simultaneously systematic and idiosyncratic random effects. In this case, we assume that

$$\lambda_i(t, X_i, Z) = (Z_0 + Z_i)\lambda_0 \exp(\beta^T X_i) \tag{7.10}$$

where Z_0 is an unobservable systemic gamma random variable, and Z_i is an unobservable gamma random variable that is specific to the obligor i.

The random variable Z_i's are mutually independent and Z_0 is independent from all the Z_i. The simulation method is almost the same as in the

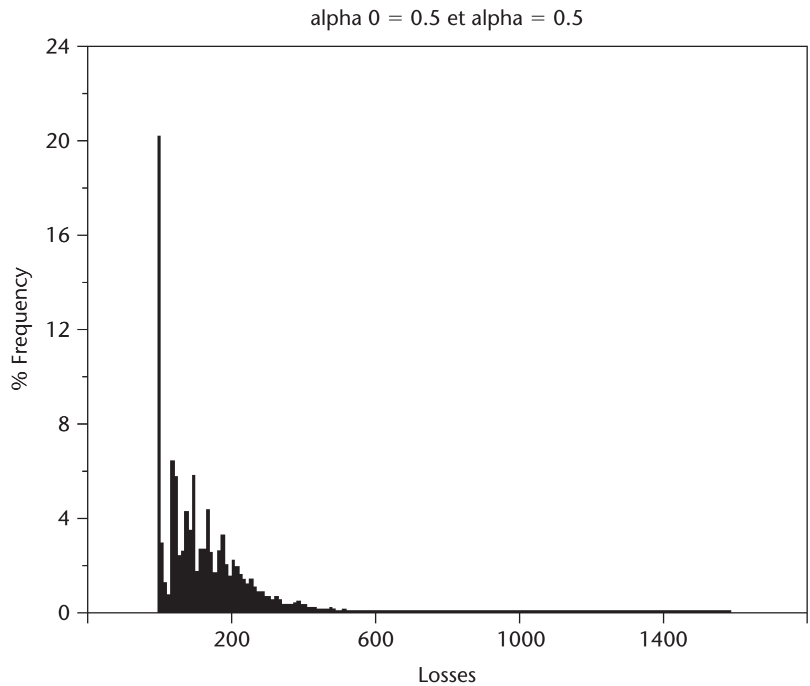


Figure 7.4 Histogram of the losses in the multi-factor intensity-based model ($T = 1$ year)

first model. We just have to draw the realizations of additional gamma random variables (one for each obligor). In practice, there are now two free parameters α_0 and α_i, related to Z_0 and Z_i respectively. This may cause some estimation complications, even if the log-likelihood of the observations can be written in closed form (Parner, 1998). In Figure 7.4, we draw the histogram of the losses obtained with model (7.10). Since we impose that the expectation of the global frailty component $Z_0 + Z_i$ equals one, we draw $Z_0 \sim \mathcal{G}(\alpha_0, \alpha_0 + \alpha)$ and $Z_i \sim$ G(α, $\alpha_0 + \alpha$). We have chosen the parameter values $\alpha_0 = 0.5$ and $\alpha = 0.5$ for every i in Figure 7.4.

In this case

$$Var(Z_0 + Z_i) = \alpha_0/(\alpha_0 + \alpha)^2 + \alpha/(\alpha_0 + \alpha)^2 = 1$$

and the correlated frailty $Z_0 + Z_i$ has the same two first moments as in Figure 7.2. The loss distributions seem to be very similar. At first glance, the introduction of specific components does not lessen too much the dependence between defaults.[9]

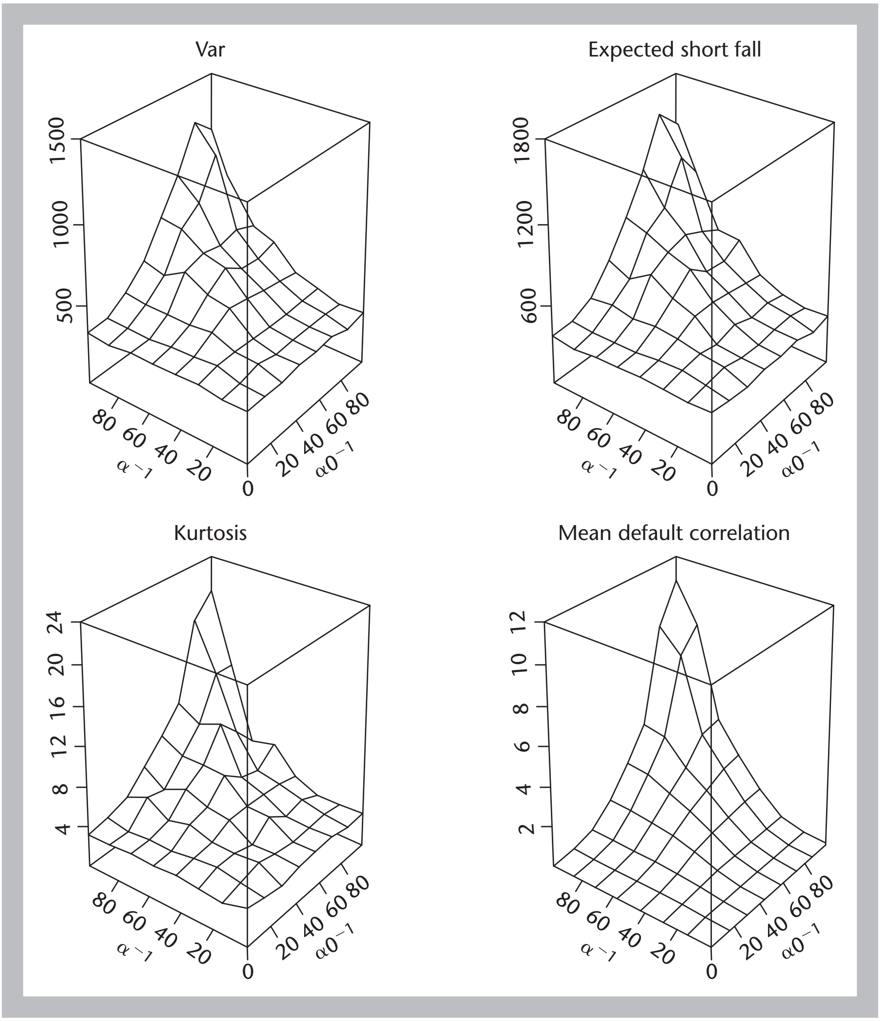


Figure 7.5 Combined effect of the parameters α_0 and α in the multi-factor intensity model

We lead many simulations with different values of the parameters α_0 and α_i in order to study their combined effects on the loss distribution (see Figure 7.5).

The variance of $Z_0 + Z_i$ varies from 0.005 to 50 when (α_0, α) varies from 0.01 to 100, which seems to be reasonable. The levels of our dependence indicators seem to be in line with those obtained in section 7.3. Note that we lose some dependence when the relative importance between Z_0 and Z_i is balanced. This is due to a diversification effect inside both components of the frailty factors. Globally, adding an idiosyncratic frailty allow more flexibility in the model, without losing the ability to reach realistic dependence levels.

Actually, the ratio $r = \alpha/\alpha_0$ provides a good measure of the dependence we obtain: the larger r, the higher the dependence indicators.

7.5.2 α-stable distributions

Properties of the family

Stable distributions allow building a rich class of probability distributions. They induce highly skewed and heavy tails features and have many interesting mathematical properties: see the survey of Samorodnitsky and Taqqu (1994), Hougaard (1986), or Mittnik and Rachev (1999) and Carr and Wu (2002) for financial applications. However, the lack of closed-form formulas for their densities and their cumulative distribution functions, despite a few exceptions, has been a major drawback that has limited their use by practitioners. To correct the ideas, we recall some basic theoretical results concerning such distributions.

Definition 1 A random variable X is said to be α-stable if for any X_1 and X_2, some independent copies of X, and for any positive numbers c_1 and c_2, there exist $c \in \mathbb{R}^+$ and $d \in \mathbb{R}$ such that:

$$cX + d \stackrel{d}{=} c_1 X_1 + c_2 X_2$$

If $d = 0$, X is said to be strictly stable.

There are other equivalent definitions of α-stable distributions (see Nolan, 2004, for a more detailed presentation of this distribution family) and we are going to invoke the following one because it is much more tractable:

Definition 2 A random variable X is said to be α-stable if its characteristic function takes the form:

$$\Phi_X(t) \stackrel{def}{=} \mathbb{E}(e^{itX}) = \begin{cases} \exp(-\gamma^\alpha |t|^\alpha (1 - i\beta \tan\left(\frac{\pi\alpha}{2}\right) sign(t)) + i\delta t) & \text{if } \alpha \neq 1 \\ \exp(-\gamma |t| (1 + i\beta \frac{2}{\pi} sign(t) \ln(|t|)) + i\delta t) & \text{if } \alpha = 1 \end{cases} \tag{7.11}$$

where $\alpha \in [0, 2], \beta \in [-1, 1], \gamma \geq 0$ and $\delta \in \mathbb{R}$.

This definition shows that an α-stable distribution generally requires four parameters as inputs:

- α, the index of stability. It is related to the tail behavior of the distribution. The smaller α, the stronger the leptokurtic feature of the distribution.
- β, the skewness parameter. If $\beta = 0$ then the distribution is symmetrical. If $\beta > 0$ then it is right skewed. Otherwise, it is left skewed.
- γ, the scale parameter.

- δ, the location parameter (When $\alpha > 1$, it measures the mean of the distribution).

There are multiple parameterizations for α-stable laws which may lead to some confusion. We keep the previous one, and we denote the α-stable distribution by $S(\alpha, \beta, \gamma, \delta)$ and its probability distribution function by f.

Definition 3 The support of an α-stable distribution is:

$$support\ (f(x)) = \begin{cases} [\delta, +\infty] & if\ \alpha < 1 \quad and \quad \beta = 1 \\ [-\infty, \delta] & if\ \alpha < 1 \quad and \quad \beta = -1 \\ \mathbb{R} & otherwise. \end{cases} \tag{7.12}$$

Because of the presence of heavy tails, all moments do not exist. Actually, we have:

Definition 4 Let $X \sim S(\alpha, \beta, \gamma, \delta)$.

$$\mathbb{E}(|X|^r) < +\infty \quad \text{if and only if } 0 < r < \alpha$$

As far as we are concerned, for example, within the framework of frailty models the Laplace transforms are key tools.

Definition 5 Let $X \sim S(\alpha, \beta, \gamma, \delta)$. Its Laplace transform is defined if and only if $\beta = 1$, in which case it equals:

$$L_X(t) \equiv E\left(e^{-tX}\right) = \exp\left(-t\delta - t^\alpha \gamma^\alpha \sec\left(\frac{\pi\alpha}{2}\right)\right), \quad t \geq 0 \tag{7.13}$$

by denoting $\sec(x) = 1/\cos(x)$. We will also need the following property:

Definition 6 Let $X \sim S(\alpha, \beta, \gamma, \delta)$ where $\alpha \neq 1$. Then for all $\alpha \neq 0$ and $b \in \mathbb{R}$ we have $aX + b \sim S(\alpha, \text{sign}(a)\,\beta, |a|\gamma, a\delta + b)$.

In particular, if $Z \sim S(\alpha, \beta, 1, 0)$ and

$$X = \begin{cases} \gamma Z + \delta & if \alpha \neq 1 \\ \gamma Z + (\delta + \frac{2\beta}{\pi}\gamma \ln(\gamma)) & if \alpha = 1 \end{cases}$$

then $X \sim S(\alpha, \beta, \gamma, \delta)$. We will simply note $S(\alpha, \beta)$ instead of $S(\alpha, \beta, 1, 0)$. Thus, by some linear transformations, we get all the α-stable laws starting from the family $S(\alpha, \beta)$.

7.5.3 Simulation of an α-stable distribution

As mentioned earlier, α-stable density functions do not admit closed forms. The usual method to obtain these functions is to inverse their characteristic functions $f(x) = \frac{1}{2\pi}\int \exp(-itx)\Phi_X(t)\,dt$. Except in a few cases,[10] the estimation of the latter expression is difficult, and will rather use the method in Chambers, Mallows and Stuck (1996). Let W be a random variable exponentially distributed with parameter $\lambda = 1$, and U a random variable uniformly distributed on $[-\frac{\pi}{2}, \frac{\pi}{2}]$ and let $\xi = \arctan(\beta \tan(\pi\alpha/2))/\alpha$ and the random variable:

$$Z = \begin{cases} \dfrac{\sin(\alpha(\xi + U))}{\sqrt[\alpha]{\cos(\alpha\xi)\cos(U)}} \left(\dfrac{\cos(\alpha\xi + (\alpha - 1)U)}{W} \right)^{\dfrac{1-\alpha}{\alpha}} & if \alpha \neq 1 \\ \dfrac{2}{\pi} \left(\left(\dfrac{\pi}{2} + \beta U \right) \tan(U) - \beta \ln \left(\dfrac{\frac{\pi}{2} W \cos U}{\dfrac{\pi}{2} + \beta U} \right) \right) & if \alpha = 1 \end{cases} \tag{7.14}$$

Then $Z \sim S(\alpha, \beta)$. To get $S(\alpha, \beta, \gamma, \delta)$, we invoke the linear transform of Definition 6.

7.5.4 α-stable intensity-based model

To simulate more heavy tailed random intensities, we are going to replace the gamma frailty random variable in (7.3) by an α-stable distributed frailty. As an intensity process is always positive and according to (7.12), we impose that $\alpha < 1$, $\beta = 1$ and $\delta = 0$ in order that the support of the frailty is $[0, +\infty]$. We keep the same simple specification as in our first intensity model: for every obligor i and every time t,

$$\lambda_i(t) = \lambda_i = Z\lambda_{0,i}$$

Therefore, $Z \sim S(\alpha, 1, \gamma, 0)$ where $\alpha \in [0, 1]$. Indeed, as the frailty variable has a multiplicative effect on the intensity, its baseline hazard function plays the role of a scale parameter. Thus, the parameter γ is unnecessary. In fact, we identify $\lambda_{0,i}$ by using the Laplace transform of the α-stable distribution (7.13), which leads to the one-year default probability:

$$\overline{p}_i = 1 - \exp\left(-\gamma^\alpha \sec\left(\frac{\pi\alpha}{2} \right) \gamma^\alpha \lambda_{0,i}^\alpha \right)$$

This implies:

$$\lambda_0 = \frac{1}{\gamma} \left(\frac{\ln\left(\frac{1}{1-\overline{p}_i} \right)}{\sec\left(\frac{\pi\alpha}{2} \right)} \right)^{\frac{1}{\alpha}}$$

Hence

$$\begin{aligned} \lambda &\overset{d}{=} \lambda_0 Z \\ &= \frac{1}{\gamma} \left(\frac{\ln\left(\frac{1}{1-\overline{p}_i} \right)}{\sec\left(\frac{\pi\alpha}{2} \right)} \right)^{\frac{1}{\alpha}} \gamma S(\alpha, 1) \qquad (7.15) \\ &= \left(\frac{\ln\left(\frac{1}{1-\overline{p}_i} \right)}{\sec\left(\frac{\pi\alpha}{2} \right)} \right)^{\frac{1}{\alpha}} S(\alpha, 1) \end{aligned}$$

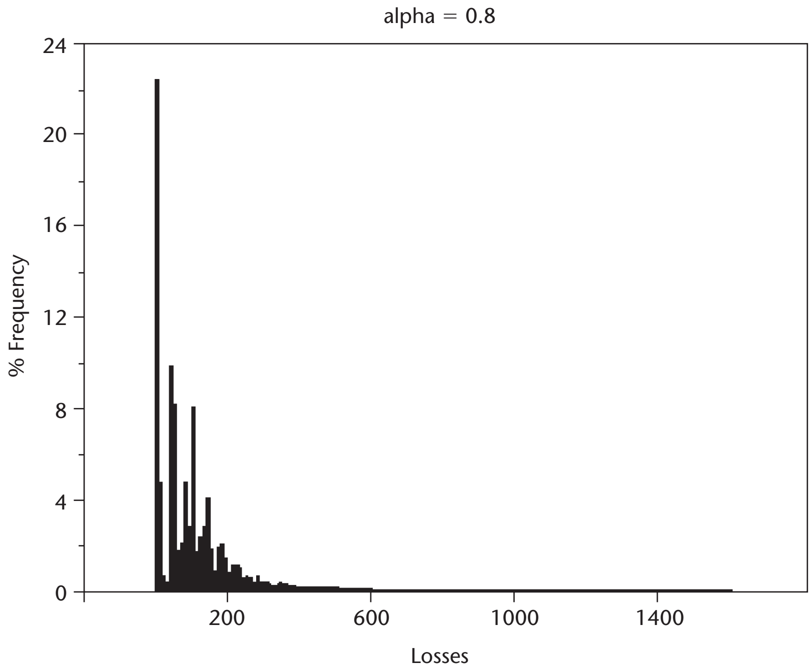


Figure 7.6 Histogram of the losses in the α–stable intensity based model

Obviously, the random intensities, and so the whole model, depend on α. In order to simulate the loss distribution, we draw a random variable $Z \sim S(\alpha,1)$ (see (7.14)) and we deduce λ from (7.15). We then follow the same steps as with the other models. For example, setting $\alpha = 0.8$ we obtain the histogram of portfolio[11] losses in Figure 7.6.

As expected, it is now easier to get large dependence levels between individual defaults inside the portfolio. Actually, the correlation between default events is even stronger than in our previous Merton-type model. Thus, α-stable frailties are a simple way to induce a strongly dependent credit-risky portfolio.

Table 7.6 presents the characteristics of the distribution for different values of the parameter α. The smaller the α, the larger the dependence between default events. The dependence indicators we get with α-stable laws are stronger than previously. Thus, it is a relatively simple way to generate highly dependent defaults, without modifying the intensity-based framework. Surprisingly, the kurtosis is increasing when the VaR and Expected Shortfall are decreasing. This can be explained by a type of degeneracy of the loss distributions: when α is very small, the losses are concentrated near the origin and very far towards the right. The implicit reference to the

Table 7.6 α-stable intensity-based model, $T = 1$ year

α	0.1	0.3	0.5	0.7	0.8	0.9	0.95
Quantile of order 99%	1588	1416	1170	1108	990	663	505
E(losses | losses > q99%)	2048	1905	1766	1654	1510	1066	844
Skewness	5.33	5.22	5.51	6.81	6.98	6.74	7.07
k kurtosis	45.59	46.10	51.11	87.51	90.26	95.60	115.48
Average correlation (%)	44.33	40.19	33.72	23.86	17.26	9.35	4.86

Gaussian distribution (when dealing with kurtosis) has no more sense in such situations.

7.6 CONCLUSION

We find some evidence that realistic and comparable dependence levels can be obtained by both intensity-style models and Merton-style models. With long time horizons, the latter approach gains a relative advantage, but the former can be strengthened by some extensions towards α-stable frailty models. Thus, the issue is not really to choose between both approaches but rather to specify conveniently a model, an intensity-based one or a Merton-style one. In practice, it is important to solve the following issues:

- What is the correlation scope that the model needs to cover?
- Observable and/or unobservable exogenous factors?
- Which distribution for such factors?
- Constant or time dependent frailties? If yes, which process is best suited?

Moreover, one of the main practical issues concerns the estimation of the key dependence parameters, typically ρ and α in our previous frameworks. Such an issue may become a hurdle for the implementation of such models. For instance, clean estimations of the simplistic frailty model (7.3) are far from trivial (see Andersen, Gill, Borgan and Keiding, 1997, for the theory, and Metayer, 2004, for a financial application). And, even more, the introduction of dynamic frailties[12] induces likelihoods without any closed form, which imposes some delicate numerical optimization procedures (simulated maximum likelihood, EM algorithm, and so for).

APPENDIX A: CALCULATION OF CORRELATION BETWEEN DEFAULT EVENTS

Our goal is the calculation of the correlation between default events and between the dates $t = 0$ and $t = T$, controlling eventually by the rating categories. Technically speaking, it is equivalent to the calculation of joint default probabilities.

To calculate the joint default probability of two obligors, say A and BB, with different ratings in the intensity-based model, we note that:

$$
\begin{aligned}
P(\tau_A < T, \tau_{BB} < T) &= E\left[E[1(\tau_A < T)1(\tau_{BB} < T)\,|\lambda]\right] \\
&= \mathrm{E}\left[\left(1-\exp\left(-\int_0^T \lambda_A\right)\right)\cdot\left(1-\exp\left(-\int_0^T \lambda_{BB}\right)\right)\right] \\
&= 1-(1-\mathrm{p_A})-(1-p_{BB}) \\
&\quad +E\left[\exp(-\int_0^T(\lambda_A+\lambda_{BB}))\right] \\
&= \mathrm{p_A}+p_{BB}-1+E\left[\exp\left(-T\left(\lambda_A^0 Z+\lambda_{BB}^0 Z\right)\right)\right] \\
&= \mathrm{p_A}+p_{BB}-1+L_{G(\alpha,\alpha)}\left(T(\lambda_A^0+\lambda_{BB}^0)Z\right) \\
&= \mathrm{p_A}+p_{BB}-1+\left(\frac{\alpha}{\alpha+T(\lambda_A^0+\lambda_{BB}^0)}\right)^{\alpha}
\end{aligned}
$$

where $\mathcal{L}_{\mathcal{G}(\alpha,\theta)}(t)$ is the Laplace transform of a gamma-distributed random variable with parameter (α, θ).

From (7.9), we deduce the default correlation coefficient between default events for firms that are rated A and BB. Finally, to get an average correlation, we calculate a mean over all the possible couples of different firms. To be specific, we calculate:

$$
\rho_m = \frac{1}{\sum_{i,j=1}^{7} n_i n_j} \sum_{i,j=1}^{7} n_i n_j \rho_{i,j}
$$

where n_i is the number of firms of rating i, and $\rho_{i,j}$ is the correlation coefficient obtained as previously explained.

To calculate the joint default probability of two obligors with different ratings in the Merton-style model, for example A and BB, we use the usual technique. According to (7.1) and (7.2) we have:

$$
\begin{pmatrix} A_A \\ A_{BB} \end{pmatrix} \sim \mathrm{N}\left(0, \begin{bmatrix} 1 & \rho^2 \\ \rho^2 & 1 \end{bmatrix}\right),
$$

which provides:

$$
\begin{aligned}
&P(\tau_A < 1\text{year}, \tau_{\mathrm{BB}} < 1\text{year}) \\
&= \frac{1}{2\pi\sqrt{1-\rho^4}} \int_{-\infty}^{D_A}\int_{-\infty}^{D_{BB}} \exp\left(-\frac{x^2+y^2-2\rho^2 xy}{2(1-\rho^4)}\right)\mathrm{dx\,dy}.
\end{aligned}
$$

We estimate numerically the latter double integral and deduce the average correlation between default events for every couple of ratings, as we made in the intensity-based model. The average correlation level is obtained by weighting conveniently such quantities.

APPENDIX B: EXTENSIONS OF THE RESULTS TO LARGE TIME HORIZONS

We use the same method as in sections 7.2 and 7.3. We choose the following default probabilities:

Average default rates over 1981–2002							
Rating	**CCC**	**B**	**BB**	**BBB**	**A**	**AA**	**AAA**
PD (%) (5 years)	61.35	33.02	14.45	3.83	0.75	0.27	0,11
PD (%) (20 years)	73.94	67.68	48.09	19.48	6.78	4.38	1.13

Source: Standard & Poor's Credit Pro.

We obtain default event correlations for the time horizons $T=5$ years and $T=20$ years (Tables 7.7, 7.8, 7.10 and 7.11), and the usual dependence indicators with $T=10$ years (Tables 7.9 and 7.12). In the latter case, particularly, the scope of values obtained in both cases is similar.

Note that when the time horizon T is increasing, it is surely questionable to assume the same values ρ and α as when $T=1$ year apply. Indeed, in the Merton-style models there is some empirical evidence that the asset correlations depend on T (see the discussion in de Servigny and Renault, 2002, for example).

Moreover, since we assumed the random default intensities λ_i are constant functions between 0 and T, their (random) levels should be less and less variable when T is increasing.[13] It should be more relevant to simulate an annual process (Z_t) for the frailty, but this does not belong in our simple framework. Thus, a realistic range of α-values is

Table 7.7 5-years default events correlations in the Merton model, with $\rho=\sqrt{0.2}$ (%)

	AAA	AA	A	BBB	BB	B	CCC
AAA	0.63	0.82	1.10	1.59	1.89	1.85	1.51
AA	0.82	1.10	1.48	2.21	2.68	2.67	2.22
A	1.10	1.48	2.04	3.12	3.90	3.98	3.40
BBB	1.59	2.21	3.12	5.05	6.67	7.11	6.37
BB	1.89	2.68	3.90	6.67	9.32	10.43	9.84
B	1.85	2.67	3.98	7.12	10.43	12.15	12.01
CCC	1.51	2.22	3.40	6.37	9.84	12.01	12.53

Table 7.8 20-years default events correlations in the Merton model, with $\rho=\sqrt{0.2}$ (%).

	AAA	AA	A	BBB	BB	B	CCC
AAA	2.60	3.68	4.02	4.64	4.41	3.78	3.50
AA	3.68	5.41	6.01	7.22	7.20	6.35	5.92
A	4.02	6.01	6.70	8.17	8.30	7.40	6.93
BBB	4.64	7.22	8.17	10.41	11.16	10.27	9.73
BB	4.41	7.20	8.30	11.16	12.81	12.28	11.81
B	3.78	6.35	7.40	10.27	12.28	12.09	11.74
CCC	3.50	5.92	6.93	9.73	11.81	11.74	11.43

Table 7.9 10-years Merton model

ρ	**0.01**	**0.1**	**0.3**	**0.4**	**0.6**	**0.7**	**0.9**	**0.95**
Q_{99}% quantile of order 99%	1,193	1,254	1,543	1,827	2,357	2,664	3,537	3,757
E(losses | losses > *q99%)*	1,253	1,339	1,663	2,020	2,704	3,089	4,145	4,426
skewness	0.15	0.22	0.46	0.75	1.05	1.20	1.53	1.5
k kurtosis	2.95	3.08	3.15	3.68	4.50	4.78	5.87	6.42
average correlation (%)	$10^{''3}$	0.24	2.26	4.19	10.62	15.62	31.60	37.72

Table 7.10 5-years default events correlations in the intensity model, with $\alpha = 1$ (%)

	AAA	**AA**	**A**	**BBB**	**BB**	**B**	**CCC**
AAA	0.04	0.05	0.08	0.16	0.30	0.36	0.31
AA	0.05	0.08	0.12	0.24	0.44	0.55	0.53
A	0.08	0.12	0.18	0.39	0.71	0.93	0.94
BBB	0.16	0.24	0.39	0.84	1.57	2.14	2.22
BB	0.30	0.44	0.71	1.57	3.02	4.19	4.31
B	0.36	0.55	0.93	2.14	4.19	6.24	7.34
CCC	0.31	0.53	0.94	2.22	4.31	7.34	13.01

Table 7.11 20-years default events correlations in the intensity model, with $\alpha = 1$ (%)

	AAA	**AA**	**A**	**BBB**	**BB**	**B**	**CCC**
AAA	0.13	0.07	0.12	0.23	0.29	0.33	0.26
AA	0.07	0.49	0.44	0.57	1.10	0.93	0.27
A	0.12	0.44	0.48	0.67	1.07	1.02	0.47
BBB	0.23	0.57	0.67	1.06	1.59	1.65	1.11
BB	0.29	1.10	1.07	1.59	3.03	2.99	2.04
B	0.33	0.93	1.02	1.65	2.99	3.46	3.41
CCC	0.26	0.27	0.47	1.11	2.04	3.41	8.39

Table 7.12 10-years intensity-based model

Var(Z) = 1/α	**0.01**	**0.1**	**0.5**	**2**	**5**	**10**	**50**	**100**
quantile of order 99%	1190	1209	1283	1508	1766	2015	2722	3026
E(losses | losses > *q*99%)	1256	1275	1364	1621	1897	2178	3003	3333
skewness	0.16	0.14	0.15	0.31	0.41	0.52	0.87	0.99
k kurtosis	2.99	2.98	2.96	2.89	2.72	2.63	2.87	3.16
average correlation (%)	0.01	0.09	0.44	1.68	3.82	6.72	18.24	24.17

becoming thinner and thinner when the time horizon is growing. That is why a straight comparison between Tables 7.8 and 7.11 particularly is not fully satisfying, because $\alpha = 1$ is probably too high is such a case.

NOTES

1. See Luciano (2004) for a discussion in the finance field. In a more general context, there is no issue to rewrite the marginal laws of the default times with intensities. At the opposite, it is more challenging to rewrite the full joint law of defaults because one needs to invoke multivariate hazard rates (Dabrowska, 1988; Fermanian, 1997). For example, a large number of intensities has to be modelized: $2^m - 1$ when m denotes the number of firms in the portfolio. In practice, such a number is unrealistic when dealing with more than 2 or 3 obligors.
2. Even if some firms or more generally some industries may be considered as negatively correlated with the "market", or rather with the vast majority of other corporates.
3. Since most of bank portfolios are composed mainly with investment grade debts, we overweight such firms.
4. The recovery rate is assumed to be zero here. This is not a limitation of our purpose. Indeed, in this chapter we do not try to study the internal source of randomness given by the exposure amounts.
5. Particularly Hull and White (2001), Schönbucher and Schubert (2003).
6. A random intensity model with constant levels between 0 and T and the same frailty for all obligors.
7. The relative sizes of the monthly default rates in Figure 7.3 are comparable with annual default rates because the former are trailed over 12 months.
8. The unobservable explanatory variables that are specific to i and that have not been taken into account previously in the vector X_i.
9. At least when we keep the balance between both Z_0 and Z_i.
10. For example, $\alpha = 2$ provides a Gaussian law and $\alpha = 1$, $\beta = 0$ provides a Cauchy distribution.
11. We are always dealing with the same portfolio from the beginning.
12. Paik *et al*. (1994) or Yue and Chan (1997), for instance.
13. because such λ_i are comparable with mean monthly default rates over a period T.

REFERENCES

Andersen, P.K., Gill, R., Borgan, O. and Keiding, N. (1997) *Statistical Models Based on Counting Processes* (New York: Springer).

Carr, P. and Wu, L. (2002) "The Finite Moment Log Stable Process and Option Pricing", Working Paper.

Chambers, J.M., Mallows, C.L. and Stuck, B.W. (1976) "A Method for Simulating Stable Random Variables", *Journal of the American Statistical Association*, 71(2): 340–4.

Clayton, D. and Cuzick, J. (1985) "Multivariate Generalizations of the Proportional Hazards Model", *Journal of the Royal Statistical Society*, *A*, 148(1): 82–117.

Crouhy, M., Galai, D. and Mark, R. (2000) "A Comparative Analysis of Current Credit Risk Models", *Journal of Banking and Finance*, 24(1–2): 59–117.

Dabrowska, D. (1988) "Kaplan–Meier on the Plane", *Annals of Statistics*, 16(4): 1475–89.

Das, S., Freed, L., Geng, G. and Kapadia, N. (2002) "Correlated Default Risk", EFA Annual Conference Paper 928 (http://ssrn.com/abstract=331901).

Duffie, D. and Singleton, K. (1999) “Modeling Term Structure of Defaultable Bonds”, *Review of Financial Studies*, 12(4): 687–720.

Elizalde, A. (2003) “Credit Risk Models I: Default Correlation in Intensity Models”, Working Paper CEMFI, Universidad Pública de Navarra, Spain.

Fermanian, J.D. (1997) “Multivariate Hazard Rates Under Random Censorship”, *Journal of Multivariate Analysis*, 62(2): 273–309.

Frey, R. and McNeil, A. (2001) Modelling Dependent Defaults. ETH E-collection.

Hougaard, P. (1986) “Survival Models for Heterogeneous Populations Derived from Stable Distributions”, *Biometrika*, 73(2): 387–96.

Hougaard, P. (2000) *Analysis of Multivariate Survival Data* (New York, NY: Statistics for Biology and Health, Springer).

Hull, J. and White, A. (2001) “Valuing Credit Default Swaps II: Modeling Default Correlations”, *Journal of Derivatives*, 8(3): 12–22.

Jarrow, R., Lando, D. and Turnbull, S. (1997) “A Markov Model for the Term Structure of Credit Risk Spreads”, *Review of Financial Studies*, 10(3): 481–523.

Koyluoglu, H. and Hickman, A. (1998) “A Generalized Framework for Credit Risk Portfolio Models”, Working Paper (New York: Wyman & Co.).

Luciano, E. (2004) “Credit Risk Assessment via Copulas: Association In-Variance and Risk-Neutrality”, Working Paper, University of Turin, Italy.

Metayer, B. (2005) “Shared Frailty Model for Rating Transitions”, University of Zurich Working Paper.

Mittnik, S. and Rachev, S.T. (1999) *Stable Models in Finance* (New York, NY: John Wiley & Sons).

Nolan, J.P. (2004) “Stable Distributions: Models for Heavy Tailed Data”, Working Paper, American University, Washington, DC.

Nyfeler, M. (2000) “Modeling Dependencies in Credit Risk Management”, Thesis, ETH, Zürich.

Paik, M.C., Tsai, W. and Ottman, R. (1994) “Multivariate Survival Analysis Using Piecewise Gamma Frailty”, *Biometrics*, 50(4): 975–88.

Parner, E. (1998) “Asymptotic Theory for the Correlated Gamma-Frailty Model”, *Annals of Statistics*, 26(1): 183–214.

Samorodnitsky, G. and Taqqu, M.S. (1994) *Stable Non-Gaussian Random Variables* (New York, NY: Chapman and Hall).

Schlögl, E. (2002) “Default Correlation Modeling”, Working Paper, University of Technology Sydney, Australia.

Schönbucher, P.J. (2001) “Factor Models for Portfolio Credit Risk”, Bonn Economics Discussion Paper, No. 16.

Schönbucher, P.J. (2003) *Credit Derivatives Pricing Models: Models, Pricing and Implementation* (New York, NY: John Wiley & Sons).

Schönbucher, P.J. and Schubert, D. (2001) “Copula-Dependent Default Risk in Intensity Models”, Bonn University Working Paper.

de Servigny, A. and Renault, O. (2002) “Default Correlation: Empirical Results”, Working Paper, Standard and Poor’s solutions.

Standard & Poor’s (2003) Special Report. Rating performances 2002. Février 2003.

Turnbull, S. (2003) “Practical Issues in Modeling Default Dependence”, Working Paper, Houston.

Yashin, A.I. and Iachine, I.A. (1995) “Genetic Analysis of Durations: Correlated Frailty Model Applied to Survival of Danish Twins”, *Genetic Epidemiology*, 12(3): 529–38.

Yu, F. (2003) “Dependent Default in Intensity-based Models”, Working Paper, University of California, Irvine.

Yue, H. and Chan, K.S. (1997) “A Dynamic Frailty Model for Multivariate Survival Data”, *Biometrics*, 53(3): 785–93.

CHAPTER 8

The Modeling of Weather Derivative Portfolio Risk

Stephen Jewson

8.1 INTRODUCTION

The companies that trade weather derivatives typically hold portfolios of between 100 and 1,000 weather derivative contracts. Different contracts have payoffs that may depend on different weather variables measured at different locations over different time periods. The payoffs between any two contracts may be highly correlated or anticorrelated (if they are based on the same or similar variables, locations or time periods), or they may be uncorrelated (if the weather variables, locations or time periods are very different). How, then, should the total financial risk in such portfolios be estimated?

In this chapter we present a number of methods for estimating this total portfolio risk and discuss some of the issues and trade-offs that arise when deciding which method to use. Several of the issues we discuss and the solutions we propose are novel in that, to our knowledge, they have not previously appeared in the literature on weather derivatives.

The chapter is structured as follows. We start with a brief description of what weather derivatives are.[1] We then explain how risk is usually defined for portfolios of weather derivatives, and follow that with a discussion of the two simplest methods for the evaluation of the risk of a weather derivative portfolio: burn analysis, and the application of the multivariate normal distribution to contract indices. In the main part of the chapter we then look beyond these basic methods and discuss a number of more complex issues: how simulations should be set up to account for sampling error; how estimates of the correlation matrix can be improved; how index non-normality can be accounted for; how the magnitude of model error can be estimated; how simulations can be re-engineered to incorporate hedging constraints between contracts; how consistency between single contract pricing and portfolio analysis can be achieved; how to calculate a simple linear approximation to the sampling error; and finally, how VaR can be estimated efficiently over short time horizons.

8.2 WHAT ARE WEATHER DERIVATIVES?

Weather derivatives are contracts between two parties that have a financial payoff that depends on some measured aspect of the weather. Since the future weather can be considered as random, the payoff of a weather derivative is also random. The economic purpose of weather derivatives is to allow companies that have profits that are affected by the weather to hedge some or all of that risk. This can be illustrated by a simple example, adapted from Jewson and Jones (2005):

A weather derivative example

ABC gas company doesn't like warm winters because they sell less natural gas to their domestic customers, who use the gas for heating their homes. ABC can lose up to £10 million in a warm winter relative to an average year. They decide to use weather derivatives to help hedge this warm winter risk. They analyse their historical revenues against historical weather data and conclude that there is a high correlation between their revenues and the total number of heating degree days measured in London between November and March (note that heating degree days, or HDDs, are a measure of the extent to which the temperature falls below 18 degrees Centrigrade, and, in this case, can be taken as a proxy for temperature on an inverted scale). Because of this high correlation they decide to base their weather derivative on a London November to March HDD index. This has the advantage that there is a well-traded market on this index, which makes it more likely that they will get a good price in the market because of the price-transparency brought about by such trading. In the first year of hedging they buy a put option, which will pay them if the number of HDDs is low (corresponding to a warm winter). A reasonable estimate of the average number of HDDs at this location and over this period is 1670 HDDs, with a standard deviation of 120HDDs, and the distribution of possible numbers of HDDs is close to normal. ABC decide to hedge themselves from 1650 HDDs downwards. They buy a put option with a strike of 1650HDDs, a tick of £50,000/HDD and a limit of £10,000,000 (this limit corresponds to 200HDDs below the strike, or 1450 HDDs). They compare quotes from a number of banks, and end up paying a premium of £2,000,000 for this contract. When the actual weather comes in at 1500 HDD they receive a payout of £7,500,000, and hence make an overall profit on the weather derivative of

£5,500,000. This roughly balances the money they lose on their gas supply business. In the second year they test a different strategy: they sell a swap with a strike of 1670 HDDs, a tick of £50,000/HDD, and limits at both ends of £10,000,000. Again the weather comes in warmer than normal, this time at 1640 HDDs. Again, they lose money on their gas business but make money on the weather derivative: this time £1,500,000.

This example illustrates the following important points. First, weather derivatives are based on weather measured at a specific location: in this case, London (in a real example it would probably be London's Heathrow Airport, weather station 03772). Second, weather derivatives are based on a weather index that has a single value per year: in this case, this index is the total number of HDDs during the winter season for this location. Third, there is a function that relates the value of the weather index to a payoff. Puts, calls and swaps are the most common functions used, but any other function is also possible. Swaps are typically traded without a premium, while options have a premium. Fourth, weather derivatives may have a limit on the financial payout. Typically, over the counter (OTC) contracts have limits while exchange traded contracts do not.

The example given above is very typical of trades in the weather market. Variations include:

(a) Using different locations: London, New York, Chicago and Tokyo are the most commonly traded locations, but many hundreds of other locations have also been used, and any location with reliable weather measurements is a potential candidate. For a hedger there may be a trade-off between using a location at which the weather correlates highly with their business, and using a commonly traded location for which better prices would be available.

(b) Using different weather variables: the bulk of the current market is based on temperature, but precipitation contracts are common, and wind contracts have also traded.

(c) Using different indices: temperature contracts are usually based on HDDs, as above, but can also be based on average temperature, the sum of daily temperatures, the number of days where the temperature exceeds a certain threshold, and so on.

(d) Using different time periods: monthly and seasonal contracts are the most common, although there are also contracts traded OTC with time periods as short as one hour.

(e) Using multiple locations or multiple variables at once in a single contract.

(f) Combining the definition of the index with financial variables such as gas price or power price.

The companies that sell weather derivatives to corporate hedgers like ABC gas company in our example are typically banks, insurance companies or hedge funds. Such companies endeavour to make a profit by selling weather derivatives using one or more of a number of possible strategies. Almost all such strategies involve selling more than one contract, and in many cases may involve trading many hundreds of contracts. They then have to consider the total financial risk in their portfolio.

8.3 DEFINING RISK FOR WEATHER DERIVATIVE PORTFOLIOS

A company that trades a weather derivative immediately assumes some financial risk, in that the ultimate outcome of the trade is uncertain. The risk from owning a portfolio of weather derivatives comes from the uncertainty associated with all the possible outcomes for the payouts of the contracts in that portfolio.

From the point of view of risk measurement, weather derivatives have similarities to both insurance contracts and financial derivatives. The risk of a portfolio of insurance contracts is typically measured in terms of the amounts of money, at various levels of probability, that the insurer may have to pay out to policy holders over the course of a year. In contrast, the risk of a portfolio of financial derivatives is typically measured in terms of how the liquidation value of the portfolio might reduce over a short time period, of maybe a day or a week. The rationale for this approach is that if the estimated risk is too large then one can consider actually liquidating the portfolio. The difference between the measurement of risk for insurance contracts and financial derivatives arises because insurance contracts cannot typically be liquidated or hedged, while financial derivatives can be.

Which of these two approaches is most appropriate for weather derivatives? In spite of the name, the insurance framework is the best starting point for understanding weather derivative risk. This is because the weather derivatives market is still rather small, and most weather derivatives cannot be liquidated or hedged in any practical way. The owner of the derivative usually has to face the possibility that they may have to hold the contract to expiry, and so the outcome at expiry is very important. The exception to this is that some contracts, particularly monthly and seasonal contracts based on temperature at locations such as London, Chicago and New York, are more heavily traded, and for these contracts it may be possible to liquidate or hedge positions. For these contracts, it may therefore be reasonable to start thinking about risk over short time-horizons, as with other financial derivatives. For most of this chapter, however, we will deal with the insurance definition of risk, which we will call “expiry risk”, since it considers the outcome of contracts at expiry. Then in section 8.12 we briefly come back to the question of how to estimate risk over shorter time horizons.

8.4 BASIC METHODS FOR ESTIMATING THE RISK IN WEATHER DERIVATIVE PORTFOLIOS

We now present the two most basic methods that one might use to estimate the risk in a weather derivative portfolio, as a starting point for our subsequent discussion. These methods are (a) burn analysis, and (b) use of the multivariate normal distribution to model weather indices.

8.4.1 Burn analysis

Burn analysis, which is the simplest method for analysing risk in a weather derivative portfolio, typically works as follows:

1. For each contract in the portfolio, 10 years of cleaned historical weather data is purchased.
2. For each of these 10 years, the historical settlement indices for each contract in the portfolio are calculated.
3. Trends (such as the global warming trend) may be removed from these historical settlement indices, if appropriate.
4. The detrended historical settlement indices are converted into historical payoffs for each contract and each historical year.
5. The historical payoffs are aggregated over the portfolio, giving a portfolio historical payoff for each of the 10 years.
6. The 10 portfolio historical payoffs thus obtained are taken as an empirical estimate for the distribution of possible payoffs for the portfolio. Quantities such as the expected payoff, the variance of payoffs or the risk of extreme losses can then be estimated.

This method is sufficiently simple that it could be implemented in a spreadsheet. However, it has two major shortcomings. First, since the estimate of the distribution of payoffs is based on only 10 points, the highest level of risk than can be estimated is 1 in 10 year risk. More years of historical data could be used in order to estimate risk at higher levels, but even in the best cases the maximum number of years of reliable data typically available is only around 50. This is not particularly satisfactory for risk managers, who often like to see estimates of the 1 in 100 or 1 in 1,000 year risk. One might be tempted to fit a distribution to the historical payoffs of the portfolio, and use the fitted distribution to extrapolate the risk to higher levels (we have occasionally called this extended burn analysis). However, such fitting and extrapolation is more guesswork than science since we have no a priori idea which distribution to fit. The payoff distribution is seldom close to normal, even for very large portfolios, because of the non-linearity of many of the contracts in a typical portfolio. The second shortcoming of burn analysis is that estimates of the greeks of the portfolio, and other important diagnostics such as the regression coefficients, or "betas", of the portfolio, cannot be calculated in an accurate way.

These shortcomings motivate the use of simulations, as we now describe.

8.4.2 Basic use of the multivariate normal

Both the shortcomings described above can be overcome by using simulations, and the simplest example of using simulations is one particular use of the multivariate normal distribution, as follows:[2]

1–3 As for burn analysis.

4 A multivariate normal distribution is fitted to the detrended historical settlement indices. The covariance matrix of the multivariate normal is estimated using the empirical covariance matrix of these indices.

5 10,000 years of simulated settlement indices are created from this fitted distribution.

6 The simulated settlement indices are converted into payoffs for each contract and each simulated year.

7 The simulated payoffs are aggregated over the portfolio, giving a portfolio simulated payoff for each simulated year.

8 The 10,000 portfolio simulated payoffs thus obtained are taken as an empirical estimate for the distribution of possible payoffs of the portfolio.

Risk can then be estimated at levels as high as 1 in 10,000 years, or even higher levels if more years of simulations are used. Of course, these estimates are only as good as the data and assumptions on which they are based, and these assumptions become more dubious the further one progresses into the tail, but at least this method gives us *some* way for estimating the probabilities of extreme losses.

In the following sections we now discuss various extensions and alternatives to the use of the multivariate normal distribution as described above.

We will refer to this basic method as the BMVN (Basic MultiVariate Normal) method.

8.5 THE INCORPORATION OF SAMPLING ERROR IN SIMULATIONS

The BMVN method described above is not strictly correct, from a statistical point of view, even in the case in which all the assumptions hold (for example, the weather indices really are multivariate normal and we know the correct model for the trends). This is because it ignores the sampling, or estimation, uncertainty on the covariance matrix, and does not propagate that uncertainty into the simulations. For instance, in this method the variance of a particular weather index is estimated using the available data, and the simulations are driven by that estimated variance. However, the estimated variance is only an estimate, and the variance of the simulations should include an extra term that takes this into account (this issue is identical to the question of how to derive prediction intervals in classical statistics). Deriving the extra terms in the expressions for the covariances can be somewhat complicated, but may make a significant difference to the final results. In the case in which the weather indices have not been detrended the extra terms are rather simple, and are typically rather small. In the case in which the weather indices have been detrended with a linear trend, the extra terms are more complex, and become much more significant. In the case in which other trend shapes are used, deriving the extra terms is complex, but very important. Jewson and Penzer (2004) give expressions for these extra terms for a number of cases. It should be noted that these extra terms are *in addition* to the standard corrections to expressions for covariances that account for changes in the number of degrees of freedom due to detrending.

8.6 ACCURATE ESTIMATION OF THE CORRELATION MATRIX

The BMVN method, step 4, involves estimating the correlation matrix among the historical weather indices using the empirical correlation matrix; that is, calculating the observed correlations between historical weather indices. Curiously, however, the empirical correlation matrix is not a particularly good estimator of the real correlation matrix, particularly for large portfolios. This is because the elements of the correlation matrix are very poorly estimated given the little data available. For example, for a portfolio of 1,000 contracts, we must estimate roughly 500,000 correlations, but with perhaps only 50,000 historical data values. This situation can be improved using a technique known as shrinkage, in which the correlation matrix is estimated using a combination of the empirical correlation matrix with a much simpler estimate such as a correlation matrix based on independence. A non-parametric approach to implementing such shrinkage for weather data, based on the Quenouille–Tukey jackknife, has been described in Jewson (2005), while a parametric approach for implementing shrinkage for financial data correlation matrices has been described in Ledoit and Wolf (2003). As far as the author is aware there has been no attempt to compare the two approaches.

8.7 DEALING WITH NON-NORMALITY

One of the most obvious assumptions in the BMVN method is that the joint distribution of the weather indices underlying the weather derivative portfolio is multivariate normal. The assumption of multivariate normality consists of the assumption that the marginals are normal and the assumption that the copula is a Gaussian copula. With respect to the first of these assumptions, empirical tests have shown that normality is a good model for the marginals for seasonal temperature contracts, but is often not satisfactory for monthly temperature contracts. It is certainly not a good model for counting indices (such as an index which counts the number of freezing days) when the counts are small.

There is a simple, and standard, method for incorporating non-normal marginals into multivariate simulations. This method was apparently first described by Iman and Conover (1982), and has recently been popularized by Wang (1998). The method works as follows:

1 Marginal distributions are fitted to each weather index.
2 Using the CDFs for these marginal distributions, and the inverse of the CDF of the normal distribution, the historical weather indices are transformed to come from a normal distribution.[3]
3 The transformed values are simulated using the multivariate normal distribution.
4 The simulated values are transformed back to the original marginal distributions, using the reverse of step 2.

The simulated values produced by this method do not have the same *linear* correlations as the original historical values, but it is usually argued that linear correlation is not a good measure of dependence anyway when considering non-normal distributions. Instead, the method yields simulated values with the correct *rank* correlations.

This method is a simple example of the use of copulas. In this case, the copula being used is the normal copula, since the historical values are transformed to a multivariate normal for the simulation step. Alternatively, one could transform to any other multivariate distribution, such as a multivariate t-distribution, or a multivariate non-parametric distribution. There is no weather-derivative specific published work in this area, however, at this point, and no evidence has been presented either for or against the general validity of the Gaussian copula for weather indices (although one would assume that there are probably cases where the Gaussian copula is not a good model).

8.8 ESTIMATING MODEL ERROR

One of the biggest shortcomings of the BMVN method is that the results may depend in a sensitive way on the assumptions in the method, and the assumptions may be wrong. For comparison, analysis of the pricing of single weather derivative contracts has shown that the results are very sensitive to the choice of the numbers of years of historical data used and the form of detrending, but less sensitive to the choice of distribution (see Jewson (2004)). These results presumably carry over to the portfolio case. This presents a serious problem for the risk manager: an apparently innocuous decision to use 30 rather than 40 years of historical data may have a large impact on the final results. What can be done about this? We offer two approaches:

1 Run scenario tests on the assumptions used. By this we mean: vary each assumption within a reasonable range, see how much the final results change, and combine all the results together. For instance, repeat all calculations with 40 rather than 30 years of data, with non-linear rather than linear trends, with non-parametric rather than parametric distributions, and with different estimators for the correlation matrix. The advantage of this approach is that it is simple to do. The disadvantage is that it is *ad hoc*, and very subjective. Two different practitioners would vary different sets of assumptions, and by different amounts, and would get different results.

2 Use a number of different models, with likelihood weighting (also called "Bayesian Model Averaging": see Hoeting, Madigan, Raftery and Volinsky, 1999). This is a slightly more formal version of the previous method that avoids some of the subjectivity. This time, for each set of assumptions, we calculate the likelihood of the historical observations. Only the sets of assumptions that give high values for the likelihood are retained, and the final simulations are combined using the likelihood values for each model. This method still involves some subjectivity in the definition of the set of models to be considered, but avoids the ad-hockery in the decision of how much each assumption should be varied.

8.9 INCORPORATING HEDGING CONSTRAINTS

Obvious constraints between the payoffs of contracts within a weather derivative portfolio are not necessarily preserved by the BMVN method, or its extension to non-normal distributions. This is an aspect of the fact that it is never possible to exactly capture the correct multivariate distribution of the weather indices, even if using copulas.

As an example, consider the following three contract portfolio: contract 1 is based on the number of freezing days in Chicago in November; contract 2 is based on the number of freezing days in Chicago in December; and contract 3 is based on the number of freezing days in Chicago in November and December. Clearly, in any particular year the index for contract 3 is the sum of the indices from contracts 1 and 2. However, the BMVN method (even with extensions to copulas) is not guaranteed to preserve this constraint i.e. may produce simulated years in which this constraint does not hold. This is unlikely to be a problem in most cases, since the errors will probably be small, but if contracts are being traded in such a way that one is relying on this constraint to get exact cancellation of risk (for instance, with long swaps as contracts 1 and 2 and a short swap as contract 3) and the simulations miss the exact cancellation and give a finite instead of zero risk, then this may be a problem.

The most obvious way to ensure that hedging constraints are captured is to use some level of atomic simulation.[4] That is, simulating at a more detailed level than the contract index level. For instance, one could consider all seasonal indices as sums of monthly indices, simulate the monthly indices, and create simulated seasonal indices by summing the simulated monthly indices. This would have the advantage that it would correctly capture constraints related to sums of monthly indices. However, it would introduce various disadvantages too: first, that monthly indices are harder to model because they are less likely to be normally distributed, second that modelling monthly rather than seasonal indices may involve estimating many more parameters, and thus introduce extra sources of parameter error, and third that simulating monthly indices may increase the dimensionality of the problem (increasing the size of the correlation matrix), which will make the whole modelling process slower.

An even more atomic approach would be to simulate daily temperatures, and there have been a number of articles written on this topic, such as those of Dischel (1998), Cao and Wei (2000), Dornier and Queruel (2000), Moreno (2000), Torro, Meneu and Valor (2001), Alaton and Djehiche and Stillberger (2001), Moreno and Roustant (2002), Caballero, Jewson and Brix (2002), Brody, Syroka and Zervos (2002) and Jewson and Caballero (2003). This is, however, a very difficult statistical problem to solve in general and the simulations can be very slow.

This particular issue highlights that there is no single best solution for modeling weather derivative portfolios. In many cases, simulating at the contract index level may be best. In other cases, in which constraints between monthly contracts are very important, simulating at the monthly level may be better. Finally, for some contracts, and for some trading situations, simulation at a daily level may be optimal.

8.10 CONSISTENCY BETWEEN THE VALUATION OF SINGLE CONTRACTS AND PORTFOLIOS

Single weather contracts can be priced using closed-form expressions or simulations. Whichever method is used, however, the estimate of the expected payoff for a single contract will not be the same as the estimate of the expected payoff of the same contract when included as part of a portfolio which is being valued using the BMVN method. If the single contract is valued using closed-form expressions then this difference arises because the BMVN method uses simulations, which necessarily introduces a small random error. If the single contract is valued using simulations then this difference arises because different simulation engines must be used for the two sets of simulations, since one is univariate and the other is multivariate.

As long as many years of simulations are being used the differences are not likely to be material relative to sampling and model error, but they can be rather inconvenient and confusing. It would therefore be useful to be able to make these two sets of results numerically consistent. One way to do this is as follows:

1 When pricing stand-alone contracts, use simulations from a univariate random number generator.

2 When modelling the portfolio, start by running simulations for the marginal distributions using the same univariate random number generator as used for pricing individual contracts, with the same seeds on a contract by contract basis.

3 Then, induce the desired correlation matrix between the independent univariate simulations of the marginals by reordering the simulated values. The reordering is based on the observed rank correlation matrix and a set of correlated multivariate normal simulations.

The result is that the simulated marginal distribution for each contract is the same in both the univariate and multivariate cases. The only disadvantages of this method appear to be (a) that simulations have to be used for the pricing of stand-alone contracts (instead of closed-form expressions, which are faster and more accurate), and (b) that it may be slightly slower than the BMVN method since it involves two stages of simulation.

8.11 ESTIMATING SAMPLING ERROR

As discussed in section 8.5, even if the assumptions used to model a set of weather indices (in terms of the shapes of trends and distributions) are correct, then the actual values of the fitted parameters are always estimated, and this induces errors into the final results that we call sampling error (sampling error is to be contrasted with model error, as discussed in section 8.8 above, which is caused by the models being wrong).

It may be useful to estimate the role of sampling error when pricing a weather derivative or valuing a weather derivative portfolio. This is because sampling error is a major contributor to overall error, and knowledge of the level of error in pricing can lead to better decisions about whether to trade, and how to set risk loading levels.

Unlike model error, sampling error can be estimated rather straightforwardly using linear theory. The case for a single weather derivative is described in Jewson (2003a). This can be generalised to the portfolio case, at least for the multivariate normal, although the derivation is somewhat involved, and, to the author's knowledge, has not been published.

8.12 ESTIMATING VaR

As discussed in section 8.3, risk in weather derivatives portfolios is principally measured by considering the distribution of outcomes of contracts at expiry. However, it is also of interest to estimate the likely fluctuations in the value of a weather derivative portfolio over much shorter time horizons. This is particularly interesting for commonly traded contracts for which liquidation or hedging may be an option. There are two common cases: either one wishes to derive the possible changes in the expected expiry value, or one wishes to derive the possible changes in the market value. With respect to the first of these two cases, a full calculation is extremely complex, depending, as it does, on modelling changes in weather forecasts, changes in weather, and the correlations between the two. But a linearized estimate, likely to be accurate for short time periods, is much simpler. Calculating such an estimate involves linearising the non-linearities in the payoff functions and modelling the short term changes in expected weather indices using Brownian motion. The resulting expressions, which can be derived for both single contracts and portfolios, are very simple (see Jewson, 2003b). With respect to the second of these cases, that of estimating possible changes in the *market* value of a portfolio of weather derivatives, one has to consider modelling the fluctuations in market prices. If one assumes that market prices are given by expectations, this is identical to the first case. If, however, one assumes that the market contains additional supply and demand dynamics that are also important, then this is a rather more difficult question. The starting point for any attack on this problem would have to be an attempt to model fluctuations in observed market prices.

8.13 CONCLUSION

We have discussed methods that can be used for the estimation of risk in portfolios of weather derivatives. The simplest, but very limited, method is burn analysis. To improve on burn analysis one can use simulations, and the simplest way to do that is to use the multivariate normal distribution for the weather indices underlying the contracts in a portfolio. However, this method has a number of shortcomings. We discuss how some of these shortcomings can be addressed. Some of the methods we describe are standard in industry. Most, however, are the subject of current research, and are yet to be applied in practice. Furthermore we have highlighted a number of areas where further research would be useful. In particular it would be beneficial to develop a better understanding of methods for accurate estimation of correlation matrices, and to explore whether there may be benefits to be had from using copulas other than the Gaussian copula.

NOTES

1. More details can be found in books on weather derivatives such as Element Re (2002), Dischel (2002) or Jewson, Brix and Ziehmann (2005).
2. The earliest references we have seen for this method are Goldman Sachs (1999) and Zeng and Perry (2002).
3. In fact there is a shortcut for this step: see Wang (1998).
4. Thanks to Seth Padowitz for the terminology, and discussions on this issue.

REFERENCES

Alaton, P., Djehiche, B. and Stillberger, D. (2002) "On Modelling and Pricing Weather Derivatives", *Applied Mathematical Finance*, 9(1): 1–20.

Brody, D., Syroka, J. and Zervos, M. (2002) "Dynamical Pricing of Weather Derivatives", *Quantitative Finance*, 2(3): 189–98.

Caballero, R., Jewson, S. and Brix, A. (2002) "Long Memory in Surface Air Temperature: Detection, Modelling and Application to Weather Derivative Valuation", *Climate Research*, 21(4): 127–40.

Cao, M. and Wei, J. (2000) "Pricing the Weather", *Risk*, 13(5): 67–70.

Dischel, R. (1998) "Black-Scholes Won't Do", *Energy and Power Risk Management Weather Risk Special Report*, October.

Dischel, R. (2002) (Ed.). *Climate Risk and the Weather Market* (London: Risk Books).

Dornier, F. and Queruel, M. (2000) "Caution to the Wind", *Energy and Power Risk Management Weather Risk Special Report*, 8(2): 30–2.

Element, R. (2002) *Weather Risk Management* (London: Palgrave Macmillan).

Goldman Sachs (1999) Kelvin Ltd. *Offering Circular.*

Hoeting, J., Madigan, D., Raftery, A. and Volinsky, C. (1999) "Bayesian Model Averaging", *Statistical Science*, 14(2): 382–401.

Iman, R. and Conover, W. (1982) "A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables", *Communications in Statistics*, 11(1): 311–34.

Jewson, S. (2003a) "Estimation of Uncertainty in the Pricing of Weather Options", Working Paper, www.ssrn.com/abstract=441286

Jewson, S. (2003b) "Horizon Value at Risk for Weather Derivatives Part 2: Portfolios", Working Paper, www.ssrn.com/abstract=478051

Jewson, S. (2004) "The Relative Importance of Trends, Distributions and the Number of Years of Data in the Pricing of Weather Options", Working Paper, www.ssrn.com/abstract=516503

Jewson, S. (2005) "Improving on the Empirical Covariance Matrix Using Truncated PCA with White Noise Residuals", Working Paper, *arXiv:physics/0506055.*

Jewson, S. and Caballero, R. (2002) "Multivariate Long-Memory Modelling of Daily Surface Air Temperatures and the Valuation of Weather Derivative Portfolios", Working Paper, www.ssrn.com/abstract=405800

Jewson, S. and Jones, S. (2005) "Weather Derivatives and Carbon Emissions Trading", in K. Tang (ed.), *The Finance of Climate Change* (London: Risk Books).

Jewson, S. and Penzer, J. (2004) "Optimal Year-Ahead Forecasting of Temperature in the Presence of a Linear Trend and the Pricing of Weather Derivatives", Working Paper, www.ssrn.com/abstract=563943

Jewson, S., Brix, A. and Ziehmann, C. (2005) *Weather Derivative Valuation* (Cambridge: Cambridge University Press).

Ledoit, O. and Wolf, M. (2003) "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection", *Journal of Empirical Finance*, 10(5): 603–21.

Moreno, M. (2000) "Riding the Temp", *Futures and Options World*, 1–12 December.

Moreno, M. and Roustant, O. (2002) "Modelisation de la Temperature: Application aux Derives Climatiques", in J. Blondeau and C. Partrat (eds), *La Reassurance, Approche Technique* (Paris: Economica).

Torro, H., Meneu, V. and Valor, E. (2001) "Single Factor Stochastic Models with Seasonality Applied to Underlying Weather Derivatives Variables", Technical Report No. 60, European Financial Management Association.

Wang, S. (1998) "Aggregation of Correlated Risk Portfolios: Models and Algorithms", *Proceedings of the Casualty Actuarial Society*, 85(163): 1–32.

Zeng, L. and Perry, K. (2002) "Managing a Portfolio of Weather Derivatives", in R.S. Dischel (ed.), *Climate Risk and the Weather Market*, pp. 241–64 (London: Risk Books).

CHAPTER 9

Optimal Investment with Inflation-Linked Products

*Taras Beletski and Ralf Korn**

9.1 INTRODUCTION

With the growing number of traded inflation linked bonds and inflation linked life insurance products there is also a growing interest in models for the evolution of inflation indexes and the inclusion of inflation linked financial products into an optimal portfolio of an investor who is otherwise investing in bonds and stocks. We will look at this problem in a model that is a combination of the standard diffusion type model of continuous-time portfolio optimization and a modeling framework for inflation indexes described in Korn and Kruse (2004) (which itself is in some aspects related to Jarrow and Yildirim, 2003).

There, an inflation index such as the harmonized consumer price index (HCPI) is modelled as a (generalized) geometric Brownian motion with a drift equal to the difference of the nominal and the real interest rate. This modeling process will be shortly described in section 9.2. As a consequence of it there will be Black–Scholes type formulae for the prices of inflation linked bonds and options on inflation. Picking up an approach of Korn and Trautmann (1999) – later generalized by Kraft (2003) – a mixed investment problem including those products and conventional investment into stocks

* The work of Taras Beletski was supported by the DFG-Graduiertenkolleg "Mathematik und Praxis". The work of Ralf Korn was supported by the Rheinland-Pfalz excellence cluster "Dependable adaptive systems and mathematical modeling".

and the money market account will be set up and solved in section 9.3. Finally, in section 9.4 we show how inflation-linked products can be used to hedge inflation dependent claims.

9.2 MODELING THE EVOLUTION OF AN INFLATION INDEX

The harmonized consumer price index (HCPI) is designed to measure inflation in the countries of the European monetary union. It is an average over 11 country-based inflation indices. As it is therefore an official index it is a natural candidate for options to be written on its future value or for linking bond payments to it. However, to value these contracts we need a model for the evolution of an inflation index over time.

To develop such a model we will base our considerations on macroeconomic foundations as there should of course be relations between (different kinds of) interest rates and inflation. The most prominent rule in this area is the so-called Fisher equation (Fisher, 1930). It states that the nominal interest rate is the sum of the real interest rate and the expected inflation:

$$r_N(t) = r_R(t) + E[i(t)] \tag{9.1}$$

where $r_N(t)$ is the nominal interest rate for the bond maturing at time t, $E[i(t)]$ is the expected (simple) inflation rate for the time horizon t and $r_R(t)$ is the real interest rate for the bond with maturity t, which corresponds to the growth of real purchasing power in the case of investment with the nominal interest rate $r_N(t)$. In the special case of constant real interest rates, thus the nominal interest rates follow the movements of expected inflation rate (a fact empirically supported by Ang and Bekaert (2003) and Nielsen (2003)). By interpreting the relative instantaneous change:

$$\frac{dI(t)}{I(t)}$$

of the inflation index $I(t)$ as the (instantaneous) rate of inflation $i(t)$ the Fisher equation suggests the following model of a generalized geometric Brownian motion for the evolution of the inflation index:

$$dI(t) = I(t)((r_N(t) - r_R(t))dt + \sigma_I dW_I(t)), \quad I(0) = i \tag{9.2}$$

where now $r_N(t)$, $r_R(t)$ are interpreted as the relevant instantaneous rates. Even more, we assume that this equation holds in equilibrium, for example, equation (9.1) is valid with respect to the risk-neutral pricing measure (the modeling of the evolution of the inflation index under a subjective measure can be done by including an additional drift rate such as for example, $\lambda\sigma_I \in \mathrm{IR}$; see also sections 9.3 and 9.4). According to the specification of the nominal and the real interest rate we can produce models of different complexity out of relation (9.2). Note also that we have a mean-reverting drift with the real rate being the mean reversion level. Especially, with this geometric Brownian motion based model we are able to derive option pricing formulae of Black–Scholes type. To understand their derivation note that our situation can be viewed as an alternative to the pricing of a foreign exchange option. The inflation index allows us to switch between an investment in the nominal and in the real currency (see also Korn and Kruse, 2004, for the formal argument):

Proposition 1 Under the assumptions of deterministic real and nominal interest rates in equation (9.2) the fair price of a European call option on the inflation index $I(T)$ at time t with strike price K and maturity T is given by:

$$C_I(t, I(t)) = I(t)\exp\left(-\int_t^T r_R(s)ds\right) N(d(t)) - K\exp\left(-\int_t^T r_N(s)ds\right) N(d(t) - \sigma_I\sqrt{T-t}) \qquad (9.3)$$

where N is the cumulative distribution function of the standard normal distribution and

$$d(t) = \frac{\ln\left(\frac{I(t)}{K}\right) + \int_t^T (r_N(s) - r_R(s))ds + \frac{1}{2}\sigma_I^2(T-t)}{\sigma_I\sqrt{T-t}} \qquad (9.4)$$

Of course, much more natural products than call options on a consumer price index such as an inflation index are inflation linked bonds. A typical such example is a coupon bond with coupons protected against inflation and the final payment of the notional being protected against inflation and deflation, for example, it consists of payments:

$$C_i\frac{I(t_i)}{I(t_0)}, \text{ at times } t_i,\ i = 1, \ldots, n, \qquad (9.5)$$

$$\max\left\{F\frac{I(t_n)}{I(t_0)}, F\right\}, \text{ at time } t_n = T \qquad (9.6)$$

Under the above assumptions of deterministic interest rates we can again derive a closed formula for its price (Korn and Kruse, 2004):

Proposition 2 Under the assumptions of deterministic real and nominal interest rates in equation (2.2) the fair price of an inflation-linked T-bond at time t with a reference date $t_0 \le t \le t_1$, face value F and coupon payments

C_i (before adjustment to the inflation) at times t_i is given by:

$$B_{IL}(t, I(t)) = \sum_{i=1}^{n} C_i \frac{I(t)}{I(t_0)} \exp\left(-\int_t^{t_i} r_R(s)ds\right) + F\left(\exp\left(-\int_t^{T} r_N(s)ds\right) + \frac{C_I(t, I(t))}{I(t_0)}\right) \tag{9.7}$$

where $t_0 \leq t < t_1 \leq \cdots \leq t_n = T$ and $C_I(t, I(t))$ is a fair price of the European call option on consumer price index $I(T)$ at time t with strike price $K = I(t_0)$ and date of maturity T.

Remark (i) Of course one could also link a deflation protection to the single coupon payments (for example, to pay at least their nominal values). Then, the price of the so constructed inflation linked bond is given as the sum of $n+1$ call options on the inflation index. A special case of the above considered coupon bond is an inflation linked zero coupon bond with deflation protection that is obtained by setting $C_i = 0$. Also, the necessary modifications for considering an inflation-linked coupon bond without deflation protection of the final payment are quite obvious.

(ii) Another possible approach for inflation modeling is the direct modeling of an (instantaneous) inflation rate similar to the short rate approach of interest rate modeling. We refer to Korn and Kruse (2004) for details on this method. A similar approach based in stead on the HJM framework is presented in Jarrow and Yildirim (2003).

9.3 OPTIMAL PORTFOLIOS WITH INFLATION LINKED PRODUCTS

In this section we will look at various optimization problems including aspects of inflation. The main problem of investment in the presence of inflation is of course that inflation itself is not a traded good. However, inflation-linked bonds or options on an inflation index are traded and can serve as investment alternatives. As they are derivatives on the underlying inflation index we can use portfolio optimization methods for portfolios with derivatives as in Korn and Trautmann (1999) or in Kraft (2003). This approach roughly consists of the following two-step procedure:

Step 1: Solve the optimal portfolio problem (P) as if all the underlyings (and in particular the inflation index) were tradable (“*the basic portfolio problem*”).

Step 2: Replicate the optimal positions of the non-tradable assets via positions in suitable derivatives and the money market account.

Of, course to use it we first have to set up the relevant optimization problem. For this, we assume that an investor is allowed to trade into a riskless bond with price given by:

$$dP_0(t) = P_0(t) r_N \, dt, \qquad P_0(0) = 1 \tag{9.8}$$

In addition, he can invest into n further risky securities which might be stocks, inflation linked bonds, or more general derivatives on inflation and/or stocks. For a given initial wealth of x and an admissible portfolio process $\pi(.) \in A(x)$ (to be specified later) including the above investment possibilities let $X^{\pi}(t)$ denote the corresponding wealth process. Then, the investor's task will be to maximize the expected utility from final wealth, for example, he tries to solve the portfolio problem (P):

$$\sup_{\pi(.)\in A(x)} E(U(X^{\pi}(T))) \tag{P}$$

where $U(x)$ is a utility function (for example, a strictly concave, monotonically increasing and differentiable function). To simplify matters we will assume that besides the riskless bond above the investor can invest into a risky stock with price dynamics given by:

$$dP_1(t) = P_1(t)(b\,dt + \sigma_I \, dW_1(t) + \sigma_2 \, dW_2(t)), \quad P_1(0) = p_1 \tag{9.9}$$

and in some inflation linked product where the inflation dynamics are given by:

$$dI(t) = I(t)((r_N - r_R + \lambda\sigma_I)dt + \sigma_I \, dW_2(t)), \qquad I(0) = i \tag{9.10}$$

with (W_1, W_2) denoting a two-dimensional Brownian motion. If in such a situation we assume that it is possible to trade in two derivatives on the stock and on the inflation with price processes given by:

$$f^{(i)}(t, P_1(t), I(t)), \quad f^{(i)} \in C^{1,2}([0,\ T) \times (0, \infty)^2), \quad i = 1,\ 2 \tag{9.11}$$

then the relevant result from Korn and Trautmann (1999) can be formulated as (where we will omit a proof here as it would be totally similar to the one given in the Korn and Trautmann (1999) for the case of optimal investment in stock derivatives).

Theorem 1 Under the assumption that the **delta-matrix** $\Psi(t) = (\Psi_{ij}(t))$, $i, j = 1, 2$ with

$$\Psi = \begin{pmatrix} f_{p_1}^{(1)}(t, P_1(t), I(t)) & f_I^{(1)}(t, P_1(t),\ I(t)) \\ f_{p_1}^{(2)}(t, P_1(t), I(t)) & f_I^{(2)}(t, P_1(t),\ I(t)) \end{pmatrix} \tag{9.12}$$

(with the subscripts denoting the corresponding partial derivatives) is regular for all t ∈ [0, T) then the option portfolio problem (P_1)

$$\max_{\varphi(.)\in B(x)} E(U(X^{\varphi}(T))) \tag{OP}$$

with

$$X^{\varphi}(t) = \varphi_0(t)P_0(t) + \varphi_1(t)f^{(1)}(t, P_1(t), I(t)) \\ + \varphi_2(t)f^{(2)}(t, P_1(t), I(t)) \quad (9.13)$$

(and B(x) the set of all admissible trading strategies in the bond and the two derivatives for an initial wealth of x) admits the following solution:

(a) The optimal final wealth B* coincides with the optimal final wealth in the corresponding basic portfolio problem (P) where the investor is assumed to be able to trade the stock and the inflation index.

(b) Let $\xi(t) = (\xi_0(t),\ \xi_1(t),\ \xi_2(t))$ denote the optimal trading strategy of the corresponding basic portfolio problem (P). Then the optimal trading strategy for the option portfolio problem $\varphi(t) = (\varphi_0(t),\ \varphi_1(t),\ \varphi_2(t))$ is given by:

$$\bar{\varphi}(t) = (\Psi(t)')^{-1} \cdot \bar{\xi}(t), \\ \varphi_0(t) = \frac{\left(X(t) - \sum_{i=1}^{d} \varphi_i(t) f^{(i)}(t, P_1(t), \ldots, P_d(t))\right)}{P_0(t)} \quad (9.14)$$

with $\bar{\varphi}(t) = (\varphi_1(t), \varphi_2(t))'$ and $\bar{\xi}(t) = (\xi_1(t), \xi_2(t))$.

Equipped with this result we are now able to solve various particular portfolio problems related to inflation explicitly. But first we recall examples of explicit solutions of the basic portfolio problem (P):

Step 1: Solving the basic portfolio problem (P). In this step we will treat the inflation index as a tradable good and solve the portfolio problem (P) for the choices of

$$U(x) \in \{\ln(x), \quad 1/\gamma\, x^{\gamma}\} \quad \text{with } \gamma < 1 \quad (9.15)$$

With the notation of

$$\sigma = \begin{pmatrix} \sigma_1 & \sigma_2 \\ 0 & \sigma_I \end{pmatrix} \quad (9.16)$$

the solutions of the basic portfolio problem (P) for these choices of the utility functions are well-known and are given by:

$$\pi^*(t) = \frac{1}{1-\gamma}(\sigma\sigma')^{-1} \begin{pmatrix} b - r_N \\ \lambda\sigma_I - r_R \end{pmatrix} \quad (9.17)$$

where the case of $\gamma = 0$ corresponds to the choice of the logarithmic utility function. If in particular the stock price is independent of the inflation index (for example, we have $\sigma_2 = 0$) then we obtain the special form of

$$\pi^*(t) = \frac{1}{1-\gamma}\begin{pmatrix} (b - r_N)/\sigma_1^2 \\ (\lambda\sigma_I - r_R)/\sigma_I^2 \end{pmatrix} \tag{9.18}$$

So in both cases, the optimal fractions of wealth invested in the risky assets are functions of the excess return suitably weighted by their volatilities. Note especially that the subjective excess return $\lambda\sigma_I$ has to be bigger than the real rate – an assumption that rarely seems to hold. So in this – hypothetical – portfolio problem a risk averse investor typically sells inflation when behaving optimally!

Step 2: Optimal portfolios with inflation linked products.
Problem 1: Inflation-linked bond and non-inflation linked bond.
In this first setting we assume that the investor has access to a market consisting of the riskless bond with price $P_0(t)$ and an inflation-linked coupon bond with coupon payments as described in relation (9.5) and a final payment as in (9.6), for example, we look at the following special case of the portfolio problem (OP):

$$\max_{\varphi(.)\in B(x)} E(U(X^{\varphi}(T))) \tag{OP1}$$

with

$$X^{\varphi}(t) = \varphi_0(t)P_0(t) + \varphi_1(t)f^{(1)}(t, I(t)) \tag{9.19}$$

and where the function $f^{(1)}$ coincides with $B_{IL}(t, I(t))$ of Proposition 9.2. As shown in Korn and Trautmann (1999) we only have to use the replication strategy in the riskless bond and the inflation index for this inflation-linked bond to determine the optimal trading strategy for the corresponding portfolio problem (OP), the dynamics of its price process $B_{IL}(t, I(t))$ play no explicit role. However, as can be directly verified, the number of the shares in the inflation index of this replication strategy has the following form:

$$\psi_1(t) = \sum_{i:t_i>t} \frac{C_i}{I(t_0)}\exp(-r_R(t_i - t)) + \frac{F\exp(-r_R(T-t))N(d(t))}{I(t_0)} \tag{9.20}$$

Combining this with the results of Step 1 and Theorem 1 leads to the optimal trading strategy in the inflation-linked bond of

$$\varphi_1(t) = \frac{\xi_1(t)}{\psi_1(t)} = \frac{\lambda\sigma_I - r_R(t)}{(1-\gamma)\sigma_I^2} \cdot \frac{X(t)}{\psi_1(t)I(t)} \tag{9.21}$$

with $X(t)$ being the optimal wealth process of the basic portfolio problem (P) (which coincides with the optimal wealth process of (OP1) by Korn and

Trautmann (1999)). To understand the qualitative behavior of the optimal strategy we look at the corresponding portfolio process:

$$\pi_1(t) = \frac{\varphi_1(t)B_{IL}(t,I(t))}{X(t)} = \frac{\lambda\sigma_I - r_R}{(1-\gamma)\sigma_I^2}\frac{B_{IL}(t,I(t))}{\psi_1(t)I(t)}$$
$$= \frac{\lambda\sigma_I - r_R}{(1-\gamma)\sigma_I^2}\theta(t,I(t)). \tag{9.22}$$

By comparing relations (9.20) and (9.5) and using

$$\psi_1(t)I(t) = B_{IL}(t,I(t)) - e^{-r_N(T-t)}F(1-\Phi(d(t)-\sigma_I\sqrt{T-t})) \tag{9.23}$$

we obtain the following relations:

$$\theta(t,I(t)) > 1, \quad \theta(t,I(t)) \to \begin{cases} +\infty, \text{ for } I(t) \to 0 \\ 1, \text{ for } I(t) \to +\infty \end{cases} \tag{9.24}$$

for example, the absolute value of the optimal portfolio process for the portfolio problem (OP1) is always bigger than the one of the corresponding basic problem (P). To interpret this, we look at the following two

Special cases

a) In the special case of an inflation-linked zero coupon bond, for example, for

$$C_i = 0, \quad i = 1, \ldots, n$$

we obtain

$$\psi_1(t) = \frac{F\exp(-r_R(T-t))N(d(t))}{I(t_0)} \tag{9.25}$$

$$B_{IL}(t,I(t)) = F\left(\exp(-r_N(T-t)) + \frac{C_I(t,I(t))}{I(t_0)}\right) \tag{9.26}$$

which would lead to the same limiting behavior of the quotient $\theta(t, I(t))$ as in relation (9.24).

In the special case of an inflation-linked bond *without deflation protection*, for example, a usual inflation-linked coupon bond where the final payment of the notional has the form:

$$F\frac{I(T)}{I(t_0)} \tag{9.6*}$$

we obtain:

$$\psi_1(t) = \sum_{i:t_i>t} \frac{C_i}{I(t_0)} \exp(-r_R(t_i - t)) + \frac{F \exp(-r_R(T-t))}{I(t_0)} \tag{9.27}$$

$$B_{IL}(t, I(t)) = \sum_{i:t_i>t} C_i \frac{I(t)}{I(t_0)} \exp(-r_R(t_i - t)) + F \frac{I(t)}{I(t_0)} \exp(-r_R(T-t)) \tag{9.28}$$

leading to:

$$\pi_1(t) = \frac{\lambda\sigma_I - r_R}{(1-\gamma)\sigma_I^2} \frac{B_{IL}(t, I(t))}{\psi_1(t) I(t)} = \frac{\lambda\sigma_I - r_R}{(1-\gamma)\sigma_I^2} \tag{9.29}$$

for example, we have the same optimal portfolio process as in the basic problem (P). This is not surprising as the inflation-linked bond without deflation protection is simply a linear product with regard to the inflation index which therefore can be identified as a tradeable good.

If we now put the insights from the two special cases together then we note that the higher absolute value of the optimal portfolio process for (OP1) compared to (P) has its reason in the protection against deflation. As for small values of the inflation index the total payment of the inflation linked bond is typically dominated by the final payment, the price of the inflation linked bond then behaves more like a nominal bond. To mimic the optimal stock position, therefore more and more units of the inflation linked bond have to be sold short.

The following numerical example illustrates the above discussion. In Figure 9.1 the simulated path of the inflation index is presented for the time interval $t = [0, 30]$, where the time unit represents one year. The inflation index is assumed to follow the geometric Brownian motion of (9.10) with the following parameters: nominal and real interest rates, $r_N = 0.07$ and $r_R = 0.05$ respectively, market price of risk $\lambda = 0.3$, volatility $\sigma_I = 0.20$ on the yearly basis and $I(0) = 100$ (the seemingly high value of $\lambda\sigma_I$ is chosen for demonstrational purposes to obtain positive values for the optimal fractions of inflation products later on).

In Figure 9.2 we present the optimal portfolio processes for three different portfolio problems, each characterized by the structure of the inflation-linked bond available as investment opportunity. These different structures are an inflation-linked zero coupon bond with deflation protection (9.26), an inflation-linked bond without deflation protection (9.28) and an inflation-linked bond with deflation protection (9.5). The inflation-linked bonds are assumed to have the following characteristics: face value $F = 100$, date of maturity $T = 30$, coupon payments $C_i = 10$ and coupon dates $t_i = i$, where

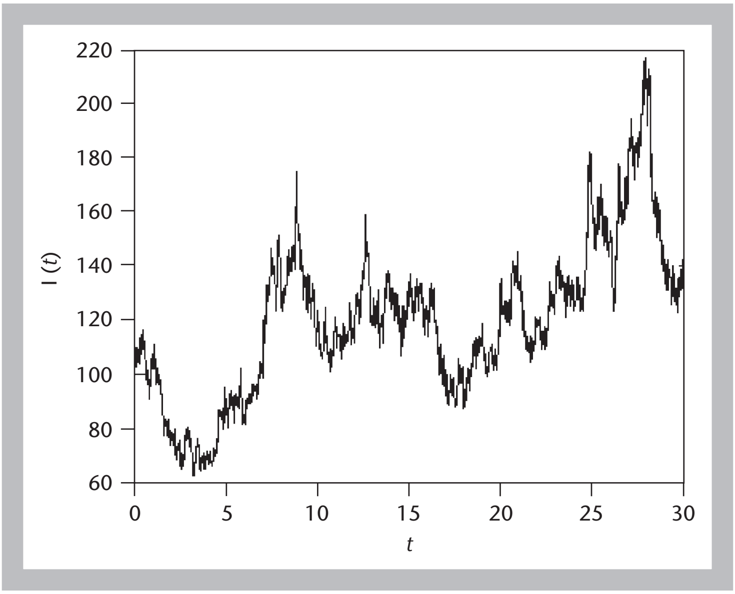


Figure 9.1 Simulated path of an inflation index as a geometric Brownian motion

$i = 1, 2, \ldots, 30$, as well as the reference inflation index of $t_0 = 100$. The optimization is done for the HARA utility function with $\gamma = 0.5$. The paths shown in Figure 9.2 of course correspond to the simulated path of the inflation index of Figure 9.1.

As already stated, the figure illustrates the fact if an inflation-linked bond without deflation protection is used the optimal portfolio process for problem (P) is constant. For the other two inflation-bonds that are deflation-protected at par one can observe that the optimal portfolio of the zero inflation-bond is always higher than the one of the inflation-bond containing coupon payments, which are not deflation-protected, due to the fact that the zero inflation-bond's structure is totally option-like, where as including coupon payments creates a mixture of the inflation index (stock) and an option-like inflation-bond.

Further, the opposite movements of the inflation index and the optimal portfolio processes of inflation-linked bonds with deflation protection can be detected. The optimal pure fraction of the bond and inflation index (taking over the role of the stock) in the portfolio is given by the solution of the basic problem (P) and coincides with the optimal portfolio of the inflation-linked bond without deflation protection. An inflation-linked bond with deflation protection can be seen as the combination of the inflation index (replicating strategy (9.20)) and the conventional bond (the remaining part of (9.5)), for example, can be replicated by the inflation index and this conventional bond.

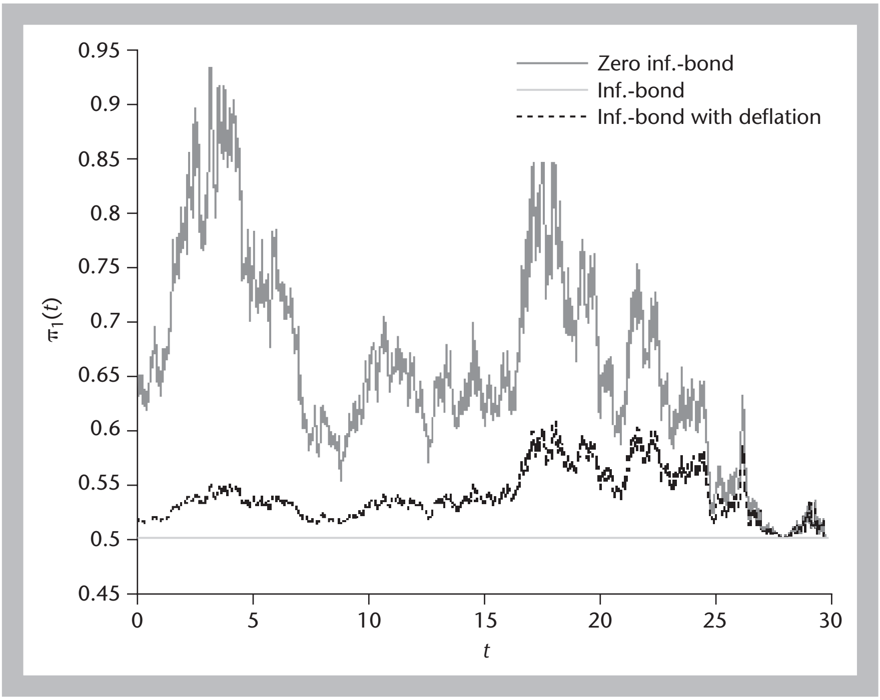


Figure 9.2 Optimal portfolio process of inflation-linked bonds with different structure characteristics

In order to maintain this optimal pure fraction of the bond and inflation index (stock) in the portfolio, one has to increase the relative fraction of the inflation-bond in the portfolio, when the inflation index is getting lower, because the replicating strategy (9.20) becomes smaller, too.

The other interesting aspect of this example is the asymptotic behavior of the optimal portfolio processes of inflation-linked bonds with deflation protection at the maturity date T. Depending on the level of the inflation index $I(t)$ the price of the inflation-linked bond $B_{IL}(t, I(t))$ approaches three different asymptotical forms. In the case of inflation, for example, when $I(t) > I(t_0)$, the asymptote is equal to $\psi(t)I(t)$. For $I(t) = I(t_0)$ we have $\psi(t)I(t) + F/2$ as an asymptote and in the case of deflation, for example, $I(t) < I(t_0)$, the asymptotic function for $B_{IL}(t, I(t))$ is $\psi(t)I(t) + F$. In these different cases the asymptotic representations of the optimal portfolio process $\pi_1(t)$ are:

$$\pi_1(t) \sim \frac{\lambda\sigma_I - r_R}{(1-\gamma)\sigma_I^2}, \quad t \to T, \quad I(t) > I(t_0)$$

$$\pi_1(t) \sim \frac{\lambda\sigma_I - r_R}{(1-\gamma)\sigma_I^2}\left(1 + \frac{F}{2\psi(t)I(t)}\right), \quad t \to T, \quad I(t) = I(t_0)$$

$$\pi_1(t) \sim \frac{\lambda\sigma_I - r_R}{(1-\gamma)\sigma_I^2}\left(1 + \frac{F}{\psi(t)I(t)}\right), \quad t \to T, \quad I(t) < I(t_0)$$

This asymptotic behavior of the optimal portfolio process $\pi_1(t)$ shows the same feature as the inverse dependence between the optimal portfolio process and inflation index $I(t)$. Having a higher probability for deflation, for example, $I(t) < I(t_0)$, the related fraction of the inflation-linked bond, for example, the optimal portfolio process $\pi_1(t)$, has a higher value compared to the situation with a lower deflation probability, for example, $I(t) > I(t_0)$.

Problem 2: Inflation-linked bond, stock and non-inflation linked bond.

In addition to the setting in Problem 1 the investor can now also invest into a stock with price given by equation (9.9). For simplicity we now assume that the final payment has no protection against deflation and that inflation and the stock price evolution are independent. More precisely, we look at the problem:

$$\max_{\varphi(.)\in B(x)} E(U(X^{\varphi}(T))) \tag{OP2}$$

with $X^{\varphi}(t) = \varphi_0(t)P_0(t) + \varphi_1(t)P_1(t) + \varphi_2(t)B_{IL}(t)$ where $B_{IL}(t)$ is given as in Equation (9.28). To solve Problem (OP2) we first determine the replication strategies of the stock (by itself) and the inflation-linked bond as:

$$\begin{aligned} \psi_1(t) &= 1 \\ \psi_2(t) &= \sum_{i:t_i>t} \frac{C_i}{I(t_0)} \exp(-r_R(t_i - t)) + \frac{F\exp(-r_R(T-t))}{I(t_0)} \end{aligned} \tag{9.30}$$

Application of Theorem 1 then yields the optimal trading strategy in stock and the inflation-linked bond as:

$$\begin{aligned} \varphi_1(t) &= \frac{\xi_1(t)}{\psi_1(t)} = \frac{b - r_N(t)}{(1-\gamma)\sigma_1^2} \cdot \frac{X(t)}{P_1(t)} \\ \varphi_2(t) &= \frac{\xi_2(t)}{\psi_2(t)} = \frac{\lambda\sigma_I - r_R(t)}{(1-\gamma)\sigma_I^2} \cdot \frac{X(t)}{\psi_1(t)I(t)} \end{aligned}$$

or in terms of the portfolio process as:

$$\pi_1(t) = \frac{b - r_N}{(1-\gamma)\sigma_1^2}, \qquad \pi_2(t) = \frac{\lambda\sigma_I - r_R}{(1-\gamma)\sigma_I^2} \tag{9.31}$$

We thus have the same optimal portfolio process as in the basic problem (P). This is not a surprise as our traded products are just linear functions of the ones traded in (P). As the portfolio process only describes the fraction of the total wealth invested in the different products we should therefore have this coincidence which of course disappears for the inflation-linked bond and the inflation index if we would compare the actual amount of units of both products we optimally have to hold in both portfolio problems.

Remarks

(a) Looking at the explicit form of the optimal fraction of wealth in the inflation-linked product as given in relation (9.31), we realize that unless we have a very strong opinion for a very high inflation rate (for example, the excess return of the inflation should be higher than the real interest rate) it would be optimal to sell inflation-linked bonds short.

(b) In the more general case of an inflation-linked bond with deflation protection and a non-vanishing correlation between the inflation index and the stock price, explicit calculation shows that we obtain the following optimal portfolio processes:

$$\begin{bmatrix} \varphi_1(t) \\ \varphi_2(t) \end{bmatrix} = (\psi(t)')^{-1}\bar{\xi}(t)$$
$$= \frac{1}{1-\gamma} X(t) \begin{bmatrix} \dfrac{b-r_N}{\sigma_1^2 P_1(t)} - \dfrac{\lambda\sigma_I - r_R}{\sigma_I^2 P_1(t)} \cdot \dfrac{\sigma_2\sigma_I}{\sigma_1^2} \\ \dfrac{\lambda\sigma_I - r_R}{\sigma_I^2 I\psi_{22}(t)} \cdot \left(1 + \dfrac{\sigma_2^2}{\sigma_1^2}\right) - \dfrac{b-r_N}{\sigma_1^2 I(t)\psi_{22}(t)} \cdot \dfrac{\sigma_2\sigma_I}{\sigma_I^2} \end{bmatrix} \tag{9.32}$$

with $X(t)$ denoting the optimal wealth process of both the basic portfolio problem (P) and the just considered optimal option portfolio problem and where the replication strategies are given by:

$$\psi(t) = \begin{bmatrix} 1 & 0 \\ 0 & \sum\limits_{i:t_i>t}^{n} \dfrac{C_i}{I(t_0)} \exp(-r_R(t_i - t)) + \dfrac{F\exp(-r_R(T-t))N(d(t))}{I(t_0)} \end{bmatrix} \tag{9.33}$$

Hence, the moral of both remarks is that the type of risk-averse investors we are considering are typically selling inflation-linked bond products short. However, there should be market participants which are interested in purchasing such inflation-linked products as otherwise there is no use in offering them at all. Therefore, in the next section we are presenting a situation where those products are needed.

9.4 HEDGING WITH INFLATION-LINKED PRODUCTS

As we have seen in the last section, a risk-averse investor is usually not attracted by inflation linked products of the type we considered here. The only exception is if he is pretty sure that there will be a tendency for a huge increase of the inflation index expressed via a high risk premium. On the other side, there are companies that have to hedge inflation to do their original business. Typical such candidates for buying the above inflation-linked products are (usually non-life) insurance companies who are facing a risk process due to their insurance business that is closely linked to the inflation as the value of the insured goods are obviously related to it. To study one of their relevant problems we consider a financial market consisting of a riskless bond, a stock and some inflation-linked product where we assume the following price processes (for simplicity we assume the inflation index to be tradable which – as seen – is equivalent to the assumption that an inflation-linked bond without deflation protection is traded):

$$dP_0(t) = P_0(t) r_N \, dt, \qquad P_0(0) = 1 \tag{9.34}$$

$$dP_1(t) = P_1(t)(\mu dt + \sigma_{11} dW_1(t)), \quad P_1(0) = p_1 \tag{9.35}$$

$$dI(t) = I(t)((r_N - r_R + \nu)dt + \sigma_{21} \, dW_1\,(t) + \sigma_{22} \, dW_2(t)) \quad I(0) = i \tag{9.36}$$

We assume that an insurer wants to hedge a payment that is inflation related (for example, payments arising from a car insurance) and has the form,

$$\tilde{B} = B \cdot I(T)$$

where B is independent of the Brownian motion W and represents the value of the insurance premium that has to be paid out if the insurance case would happen today. Further, B can be interpreted as the total outcome of the risk process of the insurance company until the time horizon. The hedging problem we are considering is

$$\min_{\pi} E(\tilde{B} - X^{\pi}(T))^2 \tag{9.37}$$

where the utility criterion can be simplified to

$$E(\tilde{B}^2) - 2E(B)E(I(T)X^{\pi}(T)) + E(X^{\pi}(T)^2) \tag{9.38}$$

for example, it is sufficient to consider the utility maximization problem with the following (random) utility function:

$$U(x) = U(x; I(T)) = -x^2 + 2cx \cdot I(T) \tag{9.39}$$

with a suitable constant c. Note that this utility function – although strictly concave – is not strictly increasing, but has an invertible derivative. As in Korn (1997) one can show that the usual procedure of the martingale approach to portfolio optimization (see Korn (1997b)) is still valid but cannot ensure a non-negative final wealth process. Let us first consider the case of:

9.4.1 Investment in bond, stock and inflation

In this setting we have a complete market. If the investor can afford it then he will use a trading strategy that will minimize the utility function $U(x)$ of (9.39) pathwise, for example, his final wealth should satisfy

$$X^{\pi}(T) = cI(T) \tag{9.40}$$

which – by completeness of the market – the investor can exactly attain with an initial wealth of

$$\tilde{x} = cE(H(T)I(T)) = I(0)E(B) \tag{9.41}$$

where $H(T)$ is given by:

$$H(T) = \exp(-(r_N + ½\|\theta\|^2)T - \theta' W(T)),$$

$$\theta := \sigma^{-1}\left(\begin{pmatrix} \mu \\ r_N - r_R + \nu \end{pmatrix} - r_N \underline{1}\right) \tag{9.42}$$

In the case of $x \geq \tilde{x}$ it is not necessary for the investor to use more money than $\tilde{x}$ for hedging activities as this would lead to a deviation from the optimal final wealth of (4.7). We therefore have proved:

> **Proposition 3** Let $\tilde{x}$ be defined as in (9.41) and assume $x \geq \tilde{x}$. Then, the optimal final wealth for the hedging problem (9.37) is given by relation (9.40), the optimal hedging strategy is to buy c units of the inflation index (respectively a suitable inflation product delivering the same final payment) and to hold it until maturity T. The corresponding minimal quadratic hedging error equals:
>
> $$Var(B)E(I(T)^2) \tag{9.43}$$
>
> The remaining money $x - \tilde{x}$ can be used for different purposes.

Proof: In light of the arguments preceding Proposition 3 only (9.43) has to be shown, but this follows directly from the final wealth of the form of relation (9.40) into the utility function (9.38).

In the case of $x < \tilde{x}$ we are in the situation of Korn (1997a) and can use the martingale approach. Note first that we have:

$$U'(x) = 2c \cdot I(T) - 2x$$

which implies (with the notation of Korn (1997a)):

$$\tilde{I}(y) = (U')^{-1}(y) = cI(T) - ½\, y$$

leading to

$$\mathrm{X}(y) = E(H(T)\tilde{I}(yH(T))) = E(cH(T)I(T)) - ½\, yE(H(T)^2)$$

$$Y(x) = (\mathrm{X})^{-1}(x) = 2(E(cH(T)I(T)) - x)/E(H(T)^2)$$

As in Korn (1997a) we then obtain:

Proposition 4 In the case of $x < \tilde{x}$ the optimal final wealth for problem (9.37) is given by:

$$B^* = I(T)E(B) - (E(B)I(0) - x)\frac{H(T)}{E(H(T)^2)} \tag{9.44}$$

with a minimal quadratic hedging error of:

$$Var(B)E(I(T)^2) + \frac{(E(B)I(0) - x)^2}{E(H(T)^2)} \tag{9.45}$$

Proof: The form of B^* follows from the main result of Korn (1997a) as we there have:

$$B^* = \tilde{I}(Y(x)H(T)) = E(B)I(T) - E(E(B)H(T)I(T) - x)\frac{H(T)}{E(H(T)^2)}$$

Using $E(H(T)I(T)) = I(0)$ and B^* instead of B in (9.38), simplifying the resulting expression leads directly to the minimal hedging error as given in (9.45).

Remark (i) Note that the hedging error above consists of two different components. First, there is the unavoidable error term $Var(B)E(I(T)^2)$ which only vanishes if the height of the premium (more precisely, the part not depending on price changes due to inflation) is exactly foreseeable. The remaining part

$$\frac{(E(B)I(0)) - x)^2}{E(H(T)^2)}$$

is non-negative, but can vanish depending on the amount of money x available for hedging activities (for example, for $x = E(B)I(0)$) which also shows that Propositions 3 and 4 are consistent for exactly this choice of x).

(ii) The form of the hedging error obtained in Proposition 3 is quite natural as due to the independence of B and the capital market, $E(B)$ is the best forecast for B made up out of the actions at the capital market, and $E(I(T)^2)$ represents the minimal possible uncertainty due to inflation.

To demonstrate the effect that the use of the inflation index has on the hedging error we also have to solve problem (9.37) if we are not allowed to trade the inflation index (or any other inflation-linked product – besides the stock, of course). We will reduce this problem to solving again a (modified) quadratic problem but now in the market that consists only of the bond and the stock.

9.4.2 Investment in bond and stock

If we can only invest in bond and stock then we cannot hedge the randomness that is caused by $W_2(.)$. We therefore introduce:

$$I(t) = \hat{I}(t)\exp(-\tfrac{1}{2}\sigma_{22}^2 t + \sigma_{22}W_2(t)) \tag{9.46}$$

and thereby directly obtain that the utility criterion in this reduced market equals:

$$E(\tilde{B}^2) - 2E(B)E(\hat{I}(T)X^{\pi}(T)) + E(X^{\pi}(T)^2) \tag{9.47}$$

for example, we can now solve a conventional utility maximization problem in the complete market made up of the stock and the bond with:

$$U(x) = U(x;\hat{I}(T)) = -x^2 + 2cx \cdot \hat{I}(T) \tag{9.48}$$

$$\hat{H}(T) = \exp(-(r_N + \tfrac{1}{2}\hat{\theta}^2)T - \hat{\theta}W_1(T)),\ \hat{\theta} = (\mu - r_N)/\sigma_{11} \tag{9.49}$$

$$\hat{x} = cE(\hat{H}(T)\hat{I}(T)) = I\,(0)\exp\left(\left(-r_R + \nu - (\mu - r_N)\,\frac{\sigma_{21}}{\sigma_{11}}\right)T\right)E\,(B) \tag{9.50}$$

Exactly the same arguments as in the case (i) lead to:

Proposition 5 Let $\hat{x}$ be defined as in (9.50).

(i) For $x \geq \hat{x}$, the optimal final wealth for the hedging problem (4.4) if only investment in bond and stock is allowed is given by:

$$E(B)\,\hat{I}(T) \tag{9.51}$$

with a corresponding minimal quadratic hedging error equalling,

$$Var(B)E(I(T)^2) + (E(B))^2 E((I(T) - \hat{I}(T))^2) \tag{9.52}$$

The remaining money $x - \hat{x}$ can be used for different purposes.

(ii) In the case of $x < \hat{x}$ the optimal final wealth for problem (9.38) is given by:

$$\hat{B}^* = \hat{I}(T)E(B) - E(E(B)\hat{H}(T)\hat{I}(T) - x)\frac{\hat{H}(T)}{E(\hat{H}(T)^2)} \tag{9.53}$$

with a minimal quadratic hedging error of:

$$Var(B)E(I(T)^2) + (E(B))^2 E((I(T) - \hat{I}(T))^2) + \\ + \frac{(E(B)E(\hat{H}(T)\hat{I}(T)) - x)^2}{E(\hat{H}(T)^2)} \tag{9.54}$$

Proof: The form of the optimal final wealth in both cases follows exactly as in the situations of Propositions 3 and 4, but now in the reduced market consisting only of the bond and stock. To show the form (9.52) of the hedging error, note:

$$\begin{aligned} E[BI(T) - E(B)\hat{I}(T)]^2 &= E[BI(T) - E(B)I(T) + E(B)(I(T) - \hat{I}(T))]^2 \\ &= Var(B)E(I(T)^2) + (E(B))^2 E((I(T) - \hat{I}(T))^2) \\ &\quad + 2E(B - E(B))E(I(T)(I(T) - \hat{I}(T))) \\ &= Var(B)E(I(T)^2) + (E(B))^2 E((I(T) - \hat{I}(T))^2) \end{aligned}$$

With this, the form of the hedging error (9.54) follows as in Proposition 4.

Remark Note that for $\sigma_{11} = 0$ the results of Proposition 3 and 4 coincide with those of Proposition 5. The form of both the hedging errors and the optimal final wealths indicate that they are the results of two succeeding projections. First, $\tilde{B}$ is projected onto the market that allows for a perfect replication of inflation (see Propositions 3 and 4) and then to the market that only allows a partly hedging of inflation via trading stock and bond (see Proposition 5). This becomes most transparent when we are in case (i) of Proposition 5 where the two sums making up the hedging error exactly correspond to this interpretation.

9.4.3 Numerical examples

We are now illuminating the gains for our hedging problem by using inflation-linked products via some numerical examples. Note that to perform our computations we need the following explicit forms of the hedging errors in the above propositions:

- The quadratic hedging error H_1 corresponding to B^* (trading in bond, stock, inflation) is given by:

$$H_1 = Var(B)I(0)^2 \exp((2(r_N - r_R + \nu) + \sigma_{21}^2 + \sigma_{22}^2)T) + \frac{(I(0)E(B) - x)^2}{\exp((\theta'\theta - 2r_N)T)}$$

$$\text{with } \theta = \begin{pmatrix} \dfrac{\mu - r_N}{\sigma_{11}} \\ \dfrac{\nu - r_R}{\sigma_{22}} - \dfrac{\mu - r_N}{\sigma_{11}} \cdot \dfrac{\sigma_{21}}{\sigma_{22}} \end{pmatrix}$$

- The quadratic hedging error H_2 corresponding to $\hat{B}^*$ ("trading in bond and stock") is given by:

$$H_2 = Var(B)I(0)^2 \exp((2(r_N - r_R + \nu) + \sigma_{21}^2 + \sigma_{22}^2)T)$$
$$+ I(0)^2 \exp((2(r_N - r_R + \nu) + \sigma_{21}^2)T)(E(B))^2(\exp(\sigma_{22}^2 T) - 1)$$
$$+ \frac{(I(0)E(B) \exp((-r_R + \nu - \theta\sigma_{21})T) - x)^2}{\exp((\theta^2 - 2r_N)T)}$$

with $\theta = \dfrac{\mu - r_N}{\sigma_{11}}$

It is now easiest to see the hedging effect of using inflation products by considering a deterministic B which we therefore assume to equal 1. We will further choose:

$$\mu = 0.1, \quad \sigma_{11} = 0.3, \quad r_N = 0.04, \quad \nu = 0.01, \quad \sigma_{22} = 0.04 \quad I(0) = 100$$

and vary x, σ_{21}, r_R and report the corresponding hedging errors in Table 9.1 where we choose $r_R = 0.03$ (panel A) and $r_R = 0.05$ (panel B) respectively.

What can be clearly seen is that with tradable inflation the hedging error is smaller than if we can only use the stock for hedging inflation. The error increases with decreasing covariance between inflation and the stock. Thus, we have demonstrated that there are indeed investors who have advantages

Table 9.1 Corresponding hedging errors

Panel A						
σ_{21}	**0.04**		**0**		**−0.04**	
x	H_1	H_2	H_1	H_2	H_1	H_2
100	0	16.69	0	16.67	0	16.69
95	15.94	21.91	20.26	26.16	23.79	31.78
90	63.76	71.23	81.06	83.61	95.12	97.42

Panel B						
σ_{21}	**0.04**		**0**		**−0.04**	
x	H_1	H_2	H_1	H_2	H_1	H_2
100	0	16.04	0	16.01	0	16.04
95	6.16	16.14	9.57	17.22	13.72	19.60
90	24.66	45.42	38.29	54.47	54.88	64.89

from buying inflation-linked products and so there are counterparts to the risk-averse investors of section 3.

9.5 CONCLUSION

We have presented a simple modeling framework for pricing and dealing with inflation-linked products. After having derived Black–Scholes-type pricing formulae in section 9.2, we considered standard portfolio problems for investing in bond, stocks and inflation-linked products which could be solved by the martingale method. As a consequence we obtained that typically risk-averse investors sell standard inflation-linked products (such as inflation-linked coupon bonds) short. The necessary counterparts for creating an active market of inflation-linked products are investors that have to hedge inflation related payments (such as insurance companies), a fact which is demonstrated in section 9.4.

Future research will be centered around more detailed models including perhaps stochastic interest rates as in Jarrow and Yildirim (2003) or in general incomplete markets.

REFERENCES

Ang, A. and Bekaert, G. (2003) "The Term Structure of Real Interest Rates and Expected Inflation", Working Paper, Columbia University, New York.

Fisher, I. (1930) *The Theory of Interest* (London/Basingstoke: Macmillan).

Jarrow, R. and Yildirim, Y. (2003) "Pricing Treasury Inflation Protected Securities and Related Derivatives Using an HJM Model", *Journal of Financial and Quantitative Analysis*, 38(2): 409–30.

Korn, R. (1997a) "Some Applications of L^2–Hedging with a Non-Negative Wealth Process", *Applied Mathematical Finance*, 4(1): 64–79.

Korn, R. (1997b) *Optimal Portfolios* (New York, NY: WorldScientific).

Korn, R. and Korn, E. (2001) *Option Pricing and Portfolio Optimization* (AMS).

Korn, R. and Kruse, S. (2004) "Einfache Verfahren zur Bewertung von inflationsgekoppelten Finanzprodukten", *Blätter der DGVFM*, XXVI(3): 351–67.

Korn, R. and Trautmann, S. (1999) "Optimal Control of Option Portfolios", *OR-Spektrum*, 21(1–2): 123–46.

Kraft, H. (2003). "Elasticity Approach to Portfolio Optimization", *Mathematical Methods of Operations Research*, 58: 159–82.

Nielsen, H. (2003) "Inflation Expectations in the EU – Results from Survey Data", CASE Working Paper.

CHAPTER 10

Model Risk and Financial Derivatives

François-Serge Lhabitant

10.1 INTRODUCTION

Since the introduction of option trading in Chicago in 1973, derivatives have shaped the evolution of capital markets by allowing efficient risk unbundling and transfer. Financial intermediaries immediately recognized that derivatives were the perfect tool to customize state-contingent payoffs for both speculators and hedgers alike. Consequently, the volume and various types of derivative contracts traded on organized exchanges as well as in over the counter markets have grown steadily. The catalysts of this success were the development of financial theory and sophisticated pricing mathematical models, the availability of real-time information, the technological innovation (in particular increasingly powerful computers) as well as the move from open-outcry trading to electronic trading. Today, we have reached the point where derivatives have become an essential feature of practically any financial contract. They have changed the way companies and individuals make investments, raise capital, and even measure, manage and understand risk.

The fundamental risks associated with derivatives (for example, credit, market, operational and legal risks, among others) are no different from those that many financial institutions and firms face in their traditional businesses. However, the risks associated with derivatives can be far more complex to assess and manage due to their dynamic nature, the asymmetry of their payoffs and the implicit leverage of derivatives positions. Unlike stocks or bond traders, derivatives traders cannot sell an option contract, hedge it, place the paperwork in a drawer and forget about it. Indeed, there is a need to constantly monitor the corresponding commitments and adjust the hedge dynamically. Therefore, the need to accurately price and hedge derivative contracts throughout their life – and, more generally, to develop models to manage their risks – rapidly became self-evident among the financial community.

It is therefore not surprising to see that modelling has also significantly evolved alongside derivatives markets. A few decades ago, the derivatives industry was lacking sophisticated models, and traders primarily relied on experience and intuition, with mixed results. Today, there is a proliferation of sophisticated models, which includes well-established models from academia, proprietary models developed for internal use by leading-edge financial institutions, and third-party applications intended for sale or distribution (for example, commercial software). A standard off-the-shelf financial package typically contains between five and twenty models to value the same option, whereas a proprietary derivatives pricing software in an investment bank might contain several hundred models. As a result, derivatives traders simultaneously utilize a seemingly disparate collection of models to perform similar or related tasks. These models may be inconsistent with each other, or even inadequate for the task they are used for. Nevertheless, their results are often taken for granted, aggregated, compared and used for making strategic and tactical decisions.

For academics, the abundance of models is clearly unsatisfactory. For some practitioners, it is simply confusing. For others, it is just a warning sign of an early developmental stage in the modelling technology. For all, it should be a signal that a new type of risk has emerged, namely: model risk. As summarized by Robert C. Merton in his Nobel Prize address, "The mathematics of financial models can be applied precisely, but the models are not at all precise in their application to the real world." The abundance of models and the imperfect assumptions and hypothetical solutions that go into them result in model errors becoming more of a probability than a remote possibility. This hazard is called model risk, and it is generally categorized as a form of operational risk.

Broadly speaking, model risk encompasses all financial losses directly attributable to the use of a quantitative model. Derman (1996) lists several examples of situations associated with model risk, for example, inapplicability of modelling, incorrect model, correct model with incorrect solution, correct model with inappropriate use, badly approximated solution, software and hardware bugs, or unstable data. As illustrated by Green and Figlewski (1999) or Gibson, Lhabitant, Pistre and Talay (1999), model risk may lead to a number of problems for financial institutions, such as derivatives trading too low or too high in price, incorrect hedging strategies, liquidity issues, etc.

Table 10.1 A few examples of model risk and its consequences

Period	Institution	Problem	Loss (M)
1970s	Merrill Lynch	Pricing of bond issues with an inadequate US interest rate curve.	US$ 70
1970	Merrill Lynch	Use of an incorrect yield curve to price stripped Government bonds	US$ 70
1987	Merrill Lynch	Incorrect pricing model for stripped mortgage-backed securities	US $350
1992	J.P. Morgan	Inadequate models of prepayments for mortgage-backed securities.	US$ 200
1997	Natwest Markets	Pricing of options based on a naïve volatility feed.	GBP 90
1997	Bank of Tokyo-Mitsubishi	Inadequate calibration of a model for swaptions.	US$ 83
1997	UBS AG	Inadequate pricing model for structured equity derivatives	CHF 120
1998	LTCM	Inadequate models for arbitrage between US and European interest rates, and excessive leverage	Unknown
2001	Lipper Convertible Hedge Fund	Inadequate mark to model for convertible bond arbitrage.	US$ 600
2003	Fourth District Institution Provident Bank	Modelling error in its auto-leasing business identified 7 years later	US$ 67
2005	Various hedge funds	Market reaction to the GM and Ford credit downgrades too improbable an event for credit risk models to capture it	Unknown

Far from remaining a theoretical concern, the percentage of losses attributed to model risk has been consistently increasing over recent years, and the deficiencies in current quantitative models continue to be brought to light by market events (see Table 10.1). In 1999, a report by Capital Markets Risk Advisors (CMRA) and Meridien Research estimated the annual model risk losses reach $5.5 billion. Since, it is almost *de rigueur* to have a major model risk-related disaster every year. The loss may be absorbable, but the acute embarrassment is not. It is, therefore, essential for model users and builders to be aware of the existence of model risk, to understand its potential sources and to implement business practices and technology solutions to mitigate it.

In this chapter, we discuss the potential impact on model risk in the particular case of derivatives pricing and hedging. Section 10.2 recalls the evolution of pricing models for derivatives. Section 10.3 illustrates the example of implied volatility – a necessary input in most pricing models that is in fact a patch against model errors – while section 10.4 discusses the role that financial models play for derivatives contracts. Section 10.5 reviews the different steps in the model building process where model risk can start. Section 10.6 presents a short technical study of how things can turn wrong when an option trader uses the wrong model to hedge his or her book, while section 10.7 contains a series of (often forgotten) common sense rules to manage and mitigate model risk. Section 10.8 summarizes our findings and opens the way to future research.

10.2 FROM MATHEMATICAL THEORY TO FINANCIAL PRACTISE

Derivatives pricing finds its roots in the doctoral thesis of Louis Bachelier (1900), which developed the first analytical model for the valuation of financial options. Unfortunately, Bachelier's theoretical contribution was too innovative for his time. Consequently, his peers essentially focused on the weaknesses of his model – normally distributed asset prices allow for negative security prices and may result in call option prices that exceed the price of the underlying asset. Therefore, Bachelier was granted his doctoral degree, but he was only offered a chair in a second-tier university, and his work remained undiscovered for more than fifty years. In a sense, Bachelier became the first publicly known victim of financial model risk, and quantitative finance went back to sleep.

One has to wait until the late 1960s to see quantitative work laying again some foundations in finance, with Markowitz's (1959) and Sharpe's (1966) works on portfolio selection and modern portfolio theory. But the major event was undoubtedly the publication of the option-pricing model developed by Black and Scholes (1973) and Merton (1973). Though mathematically complex, their formula was directly applicable, easy to understand and only required a series of rather straightforward inputs: the price of the underlying asset, the strike price, the time to maturity, the volatility of the underlying asset and the risk-free interest rate. Moreover, their model came out simultaneously with the opening of option trading at the CBOT. Needless to say, it was an immediate success. Practitioners and option traders adopted the model as a useful tool for understanding what the price of an option should be, how to make money from mispriced options, and how to hedge an option book.

Since then, the interaction between mathematical theory and financial practice has never ceased. As the mathematical awareness in the financial research community increased, financial markets became more quantified and derivatives research actively evolved in three directions in order to improve the Black, Scholes and Merton framework. The first direction involved the relaxation of some of the original underlying assumptions, with a view toward developing a theory that accounts for market imperfections (for example, transaction costs, liquidity problems, feedback effects, etc.). The second direction focused on generalizing the price dynamics of the traded assets to include broader classes of stochastic processes, such as the so-called Levy processes and their extensions. Finally, the third direction started considering the case of more complex financial products, such as exotic options and structured products. In each of these directions, new quantitative models and techniques have been developed and applied, and this trend towards more financial engineering is likely to persist. There is no turning back, despite the fact that quantitative finance is regularly pronounced dead, particularly after major market events such as the crashes of 1987, 1994 or 1998.

Unfortunately, the level of complexity in the new financial models is also rising. The new science behind quantitative finance is relying more and more on applied probability theory and numerical analysis, and uses mathematical techniques such as stochastic calculus and partial differential equations to achieve its results. It also draws on wide areas of physics, notably heat diffusion and fluid mechanics where the dynamics are similar to those of financial markets.[1] As a side effect, not surprisingly, the mathematics encapsulating many of the more innovative derivatives is less and less accessible to the majority of market participants, including senior management. This resulted in the creation of a series of new models whose role is to help senior managers understanding the risks of instruments that are themselves heavily dependent on models. It is a vicious circle, which was summarized as follows by Alan Greenspan in one of his allocutions in March 1995: “The technology that is available has increased substantially the productivity for creating losses”, and empirical evidence showed that these losses could be significant.[2]

10.3 AN ILLUSTRATION OF MODEL RISK

The concept of implied volatility is probably among the best illustrations of what we mean by model risk. Consider for instance the case of the Black and Scholes option-pricing framework. The price of an option depends upon the price of the underlying asset, the exercise price, the time to maturity, the future volatility of the underlying asset, and the risk-free interest rate. All these parameters are observable, except the volatility parameter, which needs to be estimated. Now, how can one estimate something that is unobservable?

The usual way to price an option is to plug a volatility estimate into a pricing model (as well as all the others necessary observed variables) and thus obtain the corresponding option price. Alternatively, one can also “invert the model”. Starting from a market-quoted option price, one can compute the implied volatility assuming that the Black-Scholes formula is the correct pricing model for options.[3] This implied volatility figure can then be used as an input to other models, for instance to value more complicated options on the same underlying asset for which we have no market price or illiquid markets, typically over-the-counter derivatives.

In theory, there is nothing wrong with this type of approach. In fact, derivative prices observed at any given time should contain forward looking information on volatility. This means that the models used to price and hedge derivatives must be determined partially from econometric information and partially by solving “inverse problems” (in the sense of partial differential equations) that reflect current market prices. However, the problem starts when (a) we have several option prices available from which we can obtain an implied volatility, and (b) different options on the same underlying asset display different implied volatility. This phenomenon was originally called the smile, because a graph of the implied volatility against the strike of the corresponding option would typically look like a smile, deep in and out of the money options having higher implied volatility. After the 1987 crash, the smile disappeared and usually became a frown, with implied volatility generally decreasing as the strike price increase. As a result, in the money call options and out of the money puts tend to have prices that are above their respective theoretical Black and Scholes values, while out of the money calls and in the money puts are priced below their respective Black–Scholes values. Rubinstein (1994) attributed part of this phenomenon to crash-phobia, that is, investors valuing more out of the money puts because they fear a new crash. In addition to the frown, market participants often observe a term structure effect: options with the same strike price but with different maturities also tend to display systematic volatility patterns with respect to time to maturity.

In order to account for these biases with respect to the Black and Scholes’ constant volatility assumption, market participants started to build up models that accept a volatility surface rather than a single volatility number as an input. That is, depending upon the maturity date, the degree of moneyness of their options and whether it was a put or a call, market participants were considering different levels of volatility. These different levels were obtained by inverting the Black and Scholes model to yield a local volatility figure (see Figure 10.1).

From a practical perspective, the models that use an implied volatility surface are extremely convenient. They can explain and quantify the skewness and kurtosis in the empirical distribution of stock returns. They are consistent with different types of stochastic processes for the underlying asset. Moreover, they can be recalibrated to market data several times a day and their pricing errors are extremely small, which gives a false sense of security to their users. However, from a theoretical perspective, implied volatility surface models are by definition unsatisfactory. If the Black and Scholes assumptions hold, the volatility surface should be … flat. And if the

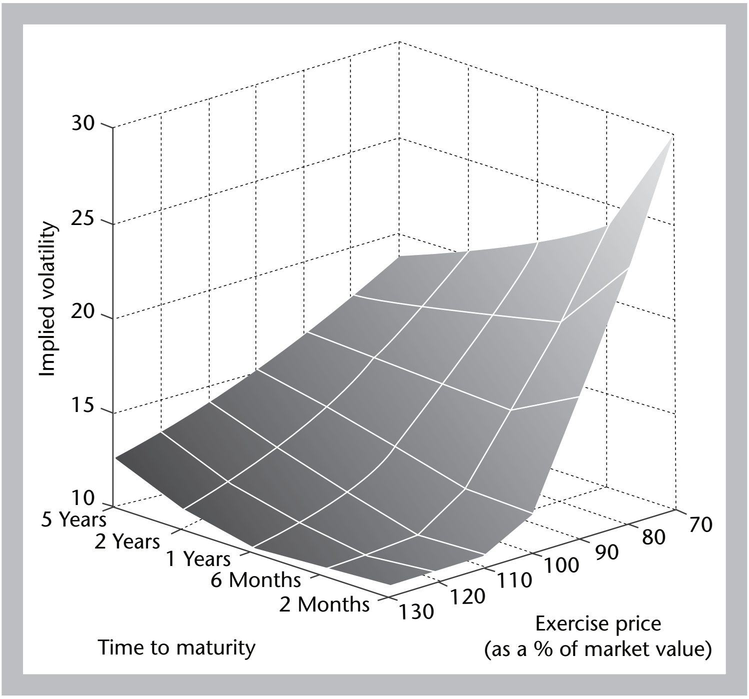


Figure 10.1 An example of the volatility surface observed on the Swiss market index

Black and Scholes assumptions do not hold, then any information implied by applying their model should by definition be considered suspicious. For a given option, the single number that we call "implied volatility" is, in some sense, just an amalgamation of all the unobservable, untradable, omitted and/or incorrectly specified parameters in the Black and Scholes model. Nevertheless, most financial institutions and researchers continue to use implied volatilities and even build more sophisticated models on top of them.[4]

10.4 THE ROLE OF MODELS FOR DERIVATIVES

Not all derivatives models were created equal. Depending on the role they are playing, models may be more likely or less likely to generate model risk. Three different perspectives should be considered.

10.4.1 A model is just shorthand

The first series of models are just convenient mathematical shorthand with no real meaning. They provide a useful tool to summarize a large amount of information in a few standardized and comparable numbers. This is the case, for instance, of the yield to maturity of a bond or the implied volatility of an option. In the latter case, a model (say the Black and Scholes formula) is a filter that turns quoted dollar values into an indicator of expected future volatility. As long as the result is taken cautiously, the quality of the model used plays a limited role. What really matters is the price the shorthand stands for, not the model itself. As an analogy, measuring and comparing distance using a biased meter does not matter as long as all distances are biased by the same factor. With such models, model risk is virtually non-existent or irrelevant. However, trouble may start if the model output is used as an input to another model (for example, relying on implied volatility to implement a hedging strategy).

10.4.2 A model is an approximation too

The second series of models are meant to serve as reasonable approximations or abstractions of some real-world behavior. They seek to explain relationships between observed phenomena and their generating process in an idealized way. They are less complicated than reality and hence easier to deal with. Their simplicity often lies in the fact that only the relevant properties of reality are represented, so that reasonable and acceptable approximations of reality are obtained for specific problems in ordinary situations. For instance, models that provide theoretical values to traders and investors and/or estimate risk exposures for risk managers enter into this category.

These models are usually popular and do not generate model risk per se. However, model risk can appear because of self-confidence. When a model seems to be consistent with the recurrence of similar expected results, we begin to become confident about its validity, assume that the model is correct, and often start using it as an extrapolation device outside of its definition range (for example, for rare and unusual situations) to predict and control the future. As an illustration, consider again the Black and Scholes model. As illustrated previously, it provides a useful pricing approximation for most at-the-money options close to expiry, but will give incorrect results for in-the-money and out-of-the-money options, resulting in the well known volatility smile.

10.4.3 A model is all that there is

In some extreme cases, a model may be the only tool available for valuation and/or decision-making. This is the case, for instance, with derivative instruments traded on OTC markets with only one market-maker, or with automated quoting and trading machines. Models in this latter category are, of course, the most likely candidates to generate model risk, since common sense is of little use.

10.5 THE MODEL-BUILDING PROCESS AND MODEL RISK-CREATION

Model risk is somehow similar to a virus: it is relatively easy to catch, and once in, it becomes extremely contagious. To understand how easy it is to introduce model risk within an institution, let us focus on the model-building process, that is, the procedure for the construction and verification of models for financial derivatives. A typical model-building process can be split into three steps: (1) model selection or creation; (2) model calibration; and (3) model usage. Each of these steps is capable of generating model risk.

10.5.1 Model selection/creation

In finance, most models have predictable regular features (deterministic) and unpredictable ones (stochastic). Since the unpredictable features are the ones which derivatives target, it is not surprising that the principal mathematical tool to build derivatives models relies heavily on probabilistic techniques and the theory of stochastic processes. As an illustration, the standard approach to option pricing consists of specifying the stochastic process followed by an underlying asset price and then deriving the option price as a function of the process parameters. Unfortunately, academic research has long stressed mathematical elegance as a key to quality.

Since analytic and closed-form solutions were the only noble outputs for option-pricing models, several researchers carefully selected the stochastic processes they were using in order to obtain closed-form solutions for their results. Sometimes, trade-offs were made between mathematical elegance and realism. Consequently, speculative prices underlying financial derivatives are not necessarily well represented by the few stochastic processes – and their related probability distributions – that are now commonly used in finance (for example, essentially normal and log-normal distributions, and diffusion-like processes). Financial time series exhibit highly non-trivial statistical features which are hard to model and even harder to explain, for example, intermittent behavior, volatility clustering (amplitudes of successive price movements are persistent, but not necessarily their signs), heavy tailed increments, and subtle dependence structures.

Nowadays, computer power is readily available, and closed-form solutions are considered to be a luxury that most practitioners cannot afford the time to find, thus preferring crude numerical schemes. Most of the time, these will work fine. However, numerical and poor semi-analytic methods inevitably carry their drawbacks. They produce discontinuities, which can be inherent numerical artefacts or genuine jumps in the portfolio sensitivities. The uniqueness (and meaningfulness) of a numerical solution should also imperatively be checked. Estimating the errors involved in a numerical scheme is a hard task, and numerical errors tend to accumulate and bias the final result.

10.5.2 Model calibration

The usefulness of a model and the value of its output are only as good as the model's ability to be effectively calibrated to its market environment. In the case of backward-looking models, information is available and there are numerous econometric techniques to estimate the necessary parameters and calibrate a model to market data. In the case of forward-looking models, validation can only occur after the fact, that is, when the authors will typically be unavailable. The calibration is then possible in only one way, by back-testing, that is using the spatial and statistical properties of the past to predict the present.

Although often neglected, the calibration stage is essential to detecting model risk. Indeed, the theory of parameter estimation generally assumes that the true model is known, and that the true model parameters must be estimated using a representative set of data. Are these properties verified in practice? Fitting a time-series model is usually straightforward nowadays in the use of appropriate computer software. However, model errors are likely to result in parameter instability. This was easy to observe in the case of simpler static models with observable parameters, but as soon as a model includes time-varying or stochastic parameters, these will absorb all the errors and output them as a simple change in value. Things get even worse when some of the quantities we are dealing with are pure abstractions, such as the expected future volatility. Even if we assume that this quantity is constant, how can we measure it?

Last but not least, instability in the calibration process can also result from numerical problems (such as near-singular matrix inversion) or from implementation problems: a model may require a large number of iterations to converge (a typical problem in Monte-Carlo simulations or in solving partial differential equations), may require a higher precision for floating-point numbers, or may use inappropriate approximations.

10.5.3 Model usage

Finally, model risk may arise during the model usage, even though all of the previous steps were correctly performed. For instance, some of the hypotheses of the model may simply not hold true in the real world, resulting in a model that performs poorly. As an illustration, a model may assume that zero-coupon bonds exist for all required maturities, while in practice the set of available maturity dates will be restricted. One should always remember that markets are driven by psychology, by supply and demand, by consensus, and not by simple diffusion processes.

10.6 WHAT IF THE MODEL IS WRONG? A CASE STUDY

When running a derivatives book, the first obvious impact of model risk is on pricing – model-based prices will diverge from observed ones. However, model risk also affects hedging, in a more subtle way. As an illustration, let us consider the case of the Black and Scholes (1973) framework. In a complete perfect market, a stock price $S(\mathrm{t})$ follows a geometric Brownian motion with constant volatility parameter and drift parameters:[5]

$$\frac{\mathrm{d}S(t)}{S(t)} = \mu \mathrm{d}t + \sigma \mathrm{d}W(t) \tag{10.1}$$

Equation (10.1) defines the true model, for example, the ones that rules the world. We denote by $C(t)$ the value at time t of a European call option with maturity T and exercise price K on this stock. By Ito's lemma, we obtain:

$$\mathrm{d}C(t) = \left(\frac{\partial C(t)}{\partial S(t)} \mu S(t) + \frac{\partial C(t)}{\partial t} + \frac{1}{2} \frac{\partial^2 C(t)}{\partial S^2(t)} \sigma^2 S^2(t) \right) \mathrm{d}t + \frac{\partial C(t)}{\partial S(t)} \sigma S(t) \mathrm{d}W(t) \tag{10.2}$$

Furthermore, we know that the call price $C(t)$ must satisfy the following partial differential equation:

$$\frac{\partial C(t)}{\partial S(t)} rS(t) + \frac{\partial C(t)}{\partial t} + \frac{1}{2} \frac{\partial^2 C(t)}{\partial S^2(t)} \sigma^2 S^2(t) - rC(t) = 0 \tag{10.3}$$

with boundary condition $C(T) = \mathrm{Max}\,[S(T) - K, 0]$.

Now, consider the case of a trader that is short one call option and needs to hedge his position. In theory, if he knew the "true" model, he could hedge perfectly in continuous-time by holding $\frac{\partial C(t)}{\partial S(t)}$ units of the underlying asset and $\left(C(t) - \frac{\partial C(t)}{\partial S(t)} S(t)\right)$ units of a zero-coupon bond maturing at time T. In the absence of arbitrage opportunities, the value of his overall portfolio $\Pi(t)$ is equal to zero, and its instantaneous variations are defined by:

$$\mathrm{d}\Pi(t) = -\mathrm{d}C(t) + \frac{\partial C(t)}{\partial S(t)} \mathrm{d}S(t) + \left(C(t) - \frac{\partial C(t)}{\partial S(t)} S(t) \right) r\mathrm{d}t \tag{10.4}$$

which can also be shown to equal zero.

Now, in reality, the trader does not know perfectly equation (10.1), and may therefore use a mis-specified and/or mis-estimated model. By mis-specified, we mean a model different from Black-Scholes, such as an arithmetic Brownian motion with time-varying parameters, or a mean-reverting diffusion process. By mis-estimated, we mean that the trader uses the Black and Scholes model, but mis-estimates the parameters μ and/or σ.

In either cases, the wrong option pricing model will give a price $\hat{C}(t)$ for the option that differs from the true (market) price $C(t)$. Moreover, the wrong option pricing model will also provide an incorrect hedge ratio $\frac{\partial \hat{C}(t)}{\partial S(t)}$. At time t, the trader's replicating portfolio will then be worth:

$$\Pi(t) = -C(t) + \frac{\partial \hat{C}(t)}{\partial S(t)} S(t) + \left(\hat{C}(t) - \frac{\partial \hat{C}(t)}{\partial S(t)} S(t) \right) \tag{10.5}$$

which is no longer necessarily equal to zero. The portfolio instantaneous variations are:

$$\mathrm{d}\Pi(t) = -\mathrm{d}C(t) + \frac{\partial \hat{C}(t)}{\partial S(t)} \mathrm{d}S(t) + \left(\hat{C}(t) - \frac{\partial \hat{C}(t)}{\partial S(t)} S(t) \right) r\mathrm{d}t \tag{10.6}$$

Using equations (10.5) and (10.6) and rearranging terms yields:

$$\begin{aligned} \mathrm{d}\Pi(t) = & \left(\hat{C}(t) - C(t) \right) r\mathrm{d}t \\ & + \left[\frac{\partial \hat{C}(t)}{\partial S(t)} - \frac{\partial C(t)}{\partial S(t)} \right] (\mu - r) S(t) \mathrm{d}t \\ & + \left[\frac{\partial \hat{C}(t)}{\partial S(t)} - \frac{\partial C(t)}{\partial S(t)} \right] \sigma S(t) \mathrm{d}W(t) \end{aligned} \tag{10.7}$$

This equation summarizes the consequences of model risk for our trader.

- The first term is a pricing error. The trader uses the model price ($\hat{C}(t)$) to determine the initial investment to set up the hedging portfolio. If $\hat{C}(t)$ differs from the market price $C(t)$, the initial investment is excessive or insufficient, and the difference is carried through time at the risk-free rate. As a consequence, the delta hedge strategy is no longer self-financing. In particular, at some point, the hedger may need to borrow and infuse external funds in order to maintain the delta-hedge. Since the amount borrowed may become larger than the value of his portfolio, this signifies that delta hedging with model risk can lead to bankruptcy.
- The second term results from (i) the difference between the true delta parameter and the delta given by the model, and (ii) the difference

between the drift of the underlying asset and the risk-free rate. Depending on the sign of these differences, at maturity, the hedging strategy may yield a profit or a loss. Therefore, the trader may end-up with a replicating portfolio that is far from what he should have in order to fulfil his liabilities. For some exotic options, delta hedging can actually even increase the risk of the option writer, as evidenced by Gallus (1996).

- The third term results again from a difference between the true delta parameter and the delta given by the model. In addition, it depends on a stochastic term ($dW(t)$), making the hedging strategy result both stochastic and path-dependent. Last, but not least, it also depends on the true level of volatility.

Clearly, in the presence of model risk, even though we assume frictionless markets, the delta hedging strategy of our trader is no longer replicating or self-financing. Even worse, it becomes path-dependent.

How can we one deal with model risk in practice when delta hedging a position? The answer is not straightforward. Rebalancing the hedge more frequently does not help, because there is still a difference between the true hedging parameters and those given by the model. A possible solution consists of looking for a super-hedging strategy, for example, a strategy such that the hedging result is guaranteed with a given probability whatever the true model[6]. Unfortunately, super-hedging strategies become rapidly expensive as the probability of being hedged increases. Another solution consists of specifying a loss function to be minimized by the hedging strategy. In this case, perfect hedging is transformed into minimum residual risk hedging. But as a consequence, pricing is not uniquely determined and risk neutrality cannot be used – different agents may have different loss functions, and therefore, reach different prices for the option being considered.

10.7 ELEVEN RULES FOR MANAGING MODEL RISK

Managing and controlling model risk is a difficult and complicated process, which should generally be performed almost on a case-by-case basis and cover at least three distinct areas: (a) the choice, testing and safekeeping of the mathematics and computer code that form the model; (b) the choice of inputs and calibration of models to market data; and (c) the management issues associated with these activities. The success of model risk management and control will often depend crucially on personal judgement and experience. Therefore, the following set of rules should not be considered as a series of recipes, but rather as a minimum checklist.

Rule 1: Define what should be a good model

Before qualifying a model as being better than another model, one needs to define precisely the model-quality metrics. What makes a model superior to another often varies according to its usage. Consider, for instance, a pricing model versus a risk management model. The former must match closely the values of known liquid securities and focuses on absolute prices today, while the latter needs to realistically represent the possible future evolution of market variables and focuses on relative variations tomorrow. Both goals are sufficiently different to result in divergences in the ranking of competing models, depending on the context. Therefore, a good pricing model will not necessarily be a good risk-management model, and vice-versa. Nor will the hypothesis or conclusions reached for the pricing model be necessarily applicable to the hedging model.

In addition to the model usage, what makes a model superior to another also varies according to the preferences of the model user and the necessary underlying assumptions. Unfortunately, practitioners often understate or neglect this aspect. For instance, it is common practice to test an option pricing model's quality by comparing the model's predicted prices with market prices, and using some sort of loss function such as the mean pricing error (with respect to the market), the mean absolute pricing error or the mean squared pricing error. The approach implicitly assumes that (i) model users display symmetric preferences (for example, equally consider under- and over-pricing, or care equally about losses and gains); (ii) the option pricing model is correct; and (iii) the market is efficient in its pricing process. If the option pricing model is rejected, we do not know if (ii) is untrue; (iii) is untrue; or (ii) and (iii) are untrue! And even if the model has been validated, the validation process may not hold true for an individual displaying asymmetric preferences (for example, downside risk aversion). It is therefore crucial to agree on the model-quality metrics prior to ranking any type of model.

Rule 2: Keep track of the models in use

Fighting model risk should start with a detailed inventory of the various models available within a financial institution. This means keeping records of which models are used, who uses them, and how they're used. For computer-based models, it also implies keeping track of who built them, who keeps the code and who is allowed to change it. It is not uncommon to see banks where the only version of the source code is stored on a magnetic tape somewhere in the archives.

An up-to-date inventory should significantly enhance productivity and reliability. Typically, in a financial institution, each time a new product is developed, there is a tendency to either quickly adapt an older model with or without authorized modifications (the spreadsheet syndrome), or rebuild a new model from scratch (the blank sheet syndrome). Neither of these approaches is efficient and both are potential model risk generators. By storing models in a common library, by documenting them and knowing their users, a significant amount of time and money can be saved and model risk is significantly reduced.

An up to date inventory will also help in understanding and explaining internal divergences. For instance, traders, risk managers and back-offices frequently use different models. This leads to an internal control problem and opens the door to conflicts regarding unexplained profit-and-loss differences. Although it would be preferable that all of them rely on the same approved model, this is rather wishful thinking, because their needs are fundamentally different. Knowing who uses what can significantly help solve such internal control problems.

Rule 3: Define a model-testing framework

This may appear as a tautology, but each financial institution should establish a complete and rigorous model-testing framework. Too often, model testing is limited to proofing some mathematical formulations and entering a few parameters in a spreadsheet to observe the model's output. This is clearly insufficient. Data mining techniques make it easy to obtain statistical proofs of nearly any relationship by selecting an appropriate historical dataset. Therefore, a rigorous model-testing framework should include:

- A dedicated model validation team, which should be independent of both the models' developers and final users to ensure impartiality and eliminate the operational risk embedded in the implementation of a model. Independent assessment is the only way to provide a welcome degree of comfort, useful suggestions and improvements, and avoid the set of incentives to realize profits early.
- A precise framework to guide all persons involved in models validation. This should include a standardized series of test procedures and data sets, as well as minimum precise requirements to qualify a model as acceptable (the model risk metric). These should not be considered as exhaustive, but rather as minima. For instance, whatever the option pricing model, a deep in the money call option should behave like a forward, while a deep out of the money call should behave like a zero-coupon bond.
- A clear formalization of internal responsibilities for validation. As a rule, if somebody is supposed to do it, nobody will do it.

It is important to realize that the role of model testing should not be reduced to validate or invalidate a model, but should also include increasing its reliability, revealing its weaknesses, confirming its strengths and promoting improvements. Consequently, it is essential that (1) any information generated during the test phase be recorded and documented; and (2) purchasing a model from an external vendor does not exempt it from the validation process.

Rule 4: Regularly challenge and revise your models

At the root of the model risk problem is that market and mathematical assumptions (for example, simplifications of market behavior) are often hard-coded and remain stagnant within the model, while things do change in real life. Consequently, models should not be carved in stone, but rather evolve and improve with time. All models used within an institution should be regularly revised and their adequacy to the current market conditions challenged. This process should include an analysis of the underlying assumptions as well as a consistency check with the best-accepted practices in the industry. In addition to this regular revision, institutions extending existing businesses or entering new ones should also make a special effort to reassess existing models, procedures, data and best practices before they adopt. Very naturally, model users should be involved in the process as they are likely to be aware of the latest developments in the field.

Rule 5: Mark to market or to market standards, not to a model

Following the Group of Thirty's (G30) recommendations, the calculation of the mark-to-market value[7] of derivative positions is widely practiced in the financial industry as a natural way to avoid model risk. Unfortunately, marking to market has its own dangers and may induce a false sentiment of security and overconfidence.

For positions where there is a conventional market price (for example, closing bid), one would expect the results of a good model to be quite close to those observable on the market. Appreciable differences should be seen as an early warning system, so that one needs to fully understand the sources of these differences to form an opinion of the model being tested. As an illustration, the 1997 disaster at NatWest could have been easily avoided by obtaining external implied volatility quotes from brokers or other institutions that trade in the marketplace and by comparing them with Natwest's values.

For more complex or illiquid derivative instruments, marking to market becomes a difficult exercise. When prices are not easily available, traders tend to use theoretical prices as a benchmark, generating an important model risk source (the mark to model syndrome). If the benchmark model is wrong, everything can go wrong. Institutions following this direction will be particularly at risk if its traders (relying on the wrong model) are themselves the unique providers of a given financial instrument on the market. Then, the market prices coincide with the incorrect model prices which means that large neutral positions could in fact generate important accumulated losses when the situation is discovered.

Although not ideal, marking to model may be acceptable if all market participants agree on a standard. However, there are fields with no consensus on a particular model. Consider for instance fixed income securities and interest rate modeling. Since the valuation of most assets relies on discounting cash flows, interest rate modeling is a very important area of finance. However, no definitive interest rate model has yet emerged.[8] This is good news for those who wish to carry out research in this line, but it is also a source of concern to investment banks and their regulators, as a mark to model gain or loss is clearly meaningless.

Rule 6: Simple is beautiful

The development of modern financial theory has come to a stage in which finance produces a rich source of challenging questions for a range of mathematical disciplines, including the theory of stochastic processes and stochastic differential equations, numerical analysis, the theory of optimization, and statistics. Theoretical results and computational tools are used, for instance, in the pricing of financial derivatives, for the development of hedging strategies associated with these derivatives, and for the assessment of risk in portfolios. Unfortunately, as the mathematics of finance reaches higher levels, the level of common sense seems to drop. Rather than starting with some idea, some concrete economic or physical or financial mechanism, and then expressing it in mathematics, researchers increasingly just write down an equation and try to solve it without any consideration of the usefulness of the overall process or its applicability to the real world. We believe this approach is clearly wrong: models should be based on concepts, information and insight, not just on advanced mathematics. Although mathematics is important to modeling, it should not be primary, but mostly complementary. Most financial models users will be fast-thinking actors in dynamic markets. Therefore, avoiding unnecessarily complicated models should be the rule. Whenever available, simple, intuitive and realistic models should always be preferred to complex ones.

For the same reason, model-users should only move to a more complex model or approach only when there's a value in doing so. In a sense, the science of modeling should be seen as an evolutionary process, a sort of chicken-or-egg problem. Better models should in turn allow for a better understanding of risks, the creation of new financial products, and, therefore, the need for additional models. As an illustration, the elegance of the Black and Scholes model is its rationality and logic. The model was not successful because prices of financial assets were actually log-normally distributed (which they may or may not be), but because the formula was easy to apply and understand, it arose as a valid first order approximation in a much wider class of models. The later Black and Scholes stochastic extensions (for example, with stochastic interest rates and/or stochastic volatility) were never as successful as the original model because they lost most of the qualities of their ancestor. As a rule, users should always understand the ideas behind a model and be comfortable with the model results. Treating a model as a black box is definitely the wrong approach.

Rule 7: Verify your data

A few years ago, the lack of reliable financial data was a major problem. It is still the case in a few areas (for example, the modeling of exotic derivatives or of credit risk). However, most of the time, we are rather awash with data. The key is turning this data into knowledge. Information should no longer be represented by data, but by data verified and organized in a meaningful way. The quality of a model's results depends heavily on the quality of its data feed. Garbage in, garbage out (GIGO) is the law, and data which is faulty to start with is likely to produce faulty conclusions after processing, and further, may ruin the benefit of sophisticated analytical models. Ensuring the integrity and accuracy of data feeds in models should therefore be key, even though it may require considerable effort and time. This implies checking both the series of data against errors, but also the semantics of the feed. Should the fair value be the price at which the firm could incrementally unwind the position, or the price at which they could sell the entire book, or the price above which they start to lose clients' interest? These questions need to be addressed at the beginning of the modeling process.

As an illustration, in the 1970s, Merrill Lynch had to book a US$70 million loss because it underpriced the interest component and overpriced the principal component of a 30-year strip issue.[9] The market identified the mis-pricing and only purchased the interest component. The problem was simply that the par-yield curve Merrill used to price both components was different from the annuity and the zero-yield curves that should have been used for each component. Oops! Wrong feed … As a rule, one should also beware of multiple data sources and non-synchronous data feeds (for example, stock indices and foreign exchange for daily close values) should also be reduced to a minimum, as they can lead to wrong pricing or create artificial arbitrage opportunities.

Rule 8: Use a model for what it is made for

Most models were initially created for a specific purpose. Things may start breaking down when a model is used outside its range of usefulness, or is not appropriate for the intended purpose. For instance, a good model for value at risk (VaR) will not necessarily be a good pricing model. The reason is that VaR estimates focus only on price variations, but not on price levels. Pricing errors are therefore not translated in the VaR. For the same reasons, a good pricing model is not necessarily a good hedging model, and vice versa. For example, using a stochastic or a deterministic volatility does not make a huge difference as far as the pricing is concerned if one gets the average volatility right. It makes a big difference as far as hedging is concerned.

Rule 9: Stress test your models

The G30 states that dealers should regularly perform simulations to determine how their portfolio would perform under stress conditions. This is often implemented through scenario analysis, which is appealing for its simplicity and wide applicability. Unfortunately, most institutions tend to focus solely on extreme market events such as the October 1987 crash. They neglect to test the impact of violations of the model hypothesis, and how sensitive the model's answers are to its assumptions. A small change in one parameter may result in dramatic changes in the model output, while a large change in another parameter may not necessarily change things at all.

Because there is no standard way of carrying out stress model risk testing and no standard set of scenarios to be considered, the danger is that one does not really suspect a model until something dramatic happens. To borrow a metaphor from a well-known movie, the threat of a North Atlantic iceberg was just a theory on 14 April 1912 – until the Titanic hit one. This is why the process should also depend on the qualitative judgement and experience of the model builder.

Rule 10: Beware of exotics!

By definition, exotic derivatives are highly subject to model risk. Firstly, exotic derivatives are not traded on liquid markets, but over the counter. Prices are therefore not the result of the equilibrium between supply and demand with numerous arbitrageurs waiting to capture any mispricing, but are rather supply driven. Secondly, exotic derivatives are often sensitive to some exotic parameters that cannot be hedged, are embedded into the model assumption, or are themselves linked to the difficulty of managing the risk. For instance, yield curve options pose vega spread volatility issues; Bermudian options create modeling problems due to their hybrid nature between American and European; and ratchet options pose difficulties associated with the existence of a volatility smile slope. None of these variables are directly hedgeable. And finally, models may produce similar plain vanilla option prices (and therefore fit to the market data), yet give markedly different prices of exotic options. This is documented for instance in Hirsa, Courtadon and Madan (2002).

Rule 11: Beware of correlations!

Correlations are found almost everywhere in finance, from portfolio construction to option pricing and hedging. As soon as there is more than one random parameter to be considered, correlations have a role to play. Unfortunately, correlations are among the most unstable parameters in real life, particularly during periods of heightened volatility. Risk managers often consider the possible effects of high return volatilities, but fail to account for the higher correlations between asset returns that would generally accompany the elevated volatility. One way to do so would be to employ information from historical periods of high volatility in order to form estimates of correlations conditional to a period of heightened volatility. These conditional correlations could then be used to evaluate the distribution of returns under a high volatility scenario. Put differently, the method used for stress testing a portfolio must not exclude the empirical feature that periods of high volatility are also likely to be periods of elevated correlation.

10.8 CONCLUSION

Acknowledging the rapid increase in sophistication of the financial community and products in recent years, most banks and trading rooms have been hiring people with the most up-to-date set of mathematical and quantitative skills. This directly resulted in a profusion of complex models – math engines that spit out risk and instruments valuations based on a flood of market data – that corporations are relying on to steer them through volatile markets.

Although most money is still made or lost because of market movements, not because of modeling, institutions are increasingly aware that no matter how advanced and refined financial models are, they are all subject to model risk and they should all be extensively tested, validated and tempered with judgment. However, intensive model-auditing, stress-testing and smart risk managers are all necessary – but they aren’t enough. All the math geniuses in the world don’t help if management either neglects to implement the procedures necessary to produce accurate calculations of risk or ignores those outputs. In all the recent derivatives losses, management can be faulted for a lack of understanding the problem. Murphy's Law holds; what can go wrong will go wrong. You can only tell when a model is wrong. It will always be more difficult to tell when a model is right.

NOTES

1. The Black – Scholes option-pricing formula, for example, can also be expressed as the solution to the heat-diffusion equation.
2. Note that the consequences of model risk are also visible in non-financial areas. For instance, a simple programming error – trying to store a 64-bit number into a 16-bit space – exploded the European Agency rocket Ariane 5 shortly after take off, destroying $7 billion of investment and 10 years of work.
3. The implied volatility is the volatility figure that one would need to plug in the Black and Scholes formula to obtain a theoretical price equal to the market price.
4. For instance, Derman and Kani (1994) construct implied binomial trees from an observed volatility smile and use it for pricing and hedging both standard and exotic options. Dupire (1994) provides an algorithm to recover a unique risk-neutral diffusion process consistent with observed (or fitted) option prices.
5. Working in the Black and Scholes framework leads to important analytic simplification without any loss of generality. The equivalent derivation in the case of a more general model can be found in Bossy *et al.* (1998).
6. See for instance Lhabitant, Martini and Reghai (2001) for options on a zero-coupon bond.
7. Marking to market is the process of regularly evaluating a portfolio on the basis of its prevailing market price or liquidation value.
8. See for instance Gibson, Lhabitant and Talay (2001) who survey more than 60 different models of interest rates.
9. In a strip issue, a bond is stripped into its regular coupon annuity payment (interest only) and principal repayment (principal only).

REFERENCES

Bachelier, L. (1900) "Theorie de la Speculation" (Thesis), *Annales Scientifiques de l'École Normale Superieure*, III-17: 21–86. [Cootner (ed.) (1964) *Random Character of Stock Market Prices*, Massachusetts Institute of Technology, pp. 17–78 or Haberman, S. and Sibett, T.A. (eds) (1995) *History of Actuarial Science*, VII, pp. 15–78, London].

Black, F. and Scholes, M. (1973) "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy*, 81: 637–59.

Cox, J.C. and Ross, S.A. (1976) "The Valuation of Options for Alternative Stochastic Processes", *Journal of Financial Economics*, 3: 145–66.

Derman, E. (April, 1996) "Model Risk", Goldman Sachs Quantitative Strategies Research Notes.

Derman, E. and Kani, I. (January, 1994) "The Volatility Smile and Its Implied Tree". Goldman Sachs Quantitative Strategies Research Notes.

Gallus, C. (1996) "Exploding Hedging Errors for Digital Options", Working Paper, Deutsche Morgan Grenfell, October.

Gibson, R., Lhabitant, F.S., Pistre, N. and Talay, D. (1999). "Interest Rate Model Risk: An Overview", *Journal of Risk*, 1(3): 37–62.

Gibson, R., Lhabitant, F.S. and Talay, D. (2001) "Modelling the Term Structure of Interest Rates: A Review of the Literature", Working Paper, http://ssrn.com/abstract=275076

Green, T.C. and Figlewski, S. (1999) "Market Risk and Model Risk for a Financial Institution Writing Options", *Journal of Finance*, 54(4): 1465–99.

Hirsa, A., Courtadon, G. and Madan, D.B. (2002) "The Effect of Model Risk on the Valuation of Barrier Options", *Journal of Risk Finance*, 4 (Winter): 1–9.

Lhabitant, F.S., Martini, C. and Reghai, A. (2001) "Pricing and Hedging Discount Bond Options in the Presence of Model Risk", *European Finance Review*, 4(1): 69–90.

Markowitz, H. (1959) *Portfolio Selection: Efficient Diversification of Investment* (New York, NY: John Wiley & Sons).

Merton, R.C. (1973) "The Theory of Rational Option Pricing", *Bell Journal of Economics and Management Science*, 4(1): 141–83.

Rubinstein, M. (1994) "Implied Binomial Trees", *Journal of Finance*, 49(3): 771–818.

Sharpe, W.F. (1966) "Mutual Fund Performance", *Journal of Business*, 39(1): 119–38.

CHAPTER 11

Evaluating Value-at-Risk Estimates: A Cross-Section Approach

Raffaele Zenti, Massimiliano Pallotta and Claudio Marsala

11.1 INTRODUCTION

Since 1998, regulatory guiding principles have required banks with significant trading activity to set aside capital to insure against extreme portfolio losses. The size of the market risk capital requirement is directly related to a measure of portfolio risk. Currently, in the regulatory framework, portfolio risk is measured in terms of its Value-at-Risk (VaR). Also in the community of asset management companies the quest for reliable risk management techniques has grown in recent years. The concept of VaR is now widespread among asset managers. This is an answer to the demand of sophisticated investors, such as pension funds and foundations, and it is also a clear response to the growing interest of asset managers for analytical tools that give better control on their portfolios.

Therefore, an important issue for risk managers and regulators is whether the VaR models in use are accurate enough. Verifying the accuracy of VaR models requires backtesting and now there is a variety of tests that examine the validity of these models. Existing methods tend to absorb a large amount of data and often show low power when used in small samples (those typically available).

In contrast with previous research on assessing the accuracy of risk metrics, which has focused on back-testing models on a time-series basis, we propose a methodology based on cross-section analysis of randomly generated portfolios. This exploits in a better way the information content of the multivariate distribution of returns used to estimate VaR.

The plan of this chapter is as follows. The next section (11.2) formalizes the notion of VaR and section 11.3 reviews existing backtesting methods. In section 11.4 we present our methodology and in section 11.5 we give some examples that show how this approach can be successfully applied. Section 11.6 concludes.

11.2 VALUE-AT-RISK

In general, VaR models attempt to forecast the time-varying distributions of portfolio returns, and different models provide different estimates. In addition, VaR estimates vary over time as market conditions and portfolio composition change. From a formal point of view, throughout the rest of this chapter, we will refer to a portfolio's value at risk as the α quantile of portfolio's return distribution:

$$VaR_{t,j}(\alpha, w_t, H) = F_{t,j}^{-1}(\alpha \mid \mathrm{I}_t, w_t) \tag{11.1}$$

where w_t is the vector of weights at time t for a given portfolio, H is the time horizon (say 10 days), $F_{t,j}^{-1}(\cdot)$ is the inverse of the cumulative probability distribution of portfolio's returns, estimated at time t using model j, and I_t is the information set on market conditions available at time t to estimate VaR. For a survey on VaR models and current risk management practice, see Jorion (2001).

11.3 REVIEW OF EXISTING METHODS FOR BACKTESTING

An assortment of tests has been proposed so far to check the accuracy of VaR models. We try to identify a taxonomy of the most popular testing methods.

11.3.1 Tests based on the hit function

These are by far the most popular. Consider the event that $R_{t,t+H}(w_t)$, the portfolio return on given period $[t, t+H]$, is less than $VaR_{t,j}(\alpha, w_t, H)$, for example, its reported VaR at time t. This event is a VaR failure.

The Hit function for model j counts failures as follows:

$$Hit_{t+H,j}(\alpha, w_t) = \begin{cases} 1 & \textit{if } R_{t,t+H} \leq VaR_{t,j}\,(\alpha, w_t, H) \\ 0 & \textit{if } R_{t,t+H} > VaR_{t,j}\,(\alpha, w_t, H) \end{cases} \tag{11.2}$$

Thus, the Hit function over time is a binary time series, $\{Hit_{t+H,j}\}$, that registers the history of VaR failures. As pointed out by Christoffersen (1998), the sequence $\{Hit_{t+H,j}\}$ should satisfies two fundamental properties.

Unconditional coverage property

A given model j satisfies this property if:

$$\Pr\lfloor Hit_{t+H,j}(\alpha, w_t) = 1\rfloor = \alpha \tag{11.3}$$

then on average the model is correct.

There are a variety of tests able to verify this property. Kupiec (1995) models $\{Hit_{t+H,j}\}$ as a sequence of independent draws from a binomial distribution with probability of occurrence equal to α. Campbell (2005) suggests a test performed directly on the sample average of failures. It is well-known, see among others Kupiec (1995) and Lopez (1999), that a key issue with these tests is their statistical power: they exhibit low power in relatively small samples (for example, 250 days, as in the regulatory framework). This implies that the chance of misclassifying an erroneous VaR model as accurate is high.

Independence property

A given model j satisfies this property if the sequence $\{Hit_{t+H,j}\}$ is independently and identically distributed (i.i.d.), that is, failures do not exhibit serial correlation. A model that exhibits the correct unconditional coverage property but that violates the independence property might display clusters of failures over time. This is undesirable, as the consequence could be an exposure to financial losses for quite a few periods in a row. The two properties, for example, unconditional coverage and independence, can be jointly expressed as:

$$Hit_{t+H,j}(\alpha, w_t) \overset{i.i.d.}{\sim} Binomial(\alpha) \tag{11.4}$$

where $Binomial(\alpha)$ is a binomial variate with probability α. In words, the sequence of failures must be a series of independent and identically distributed binomial events. The main contribution in this area is by Christoffersen (1998) that depicts a joint test of these two properties, as they are both crucial. If the two properties are tested in a standalone fashion, there is the risk of not detecting a model which is poor in one or the other property. Conversely, joint tests lack of the ability to identify models that are deficient only in one of the two properties.

These methods are quite data-intensive, since they only make use of whether or not a failure occurred. They need a lot of information. The reason is simple: with α equal to 1 percent as in the banking regulatory framework, or equal to 5 percent (more common among asset managers that use VaR for portfolio management purposes), one has to collect a lot of history in order to see some failures, if the model adopted is decent. The power of these tests is rather scarce.

11.3.2 Tests based on multiple VaR levels or the entire probability density function

Especially with small samples, a more accurate use of the information at disposal is vital: models can be more precisely tested examining more quantiles, for example, VaRs for different α_i $i = 1, 2, 3, \ldots, p$ levels, extending the procedure outlined in the previous section. In a more general fashion and broadening the idea, evaluating the entire probability density function of portfolio's return. The evaluation of the entire density forecast extracts a greater amount of information from available data, as it uses the full range of forecasted outcomes.

Among others, Chatfield (1993), Crnkovic and Drachman (1997), Diebold, Gunther and Tay (1998), Christoffersen (1998) as well as Berkowitz (2001) have proposed methods for evaluating VaR according to multiple VaR levels or the entire probability density function. The idea behind this class of testing methods is to transform ex-post portfolio's returns into a new variable, which is defined in (0,1). The transformation is made using the forecasted (for example, ex-ante) cumulative probability distribution of portfolio returns. Formally:

$$z^j_{t,t+H} = F_{t,j}(R_{t,t+H}|\mathrm{I}_t, w_t) \tag{11.5}$$

Now, for any variate $x \in \Re$ and any probability density function $\phi(x)$, the cumulative probability distribution:

$$\Phi(x) = \int_{-\infty}^{x} \phi(s)\mathrm{d}s \tag{11.6}$$

is a uniform variate on the unit interval, as $\Phi(x)$ is just a probability measure, that is uniform by definition. Hence, the testing procedures rely on the time series $\{z^j_{t,t+H}\}$, that should simultaneously exhibit two properties:

1 it should be uniformly distributed on the unit interval, as implied by (11.6); and

2 it should not exhibit serial correlation.

The two properties above can be jointly expressed as:

$$z^j_{t,t+H} \overset{i.i.d.}{\sim} Uniform(0, 1) \tag{11.7}$$

and can be tested, individually or jointly.

One limit of these approaches is that they can require large data sets in order to check the accuracy of VaR models. This is not always true, for example, the test proposed by Berkowitz (2001) is relatively parsimonious.

The crucial point about this methods is that they need the entire cumulative probability distribution of portfolio's returns estimated at time t with model j, for example, $F_{t,j}(\cdot)$. This is not a trivial requirement, as not every VaR model has the goal to predict the entire distribution, or at least it does not pretend to predict well the entire distribution. The reason is that VaR models focus on the left tail of returns' distribution. Typical examples are models based on Extreme Value Theory or Quantile Regression. They could possibly perform very well on lower quantiles, say 1 percent or 5 percent quantiles, but they cannot provide useful information on less extreme quantiles.

Tests based on a given Loss function

Alternative methods are based on loss functions that assign numerical scores to VaR estimates according to some metric that measures the impact of failures. Formally the loss function is any function of the general form:

$$L_{t+H,j}(VaR_{t,j}(\alpha, w_t, H), R_{t,t+H}, \alpha, \Theta) \tag{11.8}$$

that depends on VaR estimates and realized returns (typically according to a distance measure, because realized returns far below VaR have to be penalized), on the probability level α (as different α can be evaluated differently), and on some parameter Θ that reflects the specific concerns that this function has to take into consideration. The Hit function (11.2) is a very simple example of (11.8). Different VaR models can be evaluated based on the scores arising from (11.8).

Lopez (1999), who pioneered this approach, suggests this methodology as a flexible alternative to statistical hypothesis testing. Once a regulatory loss function is specified, he argues, VaR estimates could be compared across time and across financial institutions. Anyway, it might be a difficult task to specify a proper loss function. Another drawback comes from the fact that, in order to calibrate the assessment procedure, it is necessary to make assumptions about the distribution of portfolio's returns.

11.4 AN EXTENSION: THE CROSS-SECTION APPROACH

So far we have seen that there is an intrinsic difficulty in testing the accuracy of a VaR model: intuitively, this is because α is usually small, thus failures are rare events, and one needs large data-sets to test rare events. Data limitation is the issue. So it is crucial to increase the amount of information at our disposal when assessing a VaR model.

The starting point is that a portfolio's VaR depends on its assets' weights and a forecast of the multivariate distribution of the assets' returns. As portfolio weights are exogenous, for example are given, at the very roots of any process aimed to assess VaR model there is an evaluation of the forecasted multivariate distribution of assets returns. The problem is that all the methods outlined in section 11.3 lose the multivariate framework as they focus on portfolio's VaR and portfolio's actual returns that are conceptually univariate objects. We wish to make this point clear.

Considering an investment universe made of N assets, the actual returns $r_{t,t+H}$ on given period $[t,t+H]$ are an N-dimensional process, portfolio weights are collected in a N-dimensional column vector, but the portfolio return, for example $R_{t,t+H} = r'_{t,t+H} \cdot w_t$, is a univariate process. Hence a 'portfolio perspective' implies a reduction of dimensionality, from N to 1.

In a similar way, when estimating VaR one starts (at least in principle) with a forecast of the multivariate probability density function of returns, a very granular piece of information. However, in order to estimate a portfolio's VaR, this granular information is combined with the assets' weight, and because of this mapping, at the end of the process there is a univariate object. Again, a portfolio perspective involves a decrease of the number of dimensions we are dealing with. This has a big impact. Think about a portfolio whose weights are unequally distributed: a small number of assets have a large weight, while the others exhibit small weights. Note that this is a fairly common situation in the asset management industry, as many portfolios have approximately the same structure of indexes made of hundreds of securities. Passive mutual funds are a typical example, as well as low tracking-error portfolios, very common in the industry (just to name two real-world situations). Now, if a large number of securities has a small weight, a portfolio's VaR estimate would be determined mainly by a subset of the information content of multivariate distribution: the subset that relates to larger positions. This tends to obscure VaR model capability, as the model could be overall poor but, by chance, could be good at estimating the risk of the dominating assets. This could lead to an erroneous positive assessment, and subject to more sampling error. One could argue that small positions are not important by definition. The key point is that a portfolio's composition changes over time: a small position today could be a large position tomorrow.

Therefore the idea is to recover more information from the multivariate estimated distribution, which allows us to measure more correctly the forecasting capability of the model under consideration.

The methodology is simple and is based on the following steps:

1. At time t, considering an investment universe made of N assets, we randomly generate K portfolios (where K is large number, say some hundreds), whose weights are collected in the $K \times N$ matrix W_t (each column is a portfolio).
2. We estimate VaR for each portfolio using model j according to (1), so that we have a vector of VaRs, $\mathbf{VaR}_{t,j}(\alpha, W_t, H)$.

3 At $t + H$, it is possible to compute the actual returns for the K portfolios, collected in the vector $\mathbf{R}_{t,t+H} = W_t \cdot r_{t,t+Ht}$.

4 Armed with $\mathbf{VaR}_{t,j}$ and $\mathbf{R}_{t,t+H}$, it is possible to calculate the number (or alternatively the relative frequency) of failures *n_failures* across the K portfolios through the Hit function.

5 We apply one or more statistical tests, in order to assess VaR model j, evaluating if *n_failures* is significantly different from the theoretical frequency, under the null hypothesis that the model is correct. This can be done using most of the tests outlined in Section 11.3. The theoretical frequency of failures depends on the forecasted multivariate distribution of returns. In the scholastic case of independent returns, the total number of failures follows a binomial distribution $\Pr[n\_failures = x] = Binomial(\alpha, K)$, while in presence of some degree of comovements, it must be computed using the estimated multivariate distribution. This involves computing a cumulative distribution function, which can be done numerically, for example, by Monte Carlo simulation. Given the degree of computing power currently available in most cases this is not a major obstacle for most practitioners. For Gaussian models like RiskMetrics® this computation is rather fast.

6 It is possible to check that, over time, the VaR model under examination does not display serial correlation. This can be done, for example, by monitoring the time series of failures $\{n\_failures_t\}$, where the matrix of K portfolios is kept fixed over time, that is $W_t = W$. Basically we keep track of a large number of portfolios' failures.

It is clear that there is a lot of additional information available on the performance of one (or more) VaR models, as we focus on a high number of portfolios on each assessment date, that is, we keep a cross-section perspective. After a small number of runs of this procedure, it is possible to judge a model in a rather precise way. Of course, this can also be done using an historical backtesting method (running the model in the past). However, this requires a lot of historical data, with the associated problems, for example some securities have no price in the past, there are corporate actions that alter the history, and so on.

11.5 APPLICATIONS

11.5.1 An intuitive example

Let us look at a simple example. We assume that the data generation process (DGP) is an N-dimensional process such that $r_{t,t+H} \overset{i.i.d.}{\sim} Normal_N(0,\ \mathrm{I})$, where I is the identity matrix. We assume that N is equal to 1,000. We then estimate

Table 11.1 Proportion of failures

		Model 1	Model 2	Model 3
Run 1	POF	5.4%	4.7%	3.6%
	Kupiec statistic	0.27	2.28	10.00
	p-value	39.6%	86.9%	99.8%
Run 2	POF	5.1%	4.6%	3.7%
	Kupiec statistic	0.88	2.73	9.04
	p-value	65.2%	90.2%	99.7%
Run 3	POF	4.9%	3.8%	3.5%
	Kupiec statistic	1.49	8.13	11.02
	p-value	77.8%	99.6%	99.9%
Run 4	POF	6.5%	5.0%	4.2%
	Kupiec statistic	0.92	1.16	50.3
	p-value	66.3%	71.9%	97.5%
Run 5	POF	5.2%	4.0%	3.1%
	Kupiec statistic	0.64	6.48	15.72
	p-value	57.5%	98.9%	100.0%

VaR, with α equal to 5 percent, according to several models that differ for the hypothesis made about the DGP:

1 model 1 assumes $r_{t,t+H} \overset{i.i.d.}{\sim} Normal_N(0, \mathrm{I})$;

2 model 2 assumes $r_{t,t+H} \overset{i.i.d.}{\sim} Normal_N(0, 1.1 \cdot \mathrm{I})$, that is, variance is 10% greater than reality; and

3 model 3 assumes $r_{t,t+H} \overset{i.i.d.}{\sim} Normal_N(0, 1.2 \cdot \mathrm{I})$, that is, variance is 20% greater than reality;

We apply our procedure, generating K random portfolios with K equal to 1,000. We also create a market scenario, generating a vector of returns $r_{t,t+H}$ using the chosen DGP. We then test the unconditional coverage property using the binomial test proposed by Kupiec (1995), probably the most popular method among practitioners. We calculate the proportion of failures (POF), the Kupiec statistic and the associated p-value. We apply the procedure for five consecutive periods (5 days). Results are reported in Table 11.1.

It is apparent that VaR estimates obtained using Model 1 (corresponding to the DGP) are better than those obtained from the other models: the number of failures and the other statistics say that the performance of this model is closer to what we expect in theory. If we assume that the DGP is a completely different process, for example a multivariate t-student with 3 degrees of freedom a with correlation equal to zero, such that

Table 11.2 Proportion of failures

	Model 1	Model 2	Model 3
POF	1.8%	1.5%	1.2%
Kupiec statistic	39.24	47.03	56.04
p-value	100.0%	100.0%	100.0%

$r_{t,t+H} \overset{i.i.d.}{\sim} t-student_N(0, \mathrm{I},\ 3\ dgf)$, after a single run we see that all the models are in difficulties, as can be seen in Table 11.2.

After a small number of days, the picture becomes clearer. In order to get more information, one can use some Bayesian analysis, as outlined below.

11.5.2 Comparative Bayesian analysis of the performance of the VaR models

In the case of the POF test, we are testing the null hypothesis that the proportion of failures $p = n\_failures/K$ is equal to $Binomial(\alpha, K) = p_{\alpha,K} \approx 5.8\%$ in our case (because, for the sake of simplicity, we are assuming absence of correlation).

Hence, it would be reasonable to assume as a prior, a Beta distribution with parameters $a_0 = K \cdot p_{\alpha,K} + 1$, $b_0 = K \cdot (1 - p_{\alpha,K}) + 1$. Thus, our prior can be written as:

$$\pi(p) \propto p^{a_0 - 1}(1 - p)^{b_0 - 1} \tag{11.9}$$

Basically, we center our prior distribution on 5.8 percent, that is the theoretical POF if the model is correct.

After the first run we have some data: we observe *n_failures* out of *K* results. The likelihood function $l(n\_failures,\ K, p)$ is given by the binomial distribution:

$$l(n\_failures, K, p) = \binom{K}{n\_failures} p^{n\_failures}(1 - p)^{K - n\_failures} \tag{11.10}$$

that can be rewritten as:

$$l(n\_failures, K, p) \propto p^{n\_failures}(1 - p)^{K - n\_failures} \tag{11.11}$$

We combine the likelihood function with our prior distribution and we get the posterior distribution which turns out to be a Beta distribution (as we are using a conjugate prior):

$$\begin{aligned} f(p|n\_failures, K) &\propto \pi(p) \cdot l(n\_failures, K, p) \\ &= Beta(a_0 + n\_failures, b_0 + K - n\_failures) \\ &= Beta(a_1,\ b_1) \end{aligned} \tag{11.12}$$

so we have a Bayesian updating scheme that enables us to understand how a given VaR model performs over time. After t steps, the distribution of p is:

$$Beta_t(a_t,\ b_t) = Beta\left(a_0 + \sum_{i=1}^{t} n\_failures_i, b_0 + \sum_{i=1}^{t}(K - n\_failures_i)\right) \tag{11.13}$$

For example, Figure 11.1 shows the prior distribution and the posterior distribution, calculated using (11.13), for Model 1 and Model 2 of the previous example. The prior is common to both models, as the null is that any given model is correct, but the posterior is different: it is rather evident that Model 1 is closer to the prior distribution which corresponds to the null. As time passes, the distribution of Model 1 will eventually converge to the prior (true by definition). Thus this methodology can help in ranking different risk models.

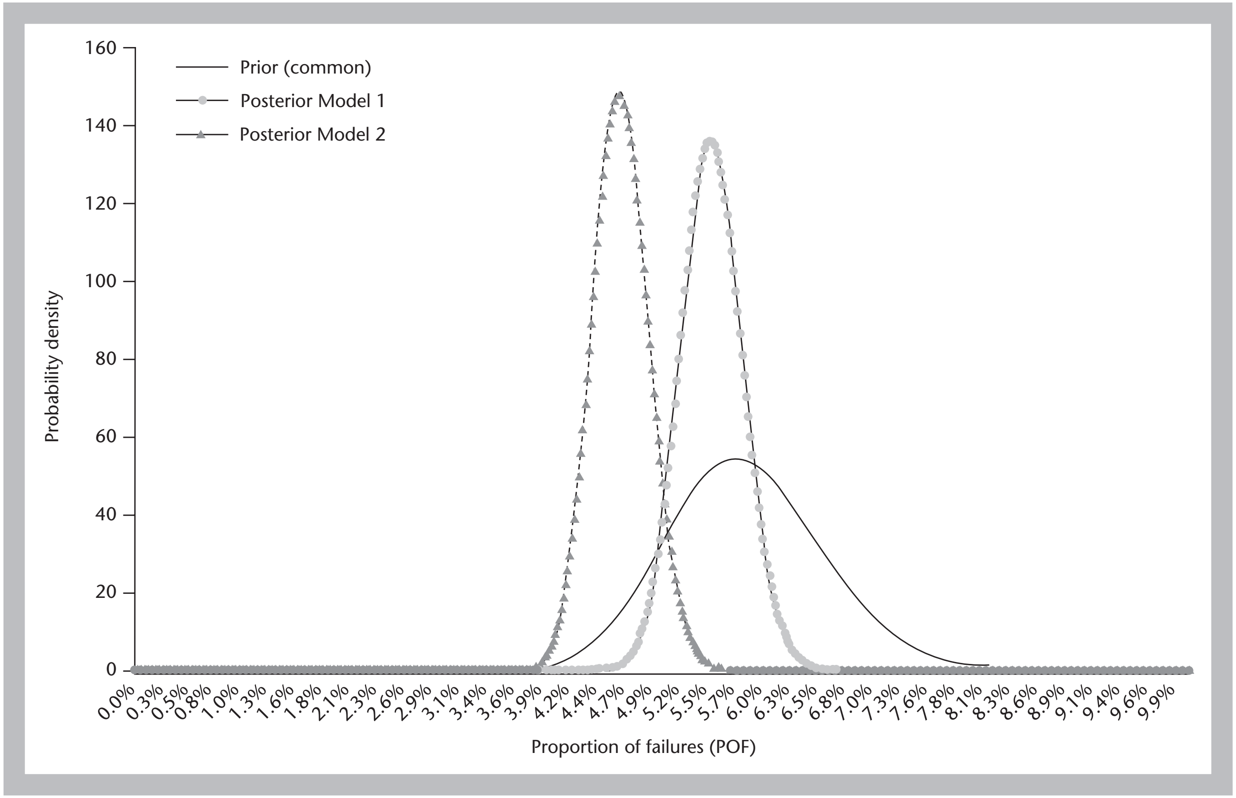
Depending on the perspective of the analysis, one could instead use a non-informative prior, for example a uniform distribution over the unit interval.

For global risk models (for example, models that can be applied across the whole spectrum of asset classes – many specialized software products claim to be global risk analysers), this procedure can be applied in a parallel fashion to several investment universes. This enables regulators and financial institutions to assess the model on several areas, for instance European equities, US equities, Far East equities, emerging markets equities, international government bonds, international corporate bonds, and so on. It is possible to understand how and where a model fails. For example, if among the randomly generated K portfolios, portfolio s exhibits a large failure, it is possible to analyse the sources of this loss, as suggested below.

11.5.3 Failures analysis

First, consider the shortfall of model j on a given period $[t, t+H]$ for portfolio s, defined by the vector of weights w_t^s:

$$D^j_{t,t+H}(\alpha, w_t^s, H) = R_{t,t+H}(w_t^s) - VaR_{t,j}(\alpha, w_t^s, H). \tag{11.14}$$


Prior (common)
Posterior Model 1
Posterior Model 2
Probability density
160
140
120
100
80
60
40
20
0
0.0% 0.3% 0.5% 0.8% 1.0% 1.3% 1.6% 1.8% 2.1% 2.3% 2.6% 2.9% 3.1% 3.4% 3.6% 3.9% 4.2% 4.4% 4.7% 4.9% 5.2% 5.5% 5.7% 6.0% 6.3% 6.5% 6.8% 7.0% 7.3% 7.6% 7.8% 8.1% 8.3% 8.6% 8.9% 9.1% 9.4% 9.6% 9.9%
Proportion of failures (POF)

Then compute the derivative $\partial D^j_{t,t+H} / \partial w^s_t$, that, for asset k is given by:

$$r^k_{t,t+H} - \frac{\partial VaR_{t,t+H}}{\partial w^s_t(k)} \tag{11.15}$$

It is easy to notice that (24.15) is simply the difference between the actual return on asset k and its marginal VaR, that is the change in portfolio VAR resulting from a marginal change in a given position. Ranking all the securities on the basis of (11.15) can be of help to understand where a model fails, and can suggest improvements to the model itself.

There are other applications. In order to chose the appropriate VaR model, several models can be put in competition and in a reasonable time one can collect a large amount of information about their behavior. As no financial model is strictly true, this kind of testing procedure should provide support as to choosing the most appropriate model.

11.6 CONCLUSION

We have presented an approach to backtesting VaR models that introduces a new perspective on models' behavior, as it is performed as a kind of cross-section analysis of randomly generated portfolios. With this approach, regulators and financial institutions can use basically all the existing techniques, augmented by extra information coming from the cross-sectional analysis.

Aside from formal testing, this kind of approach to backtesting can help to:

- understand quickly if a model does work;
- learn how and under what conditions models fail;
- track their performance over time using detailed information coming from many portfolios;
- test global risk models simultaneously on more markets and asset classes; and
- discriminate among competing VaR models (as no financial model is ever factually true), helping in the choice of the best model.

REFERENCES

Berkowitz, J. (2001) "Testing Density Forecasts with Application to Risk Management", *Journal of Business and Economic Statistics*, 19(4): 465–74.

Campbell, S.D. (2005) "A Review of Backtesting Procedures", *Finance and Economic Discussion Series* (Washington, DC: Federal Reserve Board).

Chatfield, C. (1993) “Calculating Interval Forecasts”, *Journal of Business and Economic Statistics*, 11(2): 121–35.

Christoffersen, P. (1998) “Evaluating Interval Forecasts”, *International Economic Review*, 39(4): 841–64.

Crnkovic, C. and Drachman, J. (1997) *Quality Control, in VaR: Understanding and Applying Value at Risk* (London, UK: Risk Publications).

Diebold, F.X., Gunther, T.A. and Tay, A.S. (1998) “Evaluating Density Forecasts”, *International Economic Review*, 39: 863–83.

Jorion, P. (2001) *Value at Risk* (Chicago, IL: Irwin Publishers).

Kupiec, P. (1995) “Techniques for Verifying the Accuracy of Risk Management Models”, *Journal of Derivatives*, 3: 73–84.

Lopez, J.A. (1999) “Methods for Evaluating Value at Risk Estimates”, *FRBSF Economic Review*, 2(4): 3–15.

CHAPTER 12

Correlation Breakdowns in Asset Management

Riccardo Bramante and Giampaolo Gabbi

12.1 INTRODUCTION

Over recent years, financial and real market globalization has accelerated the process of increasing positive values of correlations. This phenomenon changed many portfolio managers' practices, which are now strictly linked with sector behaviors. In order to verify whether portfolio managers can correctly estimate the eventual correlation jump over time, we provide some new evidences for correlation dynamics among equity markets.

The chapter aims to answer the following questions:

1 Is there a relation between exponential correlation changes and volatility movements?

2 Is this relation structurally constant for all the movements of correlations, or do correlation jumps show different behaviors?

3 What errors might we make in not considering correlation jumps in portfolio optimizations?

12.2 DATA AND DESCRIPTIVE STATISTICS

The data consist of daily exponential volatility and correlations for equity markets of the United States, the Euro area and Japan. The data source is JPMorgan RiskMetrics^{TM}. The way RiskMetrics computes correlations and

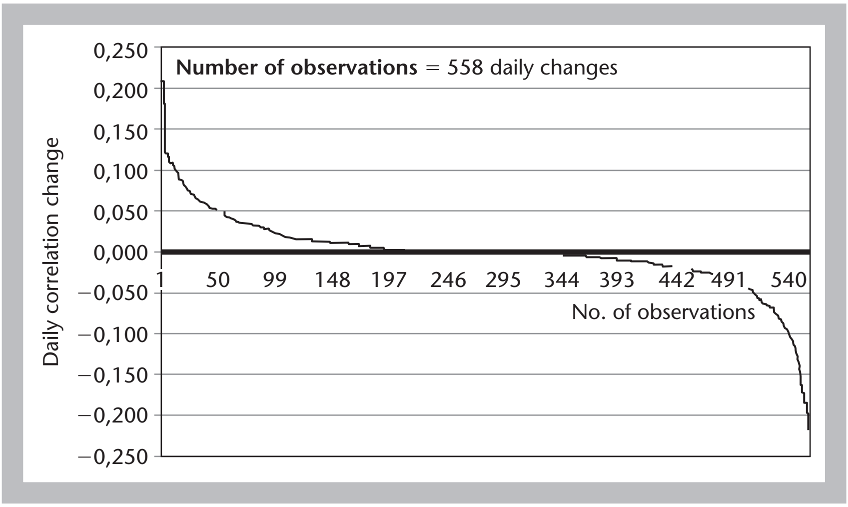


Figure 12.1 Distribution function of jumps EUR–USD, 2003–05

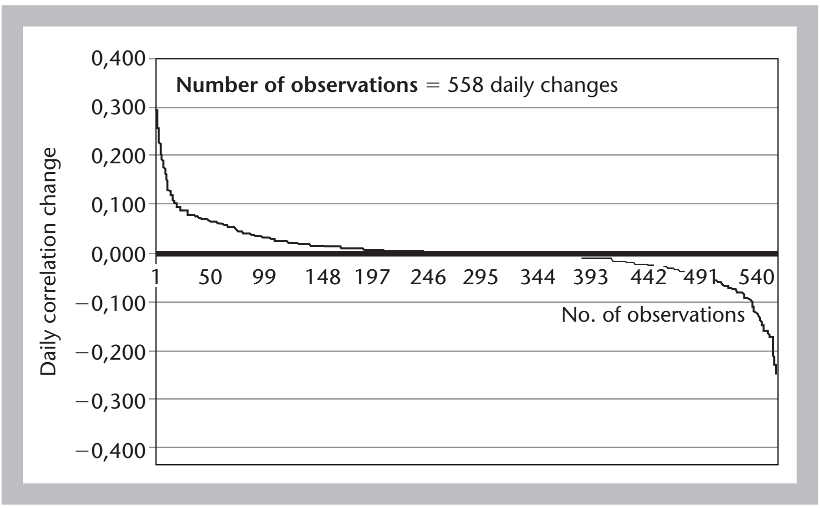


Figure 12.2 Distribution function of jumps EUR–JPY, 2003–05

volatilities is through the method of exponentially weighted daily historical observations with a decay factor of 0.94. The complete time series is recorded from 1 January 2003 to 30 September 2005 consisting of 687 data points: 558 of them, until 31 March 2005, were used for historical model estimation whereas the remaining ones were used for out of sample testing. From these data we computed all the variations, from positive to negative. Figures 12.1 to 12.3 depict the distribution function of correlation jumps in all the markets considered.

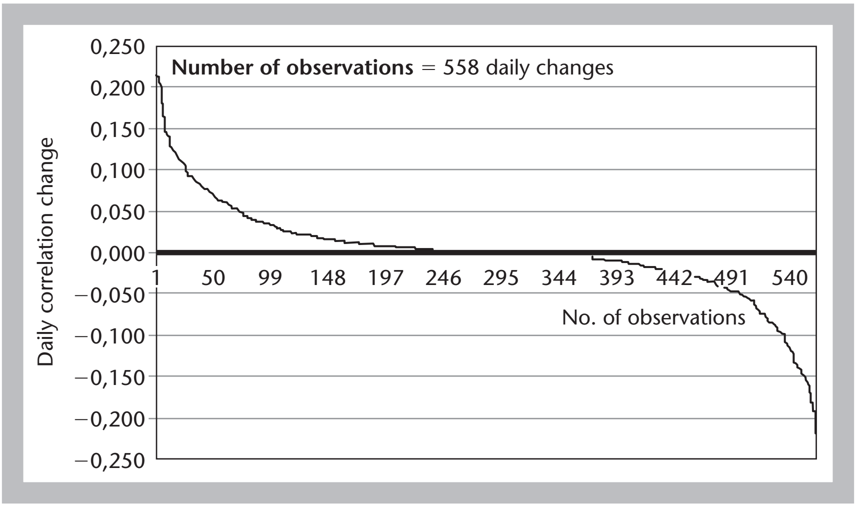


Figure 12.3 Distribution function of jumps JPY–USD, 2003–05

The correlation statistics show that historical data with skewness is near zero, while the kurtosis shows a value far from 3 in the case of the correlation between the Euro and the US market (1.83). In the other two cases the values are, respectively, 3.79 and 3.04. In the case of volatility, skewness is always far from symmetry, while kurtosis is closer to 3 (correspondingly, 3.03, 3.46 and 2.76 for the three markets considered). Afterwards we computed the absolute value of correlation jumps. The resulting time series are divided into two groups: the first is the complete time series of 558 observations; the second is that characterized by the 10 percent highest changes (56 observations).

12.3 CORRELATION JUMPS AND VOLATILITY BEHAVIOR

Firstly, we studied the behavior of correlation jumps in the equity indexes, as explained by the volatility changes observed in the originating markets. The methodology follows the idea that it could be possible to explain correlation changes through volatility differences. The model estimated is

$$\Delta\rho_{A,B_t} = \alpha + \beta \cdot \Delta\sigma_{A_t} + \gamma \cdot \Delta\sigma_{B_t} + \varepsilon_t \qquad (12.1)$$

where $\Delta\rho_{A,B_t}$ is the daily correlation difference computed at time t between the two generic markets, A and B; $\Delta\sigma_{A_t}$ is the market A daily volatility difference at time t; and $\Delta\sigma_{B_t}$ is the market B daily volatility difference at time t.

Table 12.1 Regression equation of correlation EUR–USD changes explained by volatility differences

Variable	Coefficient	Std. error	t-statistic	Prob.
DEURV	1.062701	0.128748	8.254101	0.0000
DUSDV	0.926331	0.132735	6.978780	0.0000
MA(1)	0.136762	0.042159	3.243962	0.0012
R-squared	0.180499	Mean dependent var		0.065814
Adjusted *R*-squared	0.177546	S.D. dependent var		0.113572
S.E. of regression	0.102997	Akaike info criterion		−1.702865
Sum squared residuals	5.887696	Schwarz criterion		−1.679615
Log likelihood	478.0992	F-statistic		61.12074
Durbin–Watson stat.	1.971937	Prob(F-statistic)		0.000000

Notes: The dependent variable; differences of exponential correlations between the equity euro market and the equity US market. Explanatory variables are DEURV: differences of exponential volatility of the equity euro market; DUSDV: differences of exponential volatility of the equity US market; and MA(1): moving average of first-order component. Number of observations = 558 daily changes.

For all the three correlations we also estimated the model:

$$\Delta\rho_{A,B_t} = \alpha + \beta_1 \cdot \Delta\sigma_{A_{t-1}} + \ldots + \beta_n \cdot \Delta\sigma_{A_{t-n}} + \gamma_1 \cdot \Delta\sigma_{B_{t-1}} + \ldots + \gamma_n \cdot \Delta\sigma_{B_{t-n}} + \varepsilon_t \quad (12.2)$$

where n is a time lag which was set up to 10 days during regression stepwise search.

To model the regression equation, in many cases it was useful to introduce a moving average component:

$$\Delta\rho_{A,B_t} = \alpha + \beta_1 \cdot \Delta\sigma_{A_{t-1}} + \ldots + \beta_n \cdot \Delta\sigma_{A_{t-n}} + \gamma_1 \cdot \Delta\sigma_{B_{t-1}} + \ldots + \gamma_n \cdot \Delta\sigma_{B_{t-n}} + \vartheta_1\varepsilon_{t-1} + \ldots + \vartheta_q\varepsilon_{t-q} + \varepsilon_t \quad (12.3)$$

where the order of the MA term was generally set to one.

Table 12.1 shows a relative capability (*R*-squared is around 18 percent) to explain correlation changes through volatility variables, even when a moving average factor was selected. All variables demonstrate a high value of the student *t*-test, that is a significant statistical contribution. The sum squared of errors is by the way 80.5 percent lower than the complete time series.

As for the residuals, depicted in Figure 12.4, the Durbin–Watson (DW) statistic which measures the serial correlation was very close to 2.[1]

For the Euro area market and the USA we selected the first 10 percent of higher correlation changes (56 data points). We then estimated the regression equation (12.2) as previously described and the results are displayed

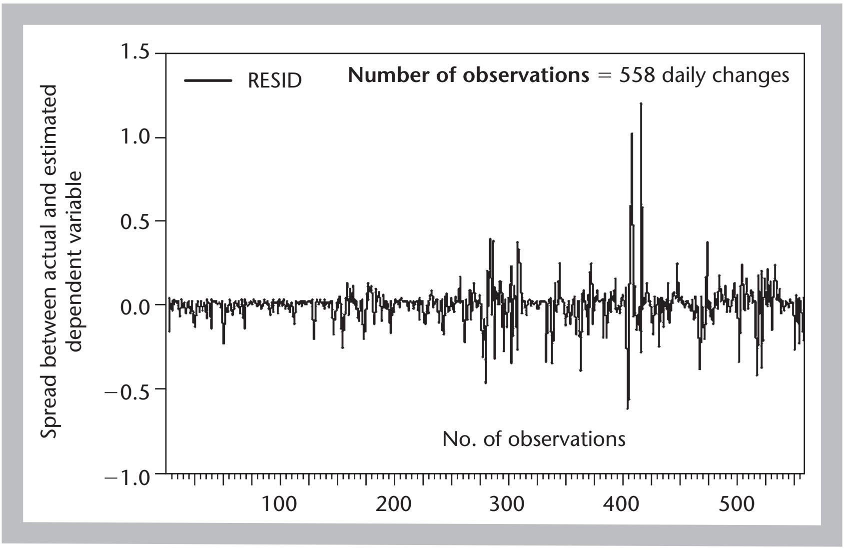


Figure 12.4 Plot of residuals of correlation EUR–USD changes explained by volatility differences

in Table 12.2. In particular, only a variable of the Euro market is statistically significant (3 days lag), while the US market explains the dependent variable with three variables (1–3 and 8 days lag). But, what is clearly higher is the *R*-squared value which reaches a level around 50 percent. In this case, the Durbin–Watson statistic shows a value of 2.26, which should be a little negative correlation. Actually, there are some limitations of the DW test as a test for serial correlation. One of them is the fact that if there are lagged dependent variables on the right-hand side of the regression equation the DW test is no longer valid.[2]

Consistent with the higher *R*-squared value, residuals are graduated within a range of (−0.35; 0.42) while the equation estimated over the complete time series has a range of (−0.55; 1.19). In Figure 12.5 the Euro–US market correlation residuals are reported. Table 12.3 proves that the correlation delta between the Euro area market and the Japanese market cannot be as well explained as the Euro area market versus the US market. In particular, the *R*-squared index is about 5.5 percent. The sum of squared residuals of the Euro–Japan market is 181.68.

Figure 12.6 demonstrates that the range of residuals is (−1.87; 9.53). In this case, the residuals of the higher correlations (Figure 12.7). Move within a range which is very similar in the negative side (−1.93) but lower in the positive side (8.11).

Even though the *R*-squared value is lower, the value of the higher correlation changes is 11 percent. In this case only two explicative variables enter the equation, respectively the variation of the volatility measured in

Table 12.2 Regression equation of higher correlation EUR–USD changes explained by volatility differences

Variable	Coefficient	Std. error	t-statistic	Prob.
C	0.256095	0.058295	4.393106	0.0001
DEURV3	3.887041	0.978974	3.970526	0.0002
DUSDV1	3.251217	0.572326	5.680705	0.0000
DUSDV3	−2.514702	0.769123	−3.269570	0.0019
DUSDV8	−2.420236	1.361542	−1.777570	0.0814
R-squared	0.500639	Mean dependent var		0.321851
Adjusted *R*-squared	0.461473	S.D. dependent var		0.199156
S.E. of regression	0.146150	Akaike info criterion		−0.923327
Sum squared residuals	1.089344	Schwarz criterion		−0.742492
Log likelihood	30.85316	F-statistic		12.78261
Durbin–Watson stat.	2.262736	Prob(F-statistic)		0.000000

Notes: The dependent variable; differences of exponential correlations between the equity euro market and the equity US market. Explanatory variables are DEURV3: equity euro market exponential volatility differences with 3 days lag; DUSDV1: equity US market exponential volatility differences with 1 day lag; DUSDV3: equity US market exponential volatility differences with 3 days lag; and DUSDV8: equity US market exponential volatility differences with 8 days lag. C: constant in regression. Number of observations = bigger 56 daily changes.

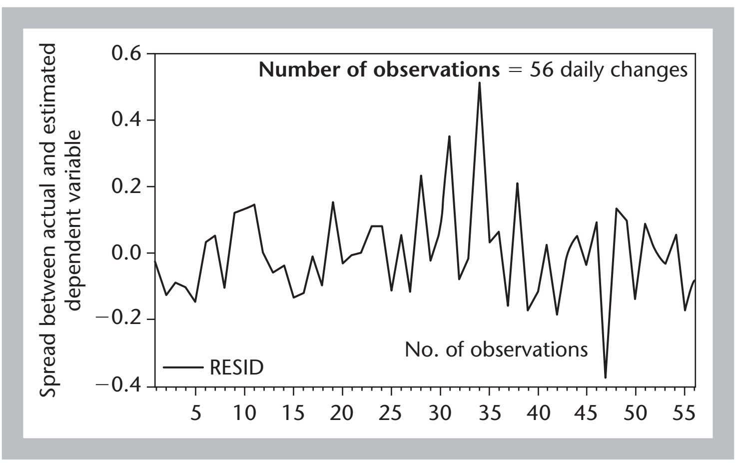


Figure 12.5 Plot of residuals of correlation EUR–USD higher changes explained by volatility differences

the Euro market 7 days before, and the variation of the volatility measured in the Japanese market 7 days before (Table 12.4).

Very similar outcomes are found for the US and Japanese market correlations. When we estimate the correlation breakdowns, we obtain a lower

Table 12.3 Regression equation of correlation EUR–JPY changes explained by volatility differences

Variable	Coefficient	Std. error	t-statistic	Prob.
DEURV	3.136068	0.655131	4.786932	0.0000
DJPYV	3.246925	0.580543	5.592912	0.0000
R-squared	0.056743	Mean dependent var		0.221726
Adjusted *R*-squared	0.055047	S.D. dependent var		0.588052
S.E. of regression	0.571638	Akaike info criterion		1.722956
Sum squared residuals	181.6841	Schwarz criterion		1.738456
Log likelihood	−478.7048	F-statistic		33.44724
Durbin–Watson stat.	1.932469	Prob(F-statistic)		0.000000

Notes: The dependent variable; differences of exponential correlations between the equity euro market and the equity Japanese market. Explanatory variables are DEURV: differences of exponential volatility of the equity euro market; and DJPYV: differences of exponential volatility of the equity Japanese market. Number of observations = 558 daily changes.

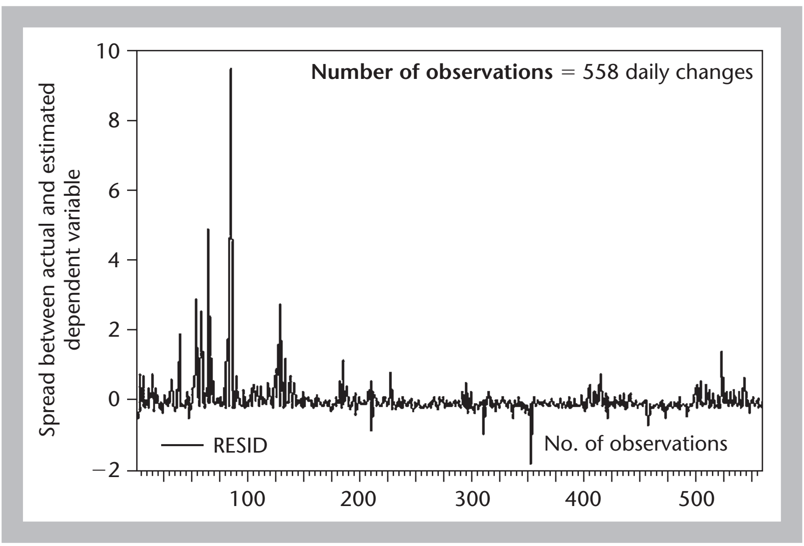


Figure 12.6 Plot of residuals of correlation EUR–JPY changes explained by volatility differences

value of the sum of squared residuals which is roughly 30 percent lower (from 43.136 to 30.321). *R*-squared is 95 percent higher in the shorter time series (0.099577 instead of 0.004663). The complete time series has been estimated twice, the first time using the two changes in volatilities. The volatility of the Japanese market is statistically non-significant (student

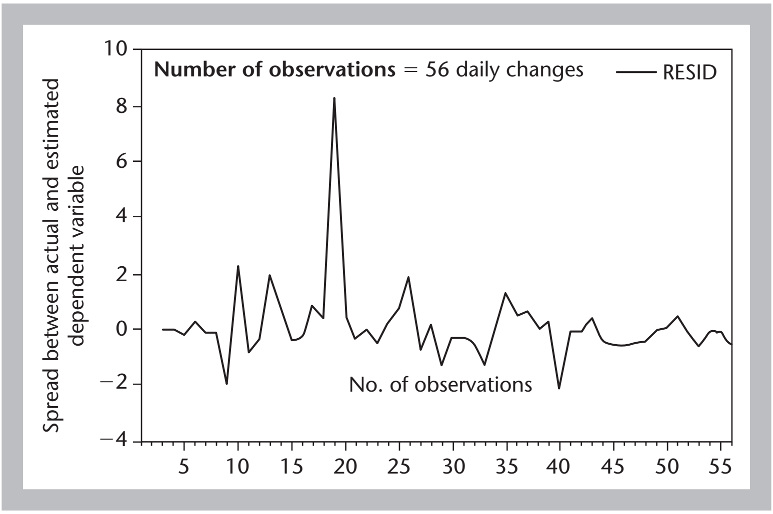


Figure 12.7 Plot of residuals of correlation EUR–JPY higher changes explained by volatility differences

Table 12.4 Regression equation of higher correlation EUR–JPY changes explained by volatility differences

Variable	Coefficient	Std. error	t-statistic	Prob.
DEURV7	22.98515	5.336414	4.307229	0.0001
DJPYV6	17.61808	5.908526	2.981807	0.0044
R-squared	0.113741	Mean dependent var		1.296226
Adjusted *R*-squared	0.096698	S.D. dependent var		1.472966
S.E. of regression	1.399940	Akaike info criterion		3.547069
Sum squared residuals	101.9112	Schwarz criterion		3.620735
Log likelihood	−93.77087	F-statistic		6.673604
Durbin–Watson stat.	1.920604	Prob(F-statistic)		0.012636

Notes: The dependent variable; differences of exponential correlations between the equity euro market and the equity US market. Explanatory variables are DEURV7: differences of exponential volatility of the equity euro market with 7 days lag; DJPYV6: differences of exponential volatility of the equity US market with 6 days lag. Number of observations = bigger 56 daily changes.

t equals 0.286315 in Table 12.5). Table 12.6 contains the regression estimates explained only by differences of exponential volatility of the equity US market and Figure 12.8 displays the plot of residuals of correlation between the USD–JPY changes explained by volatility differences. As in the other cases, Table 12.7 shows how statistics improve using only the highest correlation changes between the US and the Japanese market. Figure 12.9 displays

Table 12.5 Regression equation of correlation USD–JPY changes explained by volatility differences

Variable	Coefficient	Std. Error	t-statistic	Prob.
DUSDV	14.29453	8.641957	1.654085	0.0987
DJPYV	2.155165	7.527247	0.286315	0.7747
R-squared	0.004810	Mean dependent var		0.135836
Adjusted *R*-squared	0.003020	S.D. dependent var		8.821442
S.E. of regression	8.808111	Akaike info criterion		7.192801
Sum squared residuals	43136.04	Schwarz criterion		7.208300
Log likelihood	−2004.791	F-statistic		2.687343
Durbin–Watson stat.	1.988807	Prob(F-statistic)		0.101714

Notes: The dependent variable; differences of exponential correlations between the equity US market and the equity Japanese market. Explanatory variables are DUSDV: differences of exponential volatility of the equity US market; and DJPYV: differences of exponential volatility of the equity Japanese market. Number of observations: 558 daily changes.

Table 12.6 Regression equation of correlation USD–JPY changes explained by volatility differences

Variable	Coefficient	Std. error	t-statistic	Prob.
DUSDV	14.29993	8.634812	1.656078	0.0983
R-squared	0.004663	Mean dependent var		0.135836
Adjusted *R*-squared	0.004663	S.D. dependent var		8.821442
S.E. of regression	8.800849	Akaike info criterion		7.189364
Sum squared resid	43142.40	Schwarz criterion		7.197114
Log likelihood	−2004.833	Durbin–Watson stat		1.989381

Notes: The dependent variable; differences of exponential correlations between the equity US market and the equity Japanese market. Explanatory variables are DUSDV: differences of exponential volatility of the equity US market. Number of observations = 558 daily changes.

the plot of residuals of correlation between the USD–JPY higher changes explained by volatility differences.

Our estimates demonstrate that correlation jumps can be modeled rather than the complete time series in all the reported cases. Table 12.8 shows the measures of the improvement obtained for all the correlations. As for out of sample predictability in our daily intermarket correlation series, the results reported in Table 12.9 are coherent with those obtained in the sample period. In particular, correlation jumps mean absolute errors are everywhere lower

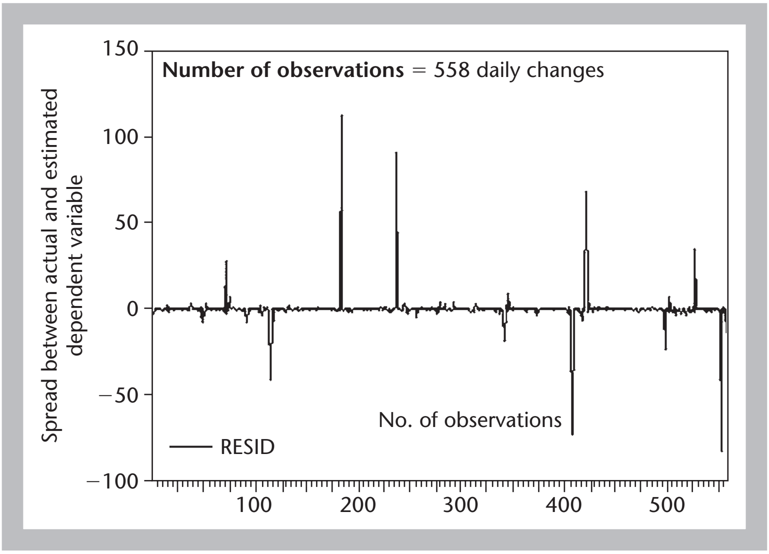


Figure 12.8 Plot of residuals of correlation USD–JPY changes explained by volatility differences

Table 12.7 Regression equation of correlation USD–JPY changes explained by volatility differences

Variable	Coefficient	Std. error	t-statistic	Prob.
DJPYV1	252.7362	131.7969	1.917619	0.0605
DJPYV5	276.4569	144.7156	1.910346	0.0614
R-squared	0.099577	Mean dependent var		13.08877
Adjusted *R*-squared	0.082902	S.D. dependent var		24.74389
S.E. of regression	23.69604	Akaike info criterion		9.203554
Sum squared residuals	30321.12	Schwarz criterion		9.275888
Log likelihood	−255.6995	F-statistic		5.971798
Durbin–Watson stat.	2.022590	Prob(F-statistic)		0.017834

Notes: The dependent variable; differences of exponential correlations between the equity US market and the equity Japanese market; Explanatory variables are DJPYV1: differences of exponential volatility of the equity Japanese market with 1 day lag; and DJPYV5: differences of exponential volatility of the equity Japanese market with 5 days lag. Number of observations = 56 daily changes.

than the corresponding errors for the whole time series. It is necessary to note, on the other hand, that the correlation higher changes analysis refers to series composed of only 10 data points and it is not useful in predicting the time point in the future where these changes will occur.

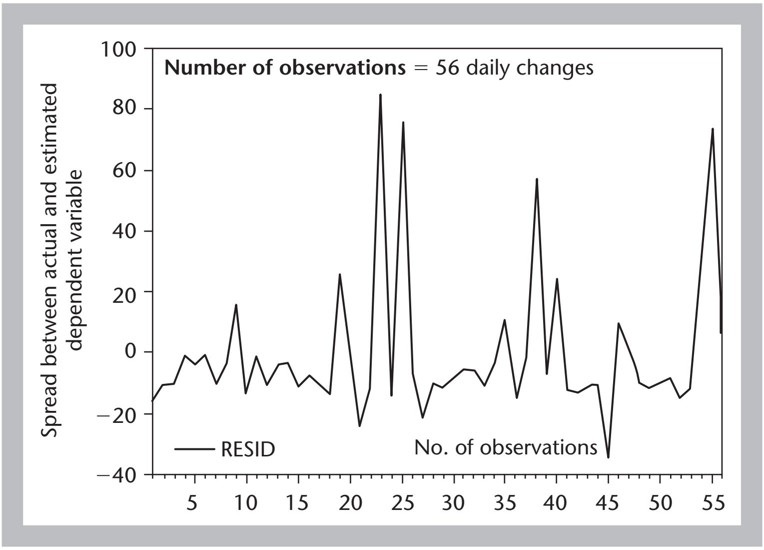


Figure 12.9 Plot of residuals of correlation USD–JPY higher changes explained by volatility differences

Table 12.8 Improvement of statistical tests applied to correlation jumps

	EUR/USD	EUR/JPY	USD/JPY
Sum squared residuals	81.50%	43.91%	29.71%
R-squared	63.95%	50.11%	95.32%

Notes: *R*-squared is the statistical measure of the success of the regression in predicting the values of the dependent variable within the sample. This measure is the fraction of the variance of the dependent variable explained by the independent variables. The statistic will equal one if the regression fits perfectly, and zero if it fits no better than the simple mean of the dependent variable. All these values are obtained as the complement to 1 of the ratio between the test of the complete time series and the same test of the higher correlations.

Table 12.9 Out of sample model efficiency: mean absolute error

	All the correlations		
MAE	**EUR/USD**	**EUR/JPY**	**USD/JPY**
	11.38%	35.05%	66.41%
	Correlations higher changes		
MAE	**EUR/USD**	**EUR/JPY**	**USD/JPY**
	8.47%	29.61%	47.15%

Note: MAE = mean absolute error.

12.4 IMPACT ON PORTFOLIO OPTIMIZATION

This last section is dedicated to a portfolio optimization simulation for the three equity markets considered. We optimize three different efficient frontiers:

1 The first frontier is determined by optimizing the historical returns along with volatilities and the correlation matrix (Table 12.10, panel A).

2 The second frontier is determined by optimizing the historical returns and volatilities. The correlation has been changed with the maximum negative jumps observed during the historical window, that is 1 January 2003 until 31 March 2005 (Table 12.10, panel B).

3 The third frontier is also determined by optimizing the historical returns and volatilities. The correlation has been changed with the maximum positive jumps observed during the historical period (Table 12.10, panel C).

The outcomes are reported in Table 12.11. There are 100 optimized portfolios, and have reported portfolio *n*. The number 1 implies that it is the less hazardous, and the number 75 avoids the concentration problem which characterizes all the optimizations *à la* Markowitz. Our results demonstrate that by introducing correlation jumps, in both directions, we would obtain portfolios with volatilities that in the left side (less risky) could have changed in a range of 3.49 percent (20.09–16.6). In the right side (portfolio no. 75) the range of volatility is 1.72 percent (23.67–21.95).

12.5 CONCLUSION

The possible occurrence of correlation jumps would considerably change the profile of the investor. Our statistical results, presented in the third section, demonstrate that not only correlations can change in the short run, but also, in some cases, these events occur with volatility shocks. Correlation breakdowns should be considered in order to evaluate:

(a) the impact for portfolios volatilities estimated by investors and their private bankers;

(b) the decisions of portfolio managers; and

(c) the monitoring process of risk managers.

The answers offered by the study are:

1 there is a relation between exponential correlation changes and volatility movements, even though it depends on the market where outcomes are estimated;

Table 12.10 Optimization inputs

	Panel A: optimization no. 1				
			Historical correlation		
	Historical return	Historical volatility	Topix	S&P 500 Index	DJ Euro Stoxx 50
Topix	6.15	20.80	1		
S&P 500 Index	7.10	25.38	0.55	1	
DJ Euro Stoxx 50	6.59	23.52	0.45	0.81	1

Notes: First column: historical average returns determined in the long run (1976–2005) for Topix and S&P 500 Composite; the Euro Stoxx has been computed for the period 1976–2004.
Second column: historical average volatility determined in the long run (1976–2005) for Topix and S&P 500 Composite; the Euro Stoxx has been computed for the period 1976–2004. The methodology is the exponential weighted with a 0.94 decay factor.
Third to fifth column: correlation matrix determined in the long run (1976–2005) for Topix and S&P 500 Composite; the Euro Stoxx has been computed for the period 1976–2004. The methodology is the exponential weighted with a 0.94 decay factor.

	Panel B: optimization no. 2				
			Historical correlation		
	Historical return	Historical volatility	Topix	S&P 500 Index	DJ Euro Stoxx 50
Topix	6.15	20.80	1		
S&P 500 Index	7.10	25.38	0.34	1	
DJ Euro Stoxx 50	6.59	23.52	0.14	0.61	1

Note: Third to fifth column: correlation matrix determined applying the max negative jumps observed in the 2003–05 period.

	Panel C: optimization no. 3				
			Historical correlation		
	Historical return	Historical volatility	Topix	S&P 500 Index	DJ Euro Stoxx 50
Topix	6.15	20.80	1		
S&P 500 Index	7.10	25.38	0.8	1	
DJ Euro Stoxx 50	6.59	23.52	0.69	0.96	1

Note: Third to fifth column: correlation matrix determined applying the max positive jumps observed in the 2003–05 period.

Table 12.11 Optimization outputs

Frontier no.		Portfolio 1	Portfolio 75
1 (no jumps)	Return volatility	6.32	6.91
		18.70	22.83
2 (negative jumps)	Return volatility	6.37	6.92
		16.60	21.95
3 (positive jumps)	Return volatility	6.28	6.90
		20.09	23.67

Notes: The table shows the return/risk values of the first optimization obtained with the historical average values; the second optimization obtained with negative correlation jumps; and the first optimization, obtained with positive correlation jumps.

2 this relation is, on a qualitative basis, demonstrated in all our analysis; the benefits obtained in the correlation-jumps dataset are statistically significant; and

3 in analysing only equity portfolios, errors we could make in ignoring daily correlation jumps are around 18 percent (3.49 over 18.7) of volatility in the less risky part of the frontier, and 7.5 percent (1.72 over 22.83) in the most risky portfolios.

NOTES

1. If the (DW) is less than 2 (until 0), there is evidence of positive serial correlation. If there is no serial correlation, the DW statistic will be around 2. Finally, if there is negative correlation, the statistic will lie somewhere between 2 and 4
2. The other limitations are: (a) the distribution of the DW statistic under the null hypothesis depends on the data matrix. The usual approach to handling this problem is to place bounds on the critical region, creating a region where the test results are inconclusive; (b) me may only test the null hypothesis of no serial correlation against the alternative hypothesis of first-order serial correlation.

REFERENCES

Brock, W.A., Dechert, W.D. and Scheinkman, J.A. (1987) "*A Test for Independence Based on the Correlation Dimension*", Working Paper, University of Houston and University of Chicago.

Erb, C.B., Harvey, C.R. and Viskanta, T.E. (1994) "Forecasting International Equity Correlations", *Financial Analysts Journal*, 50(6): 32–45.

Fong, W.M. (2003) "Correlation Jumps", *Journal of Applied Finance*, 13(2): 29–45.

Groenen, P.J.F. and Franses, P.H. (2000) "Visualizing Time-Varying Correlations across Stock Markets", *Journal of Empirical Finance*, 7(2): 155–72.

Longin, F. and Solnik, B. (1995) "Is the Correlation in International Equity Returns Constant?" *Journal of International Money and Finance*, 14(1): 3–26.

Longin, F. and Solnik, B. (2001) "Extreme Correlation of International Equity Markets", *Journal of Finance*, 56(2): 649–76.

Loretan, M. and English, W.B. (2000) "Evaluating Correlation Breakdowns During Periods of Market Volatility", *International Finance Discussion Papers no.* 658, Board of Governors of the Federal Reserve System.

Pagnoni, C. and Gabbi, G. (2001) "Tactical Asset Allocation and Time Diversification for Bond and Equity Markets", in A. Smith and B. Jones (eds), *Forecasting Financial Markets*, Presentation at the Pensions Institute: London, May 30–June 1, 2001, httpp://www.pensions-institute.org/conferences/ffm20001.pdf.

RiskMetrics (1996) *RiskMetrics – Technical Document*, 4th edn (New York, J.P. Morgan).

Solnik, B., Boucrelle, C. and Le Fur, Y. (1996) "International Correlation and Volatility", *Financial Analysts Journal*, 52(5): 17–34.

CHAPTER 13

Sequential Procedures for Monitoring Covariances of Asset Returns

Olha Bodnar

13.1 INTRODUCTION

Time variability of the expected returns as well as the volatility of asset returns can be caused by changes in the fundamental factors; for example, changes in commodity prices, macroeconomic policy, market trading activity, technological development, governmental policies, and so on. This leads to the deviation of a selected optimal portfolio from the Markowitz efficient frontier that consists of all portfolios with the highest expected return for the given level of risk or with the smallest risk for a preselected profit and, thus, is fully defined by the first two moments of asset returns (Markowitz, 1952). Changes in these characteristics are subject to structural breaks of the efficient frontier location in the mean–variance space and the optimal portfolios allocated on it.

For a long time financial studies have been concentrated on the proper estimation of the covariance matrix of asset returns. To reduce the estimation error in the covariance matrix of asset returns when the sample portfolio is constructed, Ledoit and Wolf (2003, 2004) proposed the shrinkage estimator of the covariances. In another branch of studies (Andersen, Bollerslev, Diebold and Ebens, 2001; Barndorff-Nielsen and Shephard, 2002, 2004) the returns with the higher frequency are used for calculating the covariance matrix. The question of interest is not only estimation, but also monitoring the break in the covariance structure of asset returns. We want to consider this problem in the present chapter in more detail.

Statistical control methods have been recently developed for monitoring the mean vector and covariance matrix of a random vector (see for example Bodnar, 2005, for a detailed survey). The dimension of the control problem depends on the number of assets included in the portfolio and could be extremely large. This leads to a delay in detecting changes in the portfolio structure and can lead to large losses for the investor. In order to reduce the dimensionality of the control problem the optimal portfolio weights (parts of the investor's wealth allocated into an asset) are considered in this chapter. We show that the changes in the covariance matrix of asset returns generates changes in the weights of the global minimum variance portfolio (GMVP).

Sequential procedures for monitoring financial data-sets have already been discussed in the financial literature. For instance, Theodossiou (1993) applied multivariate CUSUM control charts for predicting business failures. Financial decision strategies based on the MEWMA control scheme are discussed by Schipper and Schmid (2001) and Schmid and Tzotchev (2004). Andersson, Bock and Frisen (2003, 2005) made used of the statistical surveillance for the detection of turning points in business cycles. However, nobody up to now has adopted sequential procedures in asset management, with the exception of Yashchin, Steinand and Philips (1997). It is our aim, based on the historical values of the asset returns process, to derive sequential control schemes for monitoring changes in the covariance matrix of asset returns that could influence the selection of an optimal portfolio. In order to reduce the dimensionality of the control problem we focus essentially on the transformation of the vector of the optimal portfolio weights. A great advantage of this suggested approach is that structural breaks in the covariance matrix lead to shifts in both the mean vector and covariance matrix of this transformed vector. It also possesses several nice distributional properties and thus can be easily monitored in practice. We develop the corresponding sequential procedures.

The remainder of the chapter is organized as follows. In the next section we shed light on the relationship between the covariance matrix of asset returns and the weights of the GMVP. In Theorem 1 it is shown that the simple transformation of the estimator for the GMVP weights is multivariate t-distributed. Theorem 2 motivates the application of the control charts to the transformed vector of the portfolio weights. The link between covariances and portfolio weights is investigated. The multivariate statistical surveillance is introduced in section 13.2, while section 13.3 deals with the simultaneous control schemes. The two approaches are compared in section 13.4. As the measure of the control charts performance the average run length (ARL) is used. As no explicit formula for the ARL is available we estimate this quantity within an extensive Monte Carlo study. Final remarks are presented in section 13.5. The proofs of all results are given in the Appendix.

13.2 COVARIANCE STRUCTURE OF ASSET RETURNS AND OPTIMAL PORTFOLIO WEIGHTS

We consider a portfolio consisting of p assets. The weight of the i-th asset in the portfolio is denoted by w_i. The p-dimensional vector of portfolio weights $\mathbf{w} = (w_1, \dots, w_p)'$ specifies the investment policy. It is assumed that the whole investor's wealth is shared between the selected assets and the possibility of short-selling. Mathematically, it means that the sum of the weights is one, $\mathbf{w}'\mathbf{1} = 1$, where $\mathbf{1}$ denotes the p-dimensional vector of ones, and the portfolio weights are not obviously positive.

Suppose the vector of asset returns possesses the second moment. Its mean we denote by μ and the covariance matrix by Σ, which is assumed to be positive definite. Then the expected return of the portfolio is given by $\mathbf{w}'\mu$ and its variance is equal to $\mathbf{w}'\Sigma\mathbf{w}$. Following the seminal paper of Markowitz (1952) and his mean–variance analysis, the optimal portfolio is selected by minimizing the portfolio variance for the given level of portfolio return, or by maximizing the expected return for the given risk.

In the present study we make use of the weights of the global minimum variance portfolio (GMVP) for monitoring the covariance structure of asset returns. The main advantage of the suggested approach is that we control only the $(p-1)$ -dimensional vector of the portfolio weights instead of monitoring the $p(p+1)/2$ vector of the variances and covariances (see Bodnar (2005) for details). It leads to the significant reduction of the dimensionality of the control problem and improves considerably its power properties.

The weights of the GMVP are:

$$\mathbf{w}_M = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} \tag{13.1}$$

obtained by minimizing the portfolio's variance subject to $\mathbf{w}'\mathbf{1} = 1$. This portfolio corresponds to the case of a fully risk averse investor. Clearly more complex portfolio weights could be used for monitoring the covariance matrix of asset returns, like the weights of the optimal portfolio in the sense of maximizing the expected quadratic utility, the weights of the tangency portfolio, and others. The goal however is to keep this simple.

The situation becomes more difficult with the practical implementation of the model. The covariance matrix Σ of asset returns is usually an unknown parameter, and, thus, the investor cannot determine his portfolio policy. Instead he has to estimate the quantity by previous observations. This approach leads to heavy discussion in financial and econometric literature.

The different estimation procedures and their influences on the distributional properties of the estimator for the optimal portfolio weights has been discussed (see, for example, Scwert (1989), Ledoit and Wolf (2003, 2004), Bodnar and Schmid (2004), Kan and Zhou (2004) and references therein). For our purposes, given the sample of portfolio asset returns $\mathbf{X}_1, \ldots, \mathbf{X}_n$, the most common estimator of Σ is chosen:

$$\hat{\Sigma} = \frac{1}{n-1} \sum_{t=1}^{n} (\mathbf{X}_t - \overline{\mathbf{X}})(\mathbf{X}_t - \overline{\mathbf{X}})' = \frac{1}{n-1} \mathbf{X}(\mathbf{I} - \frac{1}{n} \mathbf{1}\mathbf{1}')\mathbf{X}' \tag{13.2}$$

with $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ and $\overline{\mathbf{X}} = \mathbf{X}\mathbf{1}/n$. Then using the standard plug-in portfolio rule, for example, replace Σ in (13.1) by $\hat{\Sigma}$, the estimator for the GMVP is given by:

$$\hat{\mathbf{w}}_M = \frac{\hat{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}'\hat{\Sigma}^{-1}\mathbf{1}} \tag{13.3}$$

Bodnar and Schmid (2004) showed that a q-dimensional vector of the estimator for the linear combinations of the GMVP weights, $\mathrm{L}\hat{\mathbf{w}}_M$, follows the multivariate t-distribution with the mean vector $\mathbf{L}\mathbf{w}_M$ and the covariance matrix $\frac{1}{n-p+1}$ $\mathbf{L}\mathbf{R}\mathbf{L}'/\mathbf{1}'\,\Sigma^{-1}\mathbf{1}$, where $\mathbf{L}$ denotes the $q \times p$ dimensional matrix of constants. This assertion we denote by $\mathbf{L}\hat{\mathbf{w}}_M \sim t_q$ $(n-p+1, \mathbf{L}\mathbf{w}_M, \frac{1}{n-p+1}$ $\mathbf{L}\mathbf{R}\mathbf{L}'/\mathbf{1}'\Sigma^{-1}\mathbf{1})$, where $\mathbf{R} = \Sigma^{-1} - \Sigma^{-}\mathbf{1}\mathbf{1}'\Sigma^{-1}/\mathbf{1}'\Sigma^{-1}\mathbf{1}$. However, it appears that the distribution of the estimator for the GMVP weights does depend on the unknown parameter Σ and thus cannot be directly monitored. To avoid the problem in the paper it is proposed to make use of the simple transformation of the estimator given by:

$$\hat{\mathbf{v}} = \sqrt{n-p}\sqrt{\mathbf{1}'\hat{\Sigma}^{-1}\mathbf{1}} \left(\mathbf{L}\hat{\Sigma}^{-1}\mathbf{L}' - \frac{\mathbf{L}\hat{\Sigma}^{-1}\mathbf{1}\mathbf{1}'\hat{\Sigma}^{-1}\mathbf{L}'}{\mathbf{1}'\hat{\Sigma}^{-1}\mathbf{1}} \right)^{-\frac{1}{2}} \mathbf{L}(\hat{\mathbf{w}}_M - \mathbf{w}_M) \tag{13.4}$$

In Theorem 1 it is shown that its finite sample distribution depends only on the current GMVP weights and is independent of Σ. We also preserve the nice properties of the t-distribution.

> **Theorem 1** Let the vectors of portfolio asset returns $\mathbf{X}_1,\ldots,\mathbf{X}_\mathrm{n}$ be independent and identical, normally distributed with the mean vector μ and covariance matrix Σ. Let $n \geq p > q \geq 1$ and $n > p + 2$. Assume Σ to be positive definite. Then the vector $\hat{\mathbf{v}}$ has a multivariate t-distribution with $n-p$ degrees of freedom, the mean vector $\mathbf{0}$ and the covariance matrix $(n-p)\mathbf{I}/(n-p-2)$.

Theorem 1 leads to several interesting procedures that are easily implemented in practice. First, it allows us to construct the confidence intervals for the GMVP weights. For example, taken $\mathbf{L} = \mathbf{e}_1 = (1,0,\ldots,0)$ as the p dimensional vector with the first element being 1 and the rest zeros, we obtain the two-sided $1-\alpha$ confidence interval for the first weight of the GMVP:

$$\left[\hat{w}_{M;1} - \frac{t_{n-p;1-\alpha/2}}{\sqrt{n-p}}\sqrt{\frac{\mathbf{e}_1'\hat{\mathbf{R}}\mathbf{e}_1}{\mathbf{1}'\hat{\Sigma}^{-1}\mathbf{1}}}, \hat{w}_{M;1} + \frac{t_{n-p;1-\alpha/2}}{\sqrt{n-p}}\sqrt{\frac{\mathbf{e}_1'\hat{\mathbf{R}}\mathbf{e}_1}{\mathbf{1}'\hat{\Sigma}^{-1}\mathbf{1}}}\right]$$

where $t_{n-p;1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard univariate t-distribution. Consequently it can be used as a tool for controlling the weights of the GMVP. It leads to a decision whether the portfolio should be adjusted or not.

Second, it permits us to apply the sequential control procedures for monitoring the efficiency of the GMVP using the distributional properties of the random vectors $\hat{\mathbf{v}}$.

Third, which is also a main feature for our purposes, the breaks in the covariance matrix of asset returns lead to the changes in the mean vector and the covariance matrix of $\hat{\mathbf{v}}$. Thus, by monitoring these two parameters we control both covariance structure of returns and the efficiency of the GMVP. To show this we need the following result. It is assumed that $\mathbf{X}_t \sim N_p(\mu,\Sigma)$ for $t \le t_0$, and $\mathbf{X}_t \sim N_p(\tilde{\mu}, \tilde{\Sigma})$ for $t > t_0$, which leads to the following GMVP weights:

$$\mathbf{w} = \Sigma^{-1}\mathbf{1}/\mathbf{1}'\Sigma^{-1}\mathbf{1} \quad \text{for} \quad t \le t_0 \quad \text{and} \quad \tilde{\mathbf{w}} = \tilde{\Sigma}^{-1}\mathbf{1}/\mathbf{1}'\tilde{\Sigma}^{-1}\mathbf{1} \quad \text{for} \quad t > t_0$$

The vector $\hat{\tilde{\mathbf{v}}}$ we define similar to $\hat{\mathbf{v}}$ using $\hat{\tilde{\Sigma}}^{-1}$ instead of $\hat{\Sigma}$ with $\hat{\tilde{\Sigma}}^{-1}$ being the estimator of $\tilde{\Sigma}$ (see (13.2). In Theorem 2 the influence of the changes in the covariance structure of assets returns on the mean vector and the covariance matrix of the vector $\hat{\tilde{\mathbf{v}}}$ is presented:

Theorem 2 Let the vectors of portfolio asset returns $\mathbf{X}_1,\ldots,\mathbf{X}_n$ be independent and identical normally distributed with the mean vector μ and covariance matrix Σ. Let $n \ge p > q \ge 1$ and $n > p+2$. Assume Σ and $\tilde{\Sigma}$ to be positive definite. Then:

(a) The expectation of $\hat{\tilde{\mathbf{v}}}$ is equal to

$$E(\hat{\tilde{\mathbf{v}}}) = \sqrt{n-p}\sqrt{\tilde{\mathbf{H}}_{22}^{(-)}}\,\tilde{b}^{-\frac{1}{2}}\,\mathbf{A}(\tilde{\mathbf{w}}_{M;q} - \mathbf{w}_{M;q})$$

where $\mathbf{A} = \text{diag}(a_{11}, \ldots, a_{qq})$ with (13.5)

$$a_{ii} = \frac{B\left(\frac{n-p-q-i}{2}+1, \frac{n-p-1}{2}\right)}{B\left(\frac{n-p-q-i+1}{2}, \frac{n-p}{2}\right)}, \; i = \overline{1,q}.$$

(b) The covariance matrix of $\hat{\hat{\mathbf{v}}}$ is equal to:

$$Var(\hat{\hat{\mathbf{v}}}) = \frac{n-p}{2}\left(\frac{\Gamma\left(\frac{n-p}{2}-1\right)}{\Gamma\left(\frac{n-p}{2}\right)}\left(\mathbf{I}+\tilde{\mathbf{H}}_{22}^{(-)}\tilde{b}^{-\frac{1}{2}}\mathbf{G}\tilde{b}^{-\frac{1}{2}\prime}\right)\right.$$
$$\left.+\left(\frac{\Gamma\left(\frac{n-p-1}{2}\right)}{\Gamma\left(\frac{n-p}{2}\right)}\right)^2\tilde{\mathbf{H}}_{22}^{(-)}\tilde{b}^{-\frac{1}{2}}\mathbf{F}\tilde{b}^{-\frac{1}{2}\prime}\right)$$

where the matrices **G** and **F** are given in the Appendix.

We make use of these results in the next section, where the multivariate and simultaneous control charts are constructed for detecting changes in the covariances of returns.

13.3 MULTIVARIATE STATISTICAL SURVEILLANCE

The covariance matrix of asset return for a given horizon of interest is estimated from the returns of higher frequency. Using non-overlapping samples of data permits us to construct a sequence of covariance matrix estimators that are independent through time. This point has already been discussed in the financial literature. For instance, Schwert (1989) proposed estimating the variance of monthly returns using daily data, while Andersen, Bollerslev, Diebold and Ebens (2001) made use of this approach for the approximation of daily variances and covariances from high-frequency return data.

For a given sample of asset returns $\mathbf{X}_1,\ldots,\mathbf{X}_n$ we constrain m subsamples of size $\tilde{n}$, that is $\{\mathbf{X}_{(1);j}\}_{j=1}^{\tilde{n}}, \{\mathbf{X}_{(2);j}\}_{j=1}^{\tilde{n}}, \ldots, \{\mathbf{X}_{(m);j}\}_{j=1}^{\tilde{n}}$, where $\mathbf{X}_{(i);j} = \mathbf{X}_{\tilde{n}(i-1)+j}, i = 1, \ldots, m$ and $j = 1, \ldots, \tilde{n}$. For the i-th subsample the estimators for Σ, $\mathbf{w}$, and $\mathbf{v}$, namely $\hat{\Sigma}_{(i)}$, $\hat{\mathbf{w}}_{(i)}$, and $\hat{\mathbf{v}}_{(i)}$, are defined using (13.2), (13.3), and (13.4) correspondingly. From Theorem 1 it follows that under the assumption of no change in the covariance matrix of returns $\hat{\mathbf{v}}_{(i)}$ are independently identically t -distributed with $\tilde{n}-p$ degrees of freedom, with the mean vector $\mathbf{0}$, and the covariance matrix $(\tilde{n}-p)\mathbf{I}/(\tilde{n}-p-2)$. We consider the $q+q(q+1)/2$ dimensional vector:

$$\eta_{(i)} = (\hat{\mathbf{v}}_{(i);1},\ldots,\hat{\mathbf{v}}_{(i);q},\hat{\mathbf{v}}_{(i);1}\hat{\mathbf{v}}_{(i);1},\hat{\mathbf{v}}_{(i);1}\hat{\mathbf{v}}_{(i);2},\ldots,\hat{\mathbf{v}}_{(i);q-1}\hat{\mathbf{v}}_{(i);q},\hat{\mathbf{v}}_{(i);q}\hat{\mathbf{v}}_{(i);q}), \quad (13.6)$$

which is used to construct control schemes for detection changes in the mean vector and the covariance matrix of $\hat{\mathbf{v}}_{(i)}$ simultaneously. In case of no breaks in the returns covariances the expected value of the vectors $\eta_{(i)}$ is:

$$\mu_\eta = E(\eta_{(i)}) = \left(0,\ldots,0,\frac{\tilde{n}-p}{\tilde{n}-p-2},0,\ldots,0,\frac{\tilde{n}-p}{\tilde{n}-p-2}\right) \quad (13.7)$$

where all elements of μ_η at positions $q+1$, $2q+1$, $3q$, $4q-2$, $5q-3,\ldots,q+q(q+1)2$ are equal to $(\tilde{n}-p)/(\tilde{n}-p-2)$ and the others are zero. Furthermore, let denote $\Theta=\{1,\, q+1,\, 2q,\, 3q-2,\, 4q-3,\ldots,q(q+1)2\}$. Then it follows that the covariance matrix of the vector $\eta_{(i)}$, $i=1,\ldots,m$, in the in-control state is:

$$\Sigma_\eta = \begin{pmatrix} \dfrac{\tilde{n}-p}{\tilde{n}-p-2}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \tilde{\Sigma}_\eta \end{pmatrix} \tag{13.8}$$

with $B(.,.)$ being the Beta function and

$$\tilde{\Sigma}_\eta = \begin{pmatrix} a & 0 & \ldots & 0 & b & \ldots & b \\ 0 & c & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & & \ddots & & \vdots & & \\ 0 & 0 & \ldots & c & 0 & \ldots & 0 \\ b & 0 & \ldots & 0 & a & \ldots & b \\ \vdots & & \ldots & & \vdots & \ddots & \\ b & 0 & \ldots & 0 & b & \ldots & a \end{pmatrix} \tag{13.9}$$

The elements are equal to $a=3(\tilde{n}-p)^2/((\tilde{n}-p-2)(\tilde{n}-p-4))$ at positions $(j,\, j)$, where $j\in\Theta$, $b=(\tilde{n}-p)^2/((\tilde{n}-p-2)(\tilde{n}-p-4))-((\tilde{n}-p)/(\tilde{n}-p-2))^2$ at positions $(j_1,\, j_2)$, where $j_1,\, j_2\in\Theta$ and $j_1\neq j_2$, $c=(\tilde{n}-p)^2/((\tilde{n}-p-2)$ $(\tilde{n}-p-4))$ at positions (j,j), where $j\in\{1,\ldots,\, q(q+1)/2\}\backslash\Theta$, and the others are zero (Fang and Zhang, Lemma 5.6.3, 1990).

To test if there is a change in the covariance matrix Σ requires checking whether the mean of the vector $\eta_{(i)}$, whose estimator is given in (13.6), departs significantly from μ_η. Hence, at time t the testing problem is given by:

$$H_{0,t}:\ E(\eta_{(t)}) = \mu_\eta \quad \text{against} \quad H_{1,t}:\ E(\eta_{(t)}) = \mu_1 \neq \mu_\eta \tag{13.10}$$

No change occurs in the covariance matrix of asset returns and, correspondingly, the GMVP is efficient up to time t if H_0 is valid for all $s\in 1,\ldots,t$. We refer to this as an in-control state. In other case the portfolio is not longer efficient, starting at the time, when the first shift has occurred. The corresponding state of the system is called out-of-control. It also means that an investor has to adjust his portfolio. As a measure of the performance of control schemes the average run length (ARL) is chosen. By this criteria the best control chart possesses the smallest out-of-control ARL for a given in-control one. Following Schipper and Schmid (2001) the in-control ARL is taken 60 that corresponds to three months returns data.

13.3.1 T^2 Control chart

The multivariate Shewhart control chart (Hotelling, 1947) is based on measure the distance between the vector of observation and the target mean of the process. Because μ_η is known (see (13.7)), the control statistic is given by:

$$T_t^2 = (\eta_{(t)} - \mu_\eta)' \Sigma_\eta^{-1} (\eta_{(t)} - \mu_\eta)$$

The null hypothesis of no changes should be rejected at time t if $T_t^2 > h_1$, which is determined under the condition that the in-control ARL is equal to a preselected value ξ. For further discussion of the properties of the T^2 chart we refer to Alt and Smith (1988) and Runger, Alt and Montgomery (1996).

13.3.2 MC1

The most frequent application of CUSUM schemes is the case when a detection of small changes in the process parameter is of interest. An important property of the CUSUM chart for univariate data is that it can be derived via the sequential probability ratio test of Wald (1947). In the multivariate case it leads to a control scheme which depends on the direction of the shift (Healy, 1987). Because this is a very unpleasant property, three other types of CUSUM charts for multivariate observations have been introduced by Crosier (1988) and Pignatiello and Runger (1990), Ngai and Zhang (2001). A more detailed discussion of these schemes can be found in Bodnar and Schmid (2006).

Let $\mathbf{S}_{m,l} = \sum_{i=m+1}^{l} (\eta_{(i)} - \mu_\eta)$ for $l, m \geq 0$. For an arbitrary vector $\mathbf{x}$ and a positive definite matrix $\mathbf{C}$ we define the norm $\|\mathbf{x}\|_\mathbf{C} = \sqrt{\mathbf{x}'\mathbf{C}^{-1}\mathbf{x}}$.

The MC1 control chart is constructed by applying the MC1 control scheme of Pignatiello and Runger (1990) to the process $\eta_{(t)}$:

$$MC1_t = \max\left(\|\mathbf{S}_{t-n_t,t}\|_{\Sigma_\eta} - kn_t, 0\right), \quad t \geq 1$$

where n_t being the number of observations since the most recent reconstruction of the CUSUM chart:

$$n_t = \begin{cases} n_{t-1} + 1 & \textit{if } MC1_{t-1} > 0 \\ 1 & \textit{if } MC1_{t-1} = 0 \end{cases} \tag{13.11}$$

for $t \geq 1$ with $MC1_0 = 0$.

When $MC1_t$ exceeds an upper control limit then the process is considered to be off-target. For the interpretation of the out-of-control signal one can consider the components of the $\mathbf{S}_{t-n_t,t}/n_t + \mu_\eta$ which indicate the direction of the shift provided that it is not a false alarm.

13.3.3 Vector valued CUSUM

Crosier (1988) proposed the multivariate CUSUM control chart, namely MCUSUM, that is based on the shrinking method. This procedure generalized the univariate proposal of Crosier (1986) to the multivariate situation by replacing the scalar quantities in the univariate CUSUM recursion by the vectors in the multivariate case. The idea is to first update the vector of cumulative sums, then to shrink it towards zero, and, finally, to use the length of the updated and shrunken CUSUM to test whether or not the process is out-of-control.

Let C_t be the length of the vector $\mathbf{S}_{t-1} + (\eta_{(t)} - \mu_\eta)$:

$$C_t = \|\mathbf{S}_{t-1} + (\eta_{(t)} - \mu_\eta)\|_{\Sigma_\eta}$$

$k > 0$ is the reference value. Then the vector-valued CUSUM scheme for the vector $\eta_{(t)}$ is given by:

$$\mathbf{S}_t = \begin{cases} \mathbf{0} & \textit{if } C_t \le k \\ (\mathbf{S}_{t-1} + \eta_{(t)} - \mu_\eta)(1 - \frac{k}{C_t}) & \textit{if } C_t > k \end{cases} \tag{13.12}$$

for $t \ge 1$ with $\mathbf{S}_0 = \mathbf{0}$. The scheme gives an out-of-control signal as soon as the length of the vector $\mathbf{S}_t$:

$$MCUSUM_t = (\mathbf{S}_{t'} \Sigma_\eta^{-1} \mathbf{S}_t)^{\frac{1}{2}} = \max\{0, C_t - k\}$$

exceeds a preselected value h_2, which is determined with the condition that the in-control ARL is equal to a fixed value ξ. In practice the last equation has to be solved by simulations. A practical advantage of the shrinking method is that the components of $\mathbf{S}_t$ give an indication in which direction the mean has shifted, provided that it is not a false alarm.

13.3.4 Projected pursuit CUSUM

An extension of the CUSUM chart, namely PPCUSUM which is based on the idea of projection pursuit, was proposed by Ngai and Zhang (2001). For the direction $\mathbf{a}_0$ with $\|\mathbf{a}_0\|_2 = 1$ (Euclidean norm), we define the CUSUM statistic by:

$$C_0^{\mathbf{a}_0} = 0, \qquad C_t^{\mathbf{a}_0} = \max\{0, C_{t-1}^{\mathbf{a}_0} + \mathbf{a}_0'(\eta_{(t)} - \mu_\eta) - k\}, t \ge 1$$

Pollak (1985) and Moustakides (2004) showed that the univariate CUSUM chart possesses certain optimality properties. If the direction of the shift would be known in the multivariate context then the CUSUM chart based on the projected observations $\mathbf{a}'\mathbf{X}_t$ would reflect this desirable behavior. The problem is that the direction is unknown and therefore the statistic cannot be directly applied. Ngai and Zhang (2001) proposed to solve this problem by estimating $\mathbf{a}_0$ by $\hat{\mathbf{a}}_0$, and approximate $C_t^{\mathbf{a}_0}$ by $C_t^{\hat{\mathbf{a}}_0}$. Here $\hat{\mathbf{a}}_0$ is the value at which $C_t^{a_0}$ attains its maximum on the unit circle, for example, $C_t^{\hat{\mathbf{a}}_0} = \max_{\|\mathbf{a}\|_2=1} C_t^{\mathbf{a}}$. They proved that $\max_{\|\mathbf{a}\|_2=1} C_t^{\mathbf{a}} = PPCUSUM_t$ with:

$$PPCUSUM_t = \max\{0, \|\mathbf{S}_{t-1,t}\|_{\Sigma_\eta} - k, \|\mathbf{S}_{t-2,t}\|_{\Sigma_\eta} - 2k, ..., \|\mathbf{S}_{0,t}\|_{\Sigma_\eta} - tk\} \tag{13.13}$$

for $t \geq 1$. $\mathbf{S}_{t-v,t}$ and $\|\mathbf{S}_{t-v,t}\|_{\Sigma_\eta}$ are defined in section 3.2.

When a process is appeared to be out-of-control at time t_0, then it exists $t_1 < t_0$ such that $\sqrt{\mathbf{S}'_{t_1,t_0}\Sigma_\eta^{-1}\mathbf{S}_{t_1,t_0}} - (t_0 - t_1)k = \max_{\|\mathbf{a}\|_2=1} C_{t_0}^{\mathbf{a}} > h_3$, where h_3 is a preselected value. Then the direction of the shift is estimated by:

$$\hat{\mathbf{a}}_0 = \frac{\Sigma_\eta^{-\frac{1}{2}}\mathbf{S}_{t_1,t_0}}{\mathbf{S}'_{t_1,t_0}\Sigma_\eta^{-1}\mathbf{S}_{t_1,t_0}}$$

13.3.5 Multivariate EWMA control chart

The EWMA control chart, first introduced by Roberts (1959), was adapted to multidimensional observations by Lowry, Woodall, Champ and Rigdon (1992). In an ARL comparison the authors showed that the properties of the multivariate EWMA chart are similar to or even better than those of the multivariate CUSUM charts of Crosier (1988) and Pignatiello and Runger (1990). Additionally, the design of the multivariate EWMA chart is much simpler than that of the multivariate CUSUM charts. Prabhu and Runger (1997) gave recommendations on the choice of the EWMA parameter.

We define the MEWMA recursion for the vector $\eta_{(t)}$ by:

$$\mathbf{Z}_t = \mathbf{R}\boldsymbol{\eta}_{(t)} + (\mathbf{I} - \mathbf{R})\mathbf{Z}_{t-1} \quad t \geq 1$$

where $\mathbf{R} = diag(r_1, r_2, ..., r_{q+q(q+1)/2})$, $0 < r_j < 1, j = 1, ..., q + q(q+1)/2$. Rewriting gives:

$$\mathbf{Z}_t = (\mathbf{I} - \mathbf{R})^t\mathbf{Z}_0 + \mathbf{R}\sum_{j=0}^{t-1}(\mathbf{I} - \mathbf{R})^t\eta_{(t-j)}$$

$$= (\mathbf{I} - \mathbf{R})^t(\mathbf{Z}_0 - \mu_\eta) + \mathbf{R}\sum_{j=0}^{t-1}(\mathbf{I} - \mathbf{R})^t(\eta_{(t-j)} - \mu_\eta) + \mu_\eta$$

since $(\mathbf{I} - \mathbf{R})^t + \mathbf{R}\sum_{j=0}^{t-1}(\mathbf{I} - \mathbf{R})^t = \mathbf{I}$. Hence, in the in-control scenario the mean of the vector $\mathbf{Z}_t$ is $E_0(\mathbf{Z}_t) = (\mathbf{I} - \mathbf{R})^t(\mathbf{Z}_0 - \mu_\eta) + \mu_\eta$. In the following it is always assumed that the process $\mathbf{Z}_t$ starts in the target value μ_η, $\mathbf{Z}_0 = \mu_\eta$.

The MEWMA chart gives an alarm when

$$Q_t = (\mathbf{Z}_t - \mu_\eta)' \Sigma_{Z_t}^{-1} (\mathbf{Z}_t - \mu_\eta) > h_4 \tag{13.14}$$

where Σ_{Z_t} is the covariance matrix of $\mathbf{Z}_t$, which is calculated by

$$\mathbf{\Sigma}_{Z_t} = \sum_{j=1}^{t} Var(\mathbf{R}(\mathbf{I}-\mathbf{R})^{t-j}\mathbf{X}_j) = \sum_{j=1}^{t} \mathbf{R}(\mathbf{I}-\mathbf{R})^{t-j}\Sigma_\eta(\mathbf{I}-\mathbf{R})^{t-j}\mathbf{R} \tag{13.15}$$

In the case $r_1 = r_2 = \ldots = r_{q+q(q+1)/2} = r$ the formula (13.15) simplifies to

$$\mathbf{\Sigma}_{Z_t} = \Sigma_\eta \frac{r(1-(1-r)^{2t})}{2-r}$$

In equation (13.14) the control limit h_4 is defined such that the in-control ARL is equal to a fixed quantity ξ. In practice this has to be done by simulations.

It is possible to monitor changes in the covariance matrix of the asset returns based on the asymptotic MEWMA control chart, namely, MEWMAas. In this case the Mahalanobis distance in the equation (13.14) is taken due to the asymptotic covariance matrix $\mathbf{\Sigma}_{Z_t;asymp} = \Sigma_\eta r/(2-r)$ instead of the exact one $\mathbf{\Sigma}_{Z_t}$.

13.4 SIMULTANEOUS STATISTICAL SURVEILLANCE

In this section we use the same notation as in section 13.3. However, instead of calculating the vector $\hat{\mathbf{v}}_{(i)}$ for the whole vector of portfolio weights $\hat{\mathbf{w}}_{(i)}$ we calculate the sequence of $\{\hat{v}_{(i)}^{(j)}\}$, $j \in 1,\ldots,p$, for each component $\hat{w}_{(i)}^{(j)}$ correspondingly. Then the two-dimensional control procedures for detecting shifts in the mean and variance of $\hat{v}_{(i)}^{(j)}$ are constructed simultaneously. We consider the sequence of

$$\eta_{(i)}^{(j)} = (\hat{v}_{(i)}^{(j)}, \hat{v}_{(i)}^{(j)} \hat{v}_{(i)}^{(j)})$$

with the in-control mean and covariance matrix given by

$$\mu_{\eta^{(j)}} = E(\eta_{(i)}^{(j)}) = \left(0, \frac{\tilde{n}-p}{\tilde{n}-p-2}\right), \quad j \in \{1,\ldots,p\} \tag{13.16}$$

and

$$\Sigma\eta(j) = \begin{pmatrix} \frac{\tilde{n}-p}{\tilde{n}-p-2} & 0 \\ 0 & \frac{3(\tilde{n}-p)^2}{(\tilde{n}-p-2)(\tilde{n}-p-4)} - \left(\frac{\tilde{n}-p}{\tilde{n}-p-2}\right)^2 \end{pmatrix} \tag{13.17}$$

The testing problem for the simultaneous control charts is presented by

$$H_{0,t} : E(\eta^{(j)}_{(t)}) = \mu_{\eta^{(j)}} \quad \text{for each} \quad j$$

against

$$H_{1,t} : E(\eta^{(j)}_{(t)}) = \mu^{(j)}_1 \neq \mu_{\eta^{(j)}} \quad \text{for some} \quad j.$$

When the simultaneous T^2 control chart is applied the null hypothesis is rejected as soon as

$$\max_{j=1,\ldots,p} \{T^{(j)\,2}_t\} > h_5, \quad \text{with} \quad T^{(j)\,2}_t = (\eta^{(j)}_{(t)} - \mu_{\eta^{(j)}})' \Sigma^{-1}_{\eta^{(j)}} (\eta^{(j)}_{(t)} - \mu_{\eta^{(j)}})$$

h_5 determined from the condition that the in-control ARL is equal to a preselected value ξ.

The simultaneous MC1 scheme is defined similar to the multivariate MC1 control chart. Let $\mathbf{S}^{(j)}_{m,l} = \sum_{i=m+1}^{l} (\eta^{(j)}_{(i)} - \mu_{\eta^{(j)}})$ for $l, m \geq 0$, and furthermore, let

$$MC1^{(j)}_t = \max\,(\|\mathbf{S}^{(j)}_{t-n_t,t}\|_{\Sigma_{\eta^{(j)}}} - kn^{(j)}_t, 0), \quad t \geq 1$$

$n^{(j)}_t$ is calculated by analogy to (13.11). The MC1 control scheme gives a signal as soon as the statistic

$$simMC1_t = \max_{j=1,\ldots,p} MC1^{(j)}_t > h_6$$

exceeds a preselected control limit h_6.

For the simultaneous MCUSUM we consider $C^{(j)}_t = ||\mathbf{S}^{(j)}_{t-1} + (\eta^{(j)}_{(t)} - \mu_{\eta^{(j)}})||_{\Sigma_{\eta^{(j)}}}$ with $\mathbf{S}^{(j)}_{t-1}$ as in (13.12). The control statistic is given by

$$simMCUSUM_t = \max_{j=1,\ldots,p} MCUSUM^{(j)}_t \tag{13.18}$$

where

$$MCUSUM^{(j)}_t = (\mathbf{S}^{(j)'}_t \Sigma^{-1}_{\eta^{(j)}} \mathbf{S}^{(j)}_t)^{\frac{1}{2}} = \max\{0, C^{(j)}_t - k\}$$

The simultaneous PPCUSUM control statistic is defined as

$$simPCUSUM_t = \max_{j=1,\ldots,p} PPCUSUM^{(j)}_t$$

with

$$PPCUSUM^{(j)}_t = \max\{0, \|\mathbf{S}^{(j)}_{t-1,t}\|_{\Sigma_{\eta^{(j)}}} - k, \ldots, \|\mathbf{S}^{(j)}_{0,t}\|_{\Sigma_{\eta^{(j)}}} - tk\}$$

A large value of $simPPCUSUM_t$ is a hint that a change in the covariance matrix of the asset returns has occurred. Furthermore, this change leads to the reconstruction of the GMVP. The control limit is determined as described above. It is obtained within an extensive Monte Carlo study.

The simultaneous MEWMA control statistic is given by

$$simMEWMA_t = \max_{j=1,\ldots,p} MEWMA_t^{(j)}$$

where $Q_t^{(j)}$ is defined as in section 13.5. Application of the asymptotic covariance matrix in constructing the quadratic forms $Q_t^{(j)}$ leads to the asymptotic analog of the MEWMA control scheme which we denote by simMEWMAas. As usual for both control designs the control limits are obtained within an extensive Monte Carlo study.

13.5 A COMPARISON OF THE MULTIVARIATE AND SIMULTANEOUS CONTROL CHARTS

In this section we compare the multivariate control charts and simultaneous control schemes.

13.5.1 Structure of the Monte Carlo Study

Without loss of generality, in this section the in-control process is taken to be a four-dimensional Gaussian process $\{\mathbf{X}_t\}$ with zero mean vector and the covariance matrix as:

$$\Sigma = \begin{pmatrix} 0.84813 & 0.3726 & 0.18718 & 0.15418 \\ 0.3726 & 1.52624 & 0.31376 & 0.35488 \\ 0.18718 & 0.31376 & 1.83864 & 0.28748 \\ 0.15418 & 0.35488 & 0.28748 & 2.06115 \end{pmatrix}$$

To calculate Σ we made use of monthly data from Morgan Stanley Capital International for equity markets returns of four developed countries (the USA, the UK, Japan and Germany). This choice is not restrictive because in the in-control state the proposed statistics $\hat{\mathbf{v}}$ has the same distribution independent of the constant matrix $\mathbf{L}$. As a result, the calculated control limits can be used for the non-singular matrix Σ. Note, that in case where the number of elements in each subsample $\{\mathbf{X}_{(i);j}\}_{j=1}^{\tilde{n}}$ is large enough, the distribution of the vector $\hat{\mathbf{v}}_{(i)}$ is very accurately approximated by the standard normal distribution. Thus, we can use the control limits that are calculated for detecting changes in the mean vector and the covariance matrix of the standard normally distributed random vector.

In our simulation study we set $\tilde{n} = 20$. This choice corresponds to estimation of the covariance matrix of the four weeks (roughly monthly) returns by the daily returns (Schwert, 1989). It follows from Theorem 1 that for each i, the random vectors $\hat{\mathbf{v}}_{(i)}$ have a multivariate standard t-distribution with 16 degrees of freedom independently distributed as their construction is based on the non-overlapping samples.

In order to obtain the performance of the proposed sequential procedures the out-of-control situation has to be determined. In our simulation study the changes are generated by the following model:

$$\mathbf{X}_t \sim N_p(\mathbf{0}, \Sigma), \quad t \leq 0$$

$$\mathbf{X}_t \sim N_p(\mathbf{0}, \boldsymbol{\Delta}\Sigma\Delta), \quad t \geq 1$$

where

$$\Delta = \begin{pmatrix} 1+a_1 & a_2 & a_2 & a_2 \\ a_2 & 1+a_1 & a_2 & a_2 \\ a_2 & a_2 & 1+a_1 & a_2 \\ a_2 & a_2 & a_2 & 1 \end{pmatrix}$$

As a measure of the performance of a control chart the average run length (ARL) is applied. All multivariate and simultaneous control schemes are calibrated to have the same in-control ARLs, namely 60. Because no explicit formula for the in-control and the out-of-control ARLs are available, a Monte Carlo study is used to estimate these quantities. We estimate the in-control ARLs based on 10^5 simulated independent realizations of the process. The control limits of all charts are determined by applying the Regula falsi to the estimated ARLs. In Table 13.1 the control limits of the multivariate charts are given for various values of the reference value k and the smoothing parameter r, while Table 13.2 contains the control limits for the simultaneous schemes. Because the vectors $\hat{\mathbf{v}}_{(i)}$ are independent and identically standard t-distributed the control limits of the charts do not depend on the covariance matrix of asset returns. In the out-of-control state they are again independent but no longer identically distributed (see Theorem 2). Consequently they are not directionally invariant. For the MEWMA charts the control limits increase as the parameter r increases. Conversely, for the CUSUM schemes the control limits decrease as k increases. The control limit for the multivariate T^2 control scheme is 356.8, while for the simultaneous T^2 it is equal to 37.42. Finally, almost in all cases the control limits of the simultaneous control charts are much smaller than the corresponding limits of the multivariate schemes.

In order to study the out-of-control behavior of the proposed control charts we take various reference values k into account. For the multivariate CUSUM charts k is chosen as an element of the set $\{1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.3\}$, while for the simultaneous schemes from $\{1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0\}$. For the MEWMA and asymptotic MEWMA charts the smoothing matrix is taken as a diagonal matrix with equal diagonal elements $r \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$.

Table 13.1 Control limits of the multivariate MEWMA, MEWMAas, MC1, MCUSUM, T^2 charts (section 13.3, in-control, ARL = 60)

Type\k = 1.3+,r=	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
MEWMA	213.41	235.84	263.56	494.55	313.70	330.83	344.69	358.55	358.55	
MEWMAas	199.55	229.29	261.53	289.25	313.70	330.83	344.40	358.26	358.55	
MC1	33.76	31.24	29.16	27.36	25.89	24.50	23.69	22.91	22.16	21.60
MCUSUM	39.62	36.45	33.72	31.47	29.62	28.26	26.74	25.85	24.79	23.97

Table 13.2 Control limits of the simultaneous MEWMA, MEWMAas, MC1, MCUSUM, T^2 charts Section 4 (in-control ARL = 60)

Type\k = 1.3+,r=	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
simMEWMA	12.02	17.32	21.65	25.88	29.21	32.23	34.41	36.07	37.10	
simMEWMAas	10.75	16.45	21.33	25.49	29.02	31.91	34.41	36.07	37.10	
simMC1	5.41	5.24	5.09	4.95	4.82	4.72	4.56	4.48	4.33	4.22
simMCUSUM	5.60	5.41	5.22	5.07	4.93	4.80	4.66	4.53	4.40	4.28

13.5.2 Behavior in the out-of-control state

For the determination of the out-of-control ARLs we made use of 10^6 independent realizations of the underlying process. In Tables 13.3 and 13.4 the out-of-control ARLs of all control charts within our study are presented. The corresponding values r and k at which the smallest out-of-control ARLs are attained, are given in brackets. These values should be taken to detect the specific shift in the mean vector of the process $\{\eta_{(t)}\}$. For a given shift the ARL of the best chart is printed boldfaced.

Table 13.3 presents the results of the multivariate charts. For the given choice of the covariance matrix the best result is attained by the MEWMA control charts. In almost all cases it provides the smallest out-of-control ARL. On the second to fourth places the multivariate asymptotic MEWMA, MC1 and MCUSUM schemes can be ranked. They are clearly worse than the nonasymptotic MEWMA approach. The multivariate MEWMAas overperforms for the small values of the out-of-control ARL while for the moderate and large values the MC1 and MCUSUM designs show a better performance. The multivariate T^2 scheme behaves considerably worse. For nearly all shifts under consideration it has a larger out-of-control ARL than the other charts.

The results of the out-of-control ARLs for the simultaneous charts can be found in Table 13.4. In all cases the best performance is obtained by the simultaneous MEWMA control scheme which overperforms the rest of the control schemes in all cases. It possesses much smaller out-of-control ARLs

Table 13.3 Minimal out-of-control ARL of the multivariate MEWMA, MEWMAas, MC1, MCUSUM, T^2 charts for different values of parameters k and r

$a_1 \backslash a_2$	−0.9	−0.6	−0.3	0.0	0.3	0.6	0.9
−0.9	**1.30** (0.1)	**1.43** (0.1)	**1.55** (0.1)	**1.61** (0.1)	**1.54** (0.1)	**1.39** (0.1)	**1.20** (0.1)
	1.42 (0.8)	1.64 (0.6)	1.86 (0.6)	1.97 (0.4)	1.92 (0.4)	1.69 (0.6)	1.39 (0.7)
	1.48 (2.3)	1.70 (2.3)	1.90 (2.3)	2.01 (2.3)	1.95 (2.3)	1.76 (2.3)	1.49 (2.3)
	1.54 (2.3)	1.78 (2.3)	2.00 (2.3)	2.12 (2.3)	2.07 (2.3)	1.87 (2.3)	1.57 (2.3)
	1.43	1.67	1.95	2.10	2.04	1.76	1.41
−0.6	**1.18** (0.1)	**1.23** (0.1)	**1.23** (0.1)	**1.13** (0.1)	**1.03** (0.1)	**1.00** (0.1)	**1.00** (0.1)
	1.25 (0.9)	1.35 (0.7)	1.38 (0.7)	1.24 (0.8)	1.06 (0.9)	**1.00** (0.9)	**1.00** (0.1)
	1.30 (2.3)	1.42 (2.3)	1.45 (2.3)	1.33 (2.3)	1.11 (2.3)	**1.00** (2.3)	**1.00** (1.5)
	1.33 (2.3)	1.47 (2.3)	1.52 (2.3)	1.39 (2.3)	1.14 (2.3)	1.01 (2.3)	**1.00** (1.4)
	1.25	1.36	1.39	1.25	1.06	**1.00**	**1.00**
−0.3	**1.08** (0.1)	**1.02** (0.1)	**1.00** (0.2)	**1.00** (0.1)	**1.02** (0.1)	**1.17** (0.1)	**1.30** (0.1)
	1.10 (0.9)	1.03 (0.8)	**1.00** (0.3)	**1.00** (0.1)	1.04 (0.9)	1.34 (0.7)	1.56 (0.6)
	1.12 (2.3)	1.05 (2.3)	**1.00** (1.8)	**1.00** (1.4)	1.08 (2.3)	1.44 (2.3)	1.65 (2.3)
	1.13 (2.3)	1.06 (2.3)	**1.00** (3.1)	**1.00** (1.5)	1.11 (2.3)	1.51 (2.3)	1.74 (2.3)
	1.10	1.03	**1.00**	**1.00**	1.04	1.35	1.60
0.0	**1.20** (0.1)	**2.61** (0.1)	**9.85** (0.1)	MEWMA	80.11 (0.1)	21.92 (0.1)	**8.03** (0.1)
	1.23 (0.9)	2.95 (0.4)	11.08 (0.2)	MEWMAas	76.02 (0.1)	22.75 (0.1)	9.74 (0.1)
	1.26 (2.3)	2.97 (2.3)	10.21 (2.2)	MC1	65.94 (1.4)	**19.22** (1.7)	8.20 (2.1)
	1.27 (2.3)	3.08 (2.3)	10.49 (2.3)	MCUSUM	**65.38** (1.4)	19.91 (1.8)	8.55 (2.3)
	1.23	3.06	13.10	T^2	92.38	53.37	23.18
0.3	**1.48** (0.1)	**1.12** (0.1)	**2.27** (0.1)	14.10 (0.1)	186.57 (0.9)	76.42 (0.1)	11.14 (0.1)
	1.57 (0.7)	1.14 (0.8)	2.65 (0.4)	15.89 (0.1)	188.19 (0.9)	70.17 (0.2)	12.67 (0.1)
	1.64 (2.3)	1.16 (2.3)	2.64 (2.3)	**13.99** (1.9)	300.51 (2.3)	44.28 (1.4)	**10.64** (1.9)
	1.67 (2.3)	1.17 (2.3)	2.77 (2.3)	14.63 (2.0)	299.16 (2.3)	**44.05** (1.4)	11.12 (2.2)
	1.58	1.15	2.79	22.76	**183.68**	142.91	39.71
0.6	**1.00** (0.1)	**3.26** (0.1)	**1.21** (0.1)	**8.22** (0.1)	248.34 (0.9)	14.75 (0.1)	**4.33** (0.1)
	1.00 (0.7)	3.80 (0.4)	1.25 (0.9)	9.72 (0.2)	247.14 (0.9)	15.97 (0.1)	5.59 (0.2)
	1.00 (2.3)	3.81 (2.3)	1.28 (2.2)	8.59 (2.2)	422.18 (2.3)	**13.19** (1.9)	4.88 (2.3)
	1.00 (2.3)	3.88 (2.3)	1.31 (2.3)	8.86 (2.3)	**181.66** (2.3)	13.61 (2.0)	5.15 (2.3)
	1.00	3.93	1.25	14.03	242.34	58.32	12.62
0.9	**1.00** (0.1)	**1.00** (0.3)	**1.00** (0.1)	**6.44** (0.1)	**3.83** (0.1)	**1.51** (0.1)	**1.17** (0.1)
	1.01 (0.9)	**1.00** (0.5)	**1.00** (0.1)	7.71 (0.2)	4.99 (0.2)	1.95 (0.4)	1.37 (0.6)
	1.03 (2.3)	**1.00** (2.3)	**1.00** (1.4)	6.90 (2.2)	4.41 (2.3)	1.98 (2.3)	1.47 (2.3)
	1.04 (2.3)	**1.00** (2.3)	**1.00** (1.4)	7.21 (2.3)	4.66 (2.3)	2.12 (2.3)	1.56 (2.3)
	1.01	**1.00**	**1.00**	11.00	10.73	2.20	1.39

Table 13.4 Minimal out-of-control ARL of the simultaneous MEWMA, MEWMAas, MC1, MCUSUM, T^2 charts for different values of parameters k and r

$a_1 \backslash a_2$	−0.9	−0.6	−0.3	0.0	0.3	0.6	0.9
−0.9	**1.75** (0.1)	**1.63** (0.1)	**1.49** (0.1)	**1.33** (0.1)	**1.19** (0.1)	**1.09** (0.1)	**1.03** (0.1)
	2.70 (0.2)	2.46 (0.2)	2.16 (0.2)	1.85 (0.3)	1.56 (0.4)	1.31 (0.6)	1.13 (0.7)
	2.77 (1.1)	2.51 (1.1)	2.18 (1.1)	1.86 (1.1)	1.56 (1.2)	1.31 (1.9)	1.13 (2.0)
	2.82 (1.1)	2.55 (1.1)	2.21 (1.3)	1.88 (1.3)	1.57 (1.5)	1.32 (2.0)	1.13 (1.9)
	3.97	3.42	2.77	2.18	1.70	1.35	1.13
−0.6	**1.72** (0.1)	**1.46** (0.1)	**1.22** (0.1)	**1.06** (0.1)	**1.00** (0.1)	**1.00** (0.1)	**1.00** (0.1)
	2.63 (0.2)	2.15 (0.2)	1.64 (0.3)	1.24 (0.4)	1.03 (0.9)	**1.00** (0.7)	**1.00** (0.1)
	2.70 (1.1)	2.16 (1.2)	1.64 (1.2)	1.24 (2.0)	1.03 (2.0)	**1.00** (1.5)	**1.00** (1.1)
	2.76 (1.1)	2.20 (1.3)	1.66 (1.3)	1.25 (1.9)	1.03 (2.0)	**1.00** (1.7)	**1.00** (1.1)
	3.83	2.75	1.82	1.26	1.03	**1.00**	**1.00**
−0.3	**1.76** (0.1)	**1.10** (0.1)	**1.00** (0.1)	**1.00** (0.1)	**1.00** (0.1)	**1.02** (0.1)	**1.03** (0.1)
	2.67 (0.2)	1.34 (0.4)	**1.00** (0.7)	**1.00** (0.1)	1.02 (0.9)	1.11 (0.6)	1.15 (0.6)
	2.74 (1.1)	1.34 (1.9)	**1.00** (1.5)	**1.00** (1.1)	1.02 (2.0)	1.12 (1.7)	1.15 (2.0)
	2.80 (1.1)	1.34 (2.0)	**1.00** (1.7)	**1.00** (1.1)	1.02 (2.0)	1.12 (1.9)	1.15 (1.9)
	3.89	1.39	**1.00**	**1.00**	1.02	1.12	1.16
0.0	**1.00** (0.1)	**1.34** (0.1)	**6.21** (0.1)	simMEWMA	**9.52** (0.1)	**3.21** (0.1)	**1.84** (0.1)
	1.00 (0.3)	1.84 (0.3)	8.14 (0.1)	simMEWMAas	11.77 (0.1)	4.67 (0.1)	2.74 (0.2)
	1.00 (1.1)	1.84 (1.1)	11.83 (1.1)	simMC1	19.42 (1.1)	5.38 (1.1)	2.81 (1.1)
	1.00 (1.3)	1.87 (1.3)	11.94 (1.1)	simMCUSUM	19.53 (1.1)	5.44 (1.1)	2.85 (1.1)
	1.00	2.17	20.95	*sim*T^2	29.03	9.56	4.03
0.3	**1.00** (0.1)	**1.17** (0.1)	**2.58** (0.1)	**8.61** (0.1)	**9.97** (0.1)	**3.46** (0.1)	**1.84** (0.1)
	1.01 (0.8)	1.51 (0.4)	3.96 (0.1)	10.88 (0.1)	12.40 (0.1)	5.09 (0.1)	2.77 (0.2)
	1.01 (2.0)	1.51 (1.7)	4.34 (1.1)	17.31 (1.1)	20.53 (1.1)	6.03 (1.1)	2.84 (1.1)
	1.01 (2.0)	1.52 (1.8)	4.42 (1.1)	17.49 (1.1)	20.90 (1.1)	6.13 (1.1)	2.91 (1.1)
	1.01	1.64	7.38	26.38	28.44	10.53	4.12
0.6	**1.00** (0.1)	**1.00** (0.1)	**1.44** (0.1)	**5.08** (0.1)	**5.11** (0.1)	**1.95** (0.1)	**1.28** (0.1)
	1.00 (0.1)	1.01 (0.9)	2.10 (0.2)	6.96 (0.1)	7.21 (0.1)	3.00 (0.2)	1.78 (0.3)
	1.00 (1.1)	1.02 (2.0)	2.11 (1.1)	9.37 (1.1)	10.12 (1.1)	3.13 (1.1)	1.79 (1.2)
	1.00 (1.1)	1.02 (2.0)	2.15 (1.1)	9.46 (1.1)	10.29 (1.1)	3.19 (1.1)	1.81 (1.2)
	1.00	1.02	2.68	16.43	16.30	4.64	2.06
0.9	**1.00** (0.2)	**1.00** (0.1)	**1.00** (0.1)	**4.07** (0.1)	**1.23** (0.1)	**1.02** (0.1)	**1.00** (0.1)
	1.00 (0.7)	**1.00** (0.1)	**1.00** (0.1)	5.79 (0.1)	1.68 (0.3)	1.10 (0.6)	1.02 (0.8)
	1.00 (1.7)	**1.00** (1.1)	**1.00** (1.1)	7.19 (1.1)	1.68 (1.2)	1.11 (1.9)	1.02 (2.0)
	1.00 (1.3)	**1.00** (1.1)	**1.00** (1.1)	7.29 (1.1)	1.70 (1.4)	1.11 (1.8)	1.02 (2.0)
	1.00	**1.00**	**1.00**	12.85	1.88	1.11	1.02

than other competitors. In second places the asymptotic MEWMA chart. For moderate and large values of the out-of-control ARLs this scheme shows a much better performance than the simultaneous MC1 and MCUSUM approaches that are in third and fourth places, while for small values all three charts behave similarly. The worst results are gives by the simultaneous T^2 control chart.

The comparison between the multivariate and the simultaneous charts leads to interesting results. If the ARL of a multivariate scheme is compared with its simultaneous counterpart, then for almost all considered shifts the simultaneous approach has the smaller out-of-control ARL for the process under consideration. While the multivariate charts have some difficulties for some shift constellations, all the changes are detected by the simultaneous control schemes. In all of our simulations the MEWMA chart provides very good results. The simultaneous MEWMA scheme shows the best performance. A possible explanation of this fact is based on the results of Woodall and Mahmoud (2005) who showed that the MC1 approach can build up a large amount of inertia. For that reason we recommend applying the MEWMA control chart for detecting changes in the covariance matrix of asset returns that could influence the optimal portfolio weights. For simplicity we have taken the smoothing values of the MEWMA chart as all equal. This chart has much more flexibility, and improvements can be expected if different values are chosen. The best results for the EWMA charts are obtained in the case of $r = 0.1$, for example, the smallest of the considered smoothing values. This result is in line with the findings of Frisen (2003), who argued that the best performance of the EWMA scheme is attained for small values of the parameter r. For further discussion of optimality for the control procedures we refer the reader to Pollak (1985), Srivastava and Wu (1997), Yakir (1997) and Moustakides (2004).

13.6 CONCLUSION

One of the most important problems in portfolio management is monitoring the covariance structure of asset returns. While Ledoit and Wolf (2003, 2004) have discussed the influence of the estimation error on the estimator for the covariance matrix, we have focused on the question of monitoring an optimal portfolio using the distributional properties of the estimator for the covariance matrix. This problem has been presented widely in the literature lately, for example Jobson and Korkie (1989), Gibbons, Ross and Shanken (1989), Britten-Jones (1999), Bodnar and Schmid (2004) and others. However, none of these proposals deals with sequential procedures.

In this chapter we derive sequential multivariate and simultaneous procedures for detecting changes in the covariance matrix of asset returns that have an influence on the weights of the global minimum variance portfolio.

All these control charts are of the residual type. To construct them we have made use of the findings presented in Theorems 1 and 2. The multivariate and simultaneous control procedures are independent of the covariance matrix of asset returns, which constitutes a great advantage of our findings. No additional information, except the portfolio weights, is required for constructing control limits and monitoring the efficiency. Finally, our findings have financial and statistical significance even for the distribution of the portfolio asset returns that do not possess the second and higher moments.

The performance of the proposed procedures is obtained within an extensive Monte Carlo study. The best results are reached by the simultaneous MEWMA control chart, and in second place we can rank the multivariate MEWMA approach. For that reason we recommend applying either the multivariate or the simultaneous MEWMA schemes.

APPENDIX

We denote $\mathbf{K}' = (\mu_1, \ldots, \mu_q, \mathbf{1})$. Let $\mathbf{H} = (\mathbf{K}\Sigma^{-1}\mathbf{K}')^{-1} = \{\mathbf{H}_{ij}\}_{i,j=1,2}$ and $\hat{\mathbf{H}} = (\mathbf{K}\hat{\Sigma}^{-1}\mathbf{K}')^{-1} = \{\hat{\mathbf{H}}_{ij}\}_{i,j=1,2}$. Let $\hat{\mathbf{H}}^{(-)} = \hat{\mathbf{H}}^{-1} = \{\hat{\mathbf{H}}^{(-)}_{ij}\}_{i,j=1,2}$, where $\hat{\mathbf{H}}^{(-)}_{22} = \mathbf{1}'\hat{\Sigma}^{-1}\mathbf{1}$. Then $\hat{\mathbf{w}}_{M;p} = \hat{\mathbf{H}}^{(-)}_{12}/\hat{\mathbf{H}}^{(-)}_{22}$.

Proof of Theorem 1

From Corollary 3.2.2 of Muirhead (1982) it holds that $(n-1)\hat{\Sigma} \sim W_p(n-1, \Sigma)$. Thus, $(n-1)^{-1}\hat{\mathbf{H}}^{(-)} \sim W^{-1}_{q+1}(n-p+2q+2, \mathbf{K}\Sigma^{-1}\mathbf{K}')$. Let us denote $\hat{\mathbf{b}} = \hat{\mathbf{H}}^{(-)}_{11} - \hat{\mathbf{H}}^{(-)}_{21}\hat{\mathbf{H}}^{(-)}_{12}/\hat{\mathbf{H}}^{(-)}_{22}$ and $\mathbf{b} = \mathbf{H}^{(-)}_{11} - \mathbf{H}^{(-)}_{21}\mathbf{H}^{(-)}_{12}/\mathbf{H}^{(-)}_{22}$. Then, from Proposition 1 of Bodnar (2004), it follows that

$$\frac{(n-1)^{-1}\hat{\mathbf{H}}^{(-)}_{12}}{(n-1)^{-1}\hat{\mathbf{H}}^{(-)}_{22}} | (n-1)^{-1}\hat{\mathbf{b}} \sim N\left(\frac{\mathbf{H}^{(-)}_{12}}{\mathbf{H}^{(-)}_{22}}, \frac{(n-1)^{-1}\hat{\mathbf{b}}}{\mathbf{H}^{(-)}_{22}}\right)$$

Thus

$$\frac{(n-1)^{-1}\hat{\mathbf{H}}^{(-)}_{12}}{(n-1)^{-1}\hat{\mathbf{H}}^{(-)}_{22}} - \frac{\mathbf{H}^{(-)}_{12}}{\mathbf{H}^{(-)}_{22}} | (n-1)^{-1}\hat{\mathbf{b}} \sim N\left(\mathbf{0}, \frac{(n-1)^{-1}\hat{\mathbf{b}}}{\mathbf{H}^{(-)}_{22}}\right)$$

Hence, $\hat{\mathbf{H}}^{(-)}_{12}/\hat{\mathbf{H}}^{(-)}_{22}$ and $\hat{\mathbf{H}}^{(-)}_{22}$ are independently distributed and

$$\sqrt{\mathbf{H}^{(-)}_{22}}\sqrt{n-1}\,\hat{\mathbf{b}}^{-\frac{1}{2}}\left(\frac{\hat{\mathbf{H}}^{(-)}_{12}}{\hat{\mathbf{H}}^{(-)}_{22}} - \frac{\mathbf{H}^{(-)}_{12}}{\mathbf{H}^{(-)}_{22}}\right) | (n-1)^{-1}\hat{\mathbf{b}} \sim N(\mathbf{0}, \mathbf{I})$$

The righthand side of the last expression does not depend on $\hat{\mathbf{b}}$. Hence, $\sqrt{\mathbf{H}^{(-)}_{22}}\sqrt{n-1}\,\hat{\mathbf{b}}^{-\frac{1}{2}}$ $(\hat{\mathbf{H}}^{(-)}_{12}/\hat{\mathbf{H}}^{(-)}_{22} - \mathbf{H}^{(-)}_{12}/\mathbf{H}^{(-)}_{22}) \sim N(\mathbf{0}, \mathbf{I})$ and is independent on $\hat{\mathbf{H}}^{(-)}_{22}$. From the other side with Theorem 3.2.11 of Muirhead $(n-1)\mathbf{H}^{(-)}_{22}/\,\hat{\mathbf{H}}^{(-)}_{22} \sim \chi^2_{n-p}$. Combining the results and using the definition of the multivariate t-distribution the statement of the theorem follows.

Proof of Theorem 2

In this proof we make use of the same notation as in proving Theorem 1. Here the tilde means that the corresponding quantities are calculated for the process with a covariance matrix $\tilde{\Sigma}$ instead of Σ. Furthermore, from Proposition 1 of Bodnar (2004) it follows that $\hat{\tilde{b}}^{-1} \sim W_q^{-1}(n-p+q, \tilde{b}^{-1})$.

(a) Hence it follows that

$$E(\hat{\tilde{\mathbf{v}}}) = E(\sqrt{n-p}\sqrt{\hat{\tilde{\mathbf{H}}}_{22}^{(-)}}\hat{\tilde{b}}^{-\frac{1}{2}}(\hat{\tilde{\mathbf{w}}}_{M;q} - \mathbf{w}_{M;q}))$$

From the independency of $\hat{\tilde{\mathbf{H}}}_{22}^{(-)}$ and $\hat{\tilde{b}}$ (Proposition 1 of Bodnar (2004)) and independency of $\hat{\tilde{\mathbf{H}}}_{22}^{(-)}$ and $\hat{\tilde{\mathbf{w}}}_{M;q}$ (see the proof of Theorem 1) the last equation transforms to:

$$E(\hat{\tilde{\mathbf{v}}}) = E\left(\frac{\sqrt{n-p}\sqrt{\hat{\tilde{\mathbf{H}}}_{22}^{(-)}}}{\sqrt{n-1}\sqrt{\tilde{\mathbf{H}}_{22}^{(-)}}}\right) E(\sqrt{n-1}\sqrt{\tilde{\mathbf{H}}_{22}^{(-)}}\hat{\tilde{b}}^{-\frac{1}{2}}(\hat{\tilde{\mathbf{w}}}_{M;q} - \mathbf{w}_{M;q})) \tag{13.19}$$

Note that $\tilde{\mathbf{H}}_{22}^{(-)}/\hat{\tilde{\mathbf{H}}}_{22}^{(-)} \sim \chi^2_{n-p}$, and it holds that:

$$E\left(\frac{\sqrt{n-p}\sqrt{\hat{\tilde{\mathbf{H}}}_{22}^{(-)}}}{\sqrt{n-1}\sqrt{\tilde{\mathbf{H}}_{22}^{(-)}}}\right) = \frac{\sqrt{n-p}}{\sqrt{2}}\frac{\Gamma\left(\frac{n-p-1}{2}\right)}{\Gamma\left(\frac{n-p}{2}\right)}$$

Let consider the second term in the product (13.19). From the proof of Theorem 1 it follows that

$$\hat{\tilde{\mathbf{w}}}_{M;q}|(n-1)^{-1}\hat{\tilde{b}} \sim N\left(\tilde{\mathbf{w}}_{M;q}, \frac{(n-1)^{-1}\hat{\tilde{b}}}{\tilde{\mathbf{H}}_{22}^{(-)}}\right)$$

Thus

$$(\hat{\tilde{\mathbf{w}}}_{M;q} - \mathbf{w}_{M;q})|(n-1)^{-1}\hat{\tilde{b}} \sim N\left(\tilde{\mathbf{w}}_{M;q} - \mathbf{w}_{M;q}, \frac{(n-1)^{-1}\hat{\tilde{b}}}{\tilde{\mathbf{H}}_{22}^{(-)}}\right)$$

Hence

$$\begin{aligned}\sqrt{\tilde{\mathbf{H}}_{22}^{(-)}}\sqrt{n-1}\hat{\tilde{b}}^{-\frac{1}{2}}(\hat{\tilde{\mathbf{w}}}_{M;q} - \mathbf{w}_{M;q})|(n-1)^{-1}\hat{\tilde{b}} \\ \sim N(\sqrt{n-1}\sqrt{\tilde{\mathbf{H}}_{22}^{(-)}}\hat{\tilde{b}}^{-\frac{1}{2}}(\tilde{\mathbf{w}}_{M;q} - \mathbf{w}_{M;q}), \mathbf{I})\end{aligned} \tag{13.20}$$

As a result

$$E(\sqrt{n-1}\sqrt{\tilde{\mathbf{H}}_{22}^{(-)}}\hat{\tilde{b}}^{-\frac{1}{2}}(\hat{\tilde{\mathbf{w}}}_{M;q} - \mathbf{w}_{M;q})) = \sqrt{n-1}\sqrt{\tilde{\mathbf{H}}_{22}^{(-)}}\hat{\tilde{b}}^{-\frac{1}{2}}(\tilde{\mathbf{w}}_{M;q} - \mathbf{w}_{M;q})$$

where $\tilde{\mathbf{w}}_{M;q} = \tilde{\Sigma}^{-1}\mathbf{1}/\mathbf{1}'\tilde{\Sigma}^{-1}\mathbf{1}$. To calculate the unconditional density we make use of Theorem 3.2.14 of Muirhead (1982), for example the fact that $\hat{\tilde{b}}^{-\frac{1}{2}} = \tilde{b}^{-\frac{1}{2}}\mathbf{T}$, where $\mathbf{T} = (t_{ij})_{i=1,\ldots,q,j=1,\ldots,i}$ is a $q \times q$ lower triangular matrix with $t_{ii}^2 \sim \chi^2_{n-p+q-i+1} i = 1, \ldots, q$,

$t_{ij} \sim N(0,1)$ $i = 1, \ldots, q$, $j = 1, \ldots, i$, and t_{ii} and t_{ij} are mutually independently distributed. Hence, the expectation of the second term is equal to

$$E(\sqrt{n-1}\sqrt{\tilde{\mathbf{H}}_{22}^{(-)}}\hat{\tilde{b}}^{-\frac{1}{2}}(\hat{\tilde{\mathbf{w}}}_{M;q} - \mathbf{w}_{M;q})) = \sqrt{\tilde{\mathbf{H}}_{22}^{(-)}}\tilde{b}^{-\frac{1}{2}} \begin{pmatrix} E(t_{11}) & & \\ & \ddots & \\ & & E(t_{qq}) \end{pmatrix} (\tilde{\mathbf{w}}_{M;q} - \mathbf{w}_{M;q})$$

Denote $\mathbf{A} = \boldsymbol{diag}(a_{11}, \ldots, a_{qq})$ with $a_{ii} = \frac{B\left(\frac{n-p-q-i+1+1}{2}, \frac{n-p-1}{2}\right)}{B\left(\frac{n-p-q-i+1}{2}, \frac{n-p}{2}\right)}, i = \overline{1, q}$. The first part of the theorem is proved.

(b) Here we denote $\mathbf{c} = \sqrt{n-1}\sqrt{\tilde{\mathbf{H}}_{22}^{(-)}}\hat{\tilde{b}}^{-\frac{1}{2}}(\hat{\tilde{\mathbf{w}}}_{M;q} - \mathbf{w}_{M;q})$. Then it holds that:

$$E(\hat{\tilde{\mathbf{v}}}\hat{\tilde{\mathbf{v}}}') = E\left(\frac{(n-p)\hat{\tilde{\mathbf{H}}}_{22}^{(-)}}{(n-1)\tilde{\mathbf{H}}_{22}^{(-)}}\right) E(\mathbf{cc}') = \frac{(n-p)}{2} \frac{\Gamma\left(\frac{n-p-1}{2}\right)}{\Gamma\left(\frac{n-p}{2}\right)} (Var(\mathbf{c}) + E(\mathbf{c})E(\mathbf{c})')$$

To evaluate the covariance matrix of the random vector $\mathbf{c}$ we apply the following formula of conditional variance:

$$Var(\mathbf{c}) = E(Var(\mathbf{c}|(n-1)^{-1}\hat{\tilde{b}})) + Var(E(\mathbf{c}|(n-1)^{-1}\hat{\tilde{b}})) \tag{13.21}$$

From equation (26.20) it follows that $Var(\mathbf{c}|(n-1)^{-1}\hat{\tilde{b}}) = \mathbf{I}$ and, respectively,

$$E(Var(\mathbf{c}|\,(n-1)^{-1}\hat{\tilde{b}})) = \mathbf{I} \tag{13.22}$$

Let consider the second term in equation (13.21). It follows from the proof of part (a) that

$$E(\mathbf{c}|(n-1)^{-1}\hat{\tilde{b}}) = \sqrt{n-1}\sqrt{\tilde{\mathbf{H}}_{22}^{(-)}}\tilde{b}^{-\frac{1}{2}}\mathbf{T}(\tilde{\mathbf{W}}_{M;q} - \mathbf{w}_{M;q})$$

where $\mathbf{T}$ is a lower triangle random matrix.

Let us denote $\mathbf{W} = (\tilde{\mathbf{w}}_{M;q} - \mathbf{w}_{M;q})(\tilde{\mathbf{w}}_{M;q} - \mathbf{w}_{M;q})' = (w_{ij})_{i,j=\overline{1,\ldots,q}}$. Then

$$Var(E(\mathbf{c}|(n-1)^{-1}\hat{\tilde{b}})) = \tilde{\mathbf{H}}_{22}^{(-)}\tilde{b}^{-\frac{1}{2}}E(\mathbf{TWT}')\tilde{b}^{-\frac{1}{2}'}$$
$$-\tilde{\mathbf{H}}_{22}^{(-)}\tilde{b}^{-\frac{1}{2}}E(\mathbf{T})\mathbf{W}E(\mathbf{T}')\tilde{b}^{-\frac{1}{2}'} = \tilde{\mathbf{H}}_{22}^{(-)}\tilde{b}^{-\frac{1}{2}}(\mathbf{G} - \mathbf{F})\tilde{b}^{-\frac{1}{2}'}$$

where $\mathbf{G} = (g_{ij})_{i,j=\overline{1,\ldots,q}}$ and $\mathbf{F} = (f_{ij})_{i,j=\overline{1,\ldots,q}}$ with

$$g_{ij} = E\left(\sum_{l=1}^{i}\sum_{k=1}^{j} t_{il}\, w_{lk}\, t_{jk}\right) = w_{ij}\, E(t_{ii}\, t_{jj})$$
$$= \begin{cases} 2w_{ii}\dfrac{\Gamma\left(\frac{n-p+q-i+1}{2}+1\right)}{\Gamma\left(\frac{n-p+q-i+1}{2}\right)}, & \text{if } i = j \\[2ex] 2w_{ij}\dfrac{\Gamma\left(\frac{n-p+q-i+1+1}{2}\right)}{\Gamma\left(\frac{n-p+q-i+1}{2}\right)}\dfrac{\Gamma\left(\frac{n-p+q-j+1+1}{2}\right)}{\Gamma\left(\frac{n-p+q-j+1}{2}\right)} & \text{otherwise} \end{cases}.$$

Here we have applied Theorem 3.2.14 of Muirhead (1982), for example the fact that t_{ii} and t_{ij} are mutually independently distributed. Similar calculations lead to:

$$f_{ij} = \sum_{l=1}^{i} \sum_{k=1}^{j} E(t_{il}) w_{lk} E(t_{jk}) = w_{ij} E(t_{ii}) E(t_{jj})$$
$$= 2w_{ij} \frac{\Gamma\left(\frac{n-p+q-i+1+1}{2}\right)}{\Gamma\left(\frac{n-p+q-i+1}{2}\right)} \frac{\Gamma\left(\frac{n-p+q-j+1+1}{2}\right)}{\Gamma\left(\frac{n-p+q-j+1}{2}\right)}$$

Thus

$$E(\mathbf{cc}') = Var(\mathbf{c}) + E(\mathbf{c})E(\mathbf{c})' = \mathbf{I} + \tilde{\mathbf{H}}_{22}^{(-)} \tilde{b}^{-\frac{1}{2}} \mathbf{G} \tilde{b}^{-\frac{1}{2}'} \tag{13.23}$$

As a result

$$Var(\hat{\hat{\mathbf{v}}}) = E(\hat{\hat{\mathbf{v}}}\hat{\hat{\mathbf{v}}}') - E(\hat{\hat{\mathbf{v}}})E(\hat{\hat{\mathbf{v}}})'$$
$$= \frac{(n-p)}{2} \left(\frac{\Gamma\left(\frac{n-p}{2} - 1\right)}{\Gamma\left(\frac{n-p}{2}\right)} (\mathbf{I} + \tilde{\mathbf{H}}_{22}^{(-)} \tilde{b}^{-\frac{1}{2}} \mathbf{G} \tilde{b}^{-\frac{1}{2}'}) \right.$$
$$\left. + \left(\frac{\Gamma\left(\frac{n-p-1}{2}\right)}{\Gamma\left(\frac{n-p}{2}\right)} \right)^2 \tilde{\mathbf{H}}_{22}^{(-)} \tilde{b}^{-\frac{1}{2}} \mathbf{F} \tilde{b}^{-\frac{1}{2}'} \right)$$

The proof is complete.

REFERENCES

Alt, F.B. and Smith, N.D. (1988) "Multivariate Process Control", in P.R. Krishnaiah and C.R. Rao (eds), *Handbook of Statistics*, Vol. 7 (Amsterdam: North-Holland).

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Ebens, H. (2001) "The Distribution of Realized Exchange Rate Volatility", *Journal of the American Statistical Association*, 96(453): 42–55.

Andersson, E., Bock, B. and Frisen, M. (2003) "Detection of Turning Points in Business Cycles", *Journal of Business Cycle Measurement and Analysis*, 1: 93–108.

Andersson, E., Bock, B. and Frisen, M. (2005) "Statistical Surveillance of Cyclical Processes with Application to Turns in Business Cycles", *Journal of Forecasting*, 24(2): 465–90.

Barndorff-Nielsen, O.E. and Shephard, N. (2002) "Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models", *Journal of the Royal Statistical Society*, Ser. B, 64(2): 253–80.

Barndorff-Nielsen, O.E. and Shephard, N. (2004) "Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression and Correlation in Financial Economics", *Econometrica*, 72(3): 885–925.

Bodnar, O. (2005) "Control Charts for Multivariate Financial Time Series", PhD thesis, Frankfurt (Oder), Germany.

Bodnar, O. and Schmid, W. (2006) "CUSUM Control Schemes for Multivariate Time Series", in H.-J. Lenz and P.-Th. Wilrich (eds), *Frontiers of Statistical Process Control*, vol. 8 (Heidelberg: Physica).

Bodnar, T. (2004) "Optimal Portfolios in an Elliptical Model – Statistical Analysis and Tests for Efficiency", PhD thesis, Europa University Viadrina, Frankfurt (Oder), Germany.

Bodnar, T. and Schmid, W. (2004) "A Test for the Weights of the Global Minimum Variance Portfolio in an Elliptical Model", EUV Working Paper 2, Europa University Viadrina, Frankfurt (Oder), Germany.

Britten-Jones, M. (1999) "The Sampling Error in Estimates of Mean–Variance Efficient Portfolio Weights", *Journal of Finance*, 54(2): 655–71.

Crosier, R.B. (1986) "A New Two-Sided Cumulative Sum Quality Control Scheme", *Technometrics*, 28(3): 187–94.

Crosier, R.B. (1988) "Multivariate Generalizations of Cumulative Sum Quality Control Schemes", *Technometrics*, 30(3): 291–303.

Fang, K.T. and Zhang, Y.T. (1990) *Generalized Multivariate Analysis* (Berlin, Germany: Springer-Verlag and Beijing, China: Science Press).

Frisen, M. (2003) "Statistical Surveillance: Optimality and Methods", *International Statistical Review*, 71(2): 403–34.

Gibbons, M.R., Ross, S.A. and Shanken, J. (1989) "A Test of the Efficiency of a Given Portfolio", *Econometrica*, 57(5): 1121–52.

Healy, J.D. (1987) "A Note on Multivariate CUSUM Procedure", *Technometrics*, 29(4): 409–12.

Hotelling, H. (1947) "Multivariate Quality Control – Illustrated by the Air Testing of Sample Bombsights", in C. Eisenhart, M.W. Hastay and W.A. Wallis (eds), *Techniques of Statistical Analysis* (New York, NY: McGraw-Hill).

Jobson, J.D. and Korkie, B. (1989) "A Performance Interpretation of Multivariate Tests of Asset Set Intersection, Spanning, and Mean–Variance Efficiency", *Journal of Financial and Quantitative Analysis*, 24(2): 185–204.

Kan, R. and Zhou, G. (2001) "Tests of Mean–Variance Spanning", Working paper, Washington University.

Kim, D. and Kon, S. (1999) "Structural Change and Time Dependence in Models of Stock Returns", *Journal of Empirical Finance*, 6(3): 283–308.

Ledoit, O. and Wolf, M. (2003) "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection", *Journal of Empirical Finance*, 10(5): 603–21.

Ledoit, O. and Wolf, M. (2004) "A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices", *Journal of Multivariate Analysis*, 88(2): 365–411.

Lowry, C.A., Woodall, W.H., Champ, C.W. and Rigdon, S.E. (1992) "A Multivariate Exponentially Weighted Moving Average Control Charts", *Technometrics*, 34(1): 46–53.

Markowitz, H. (1952) "Portfolio Selection", *Journal of Finance*, 7(1): 77–91.

Moustakides, G.V. (2004) "Optimality of the CUSUM Procedure in Continuous Time", *Annals of Statistics*, 32(1): 302–15.

Muirhead, R.J. (1982) *Aspects of Multivariate Statistical Theory* (New York, NY: John Wiley & Sons).

Ngai, H.-M. and Zhang, J. (2001) "Multivariate Cumulative Sum Control Charts Based on Projection Pursuit", *Statistica Sinica*, 11(3): 747–66.

Pignatiello, Jr., J.J. and Runger, G.C. (1990) "Comparison of Multivariate CUSUM Charts", *Journal of Quality Technology*, 22(2): 173–86.

Pollak, M. (1985) "Optimal Detection of a Change in Distribution", *Annals of Statistics*, 13(1): 206–27.

Prahbu, S.S. and Runger, G.C. (1997) "Designing a Multivariate EWMA Control Chart", *Journal of Quality Technology*, 29(1): 8–15.

Roberts, S.W. (1959) "Control Charts Tests Based on Geometric Moving Averages", *Technometrics*, 1(3): 239–50.

Runger, G., Alt, F.B. and Montgomery, D. (1996) "Contributors to a Multivariate Statistical Process Control Chart Signal", *Communications in Statistics – Theory and Methods*, 25(10): 2203–13.

Schipper, S. and Schmid, W. (2001) "Sequential Methods for Detecting Changes in the Variance of Economic Time Series", *Sequential Analysis*, 20(4): 235–62.

Schmid, W. and Tzotchev, D. (2004) "Sequential Monitoring of the Parameters of a One Factor Cox–Ingersoll–Ross Model", *Sequential Analysis*, 23(1): 1–40.

Schwert, G.W. (1989) "Why Does Stock Market Volatility Change Over Time?", *Journal of Finance*, 44(5): 1115–53.

Sliwa, P., and Schmid, W. (2005) "Monitoring the Cross-Covariances of a Multivariate Time Series", *Metrika*, 61(1): 89–115.

Srivastava, M.S. and Wu, Y.H. (1997) "Evaluation of Optimum Weights and Average Run Lengths in EWMA Control Schemes", *Communications in Statistics – Theory and Methods*, 26(6): 1253–67.

Theodossiou, P.T. (1993) "Predicting Shifts in the Mean of a Multivariate Time Series Process: An Application in Predicting Business Failures", *Journal of the American Statistical Association*, 88(422): 441–9.

Tu, J. and Zhou, G. (2004) "Data-Generating Process Uncertainty: What Difference Does It Make in Portfolio Decisions?", *Journal of Financial Economics*, 72(2): 385–421.

Wald, A. (1947) *Sequential Analysis* (New York, NY: John Wiley & Sons).

Woodall, W.H. and Mahmoud, M.A. (2005) "The Inertial Properties of Quality Control Charts", *Technometrics*, 47(4): 425–36.

Yakir, B. (1997) "A Note on Optimal Detection of a Change in Distribution", *Annals of Statistics*, 25(5): 2117–26.

Yashchin, E., Steinand, D. and Philips, T. (1997) "Monitoring Active Portfolios Using Statistical Process Control", in H. Amman *et al*. (eds), *Computational Approaches to Economic Problems* (Norwell, MA: Kluwer).

CHAPTER 14

An Empirical Study of Time-Varying Return Correlations and the Efficient Set of Portfolios

Thadavillil Jithendranathan

14.1 INTRODUCTION

Modern portfolio theory was first introduced in 1952 (Markowitz, 1952), and since then it has been the mainstay of asset allocation models. In the mean-variance paradigm of Markowitz, an efficient set of portfolios is estimated by maximizing the expected return of the portfolio and minimizing its risk, as measured by the standard deviation. For practical purposes, efficient portfolio construction requires estimation of expected returns and variances of expected returns of individual assets in the portfolio, as well as the covariance matrix of the asset returns. The most widely used method of estimating these inputs into a portfolio model is to use the past return data for a period of five years and use the historic average values of returns, variances and co-variances as proxies for expected values. One of the implicit assumptions in this method of efficient portfolio construction is that the variances and co-variances are time-invariant during the holding period of the portfolio (Jobson and Korkie, 1981).

Despite its theoretical appeal, practitioners are generally cautious in applying mean-variance optimization models in practice. As pointed out by Michaud (1989), the optimization tends to give higher weights for securities with large expected returns, low variances and negative correlations with other securities in the portfolio. Other studies have indicated that the forecasted returns by the optimization model are highly sensitive to changes in the expected returns and co-variances (Best and Grauer, 1991; Chopra and Ziemba, 1993).

Several methods have been suggested to reduce the sensitivity of the errors of the mean-variance optimization model (Jorion, 1986; Jorion, 1991; Fletcher and Hiller, 2001). Most of these studies focus on reducing the sensitivity of the optimization model to the input parameters. In this chapter I try to improve the *ex post* returns of efficient portfolios by using time-varying variances and co-variances.

Several studies into the nature of variances and co-variances of asset returns have indicated that variances and co-variances do change over time (Goetzmann, Li and Rouwenhorst, 2005; Forbes and Rigbon, 2002). If the variances and co-variances are time varying, then the next question is which is the best method of estimating these. The most popular method is to use a moving average specification in which the correlations are estimated using a moving window of time. The drawback of this method is that it gives equal weight to all the observations during the time period used in the moving average calculations. The other method of estimating the time varying correlations is to use multivariate GARCH models. The first set of models of this genre is based on the Constant Correlation Coefficient model of Bollerslev (1990). But the assumption that the correlation coefficient was constant remained the main weakness of theses models. The second set of GARCH models are based on the multivariate GARCH models introduced by Kroner and Ng (1998). Even though these multivariate GARCH models are appealing from a theoretical standpoint, computationally they suffered from the problem of estimating too many coefficients at the same time. Engle (2002) introduced a new class of multivariate GARCH models called "Dynamic Conditional Correlation Models", which combined flexibility of the univariate models with the theoretical appeal of time-varying correlations. In this chapter I use this technique to estimate the time-varying correlations.

The main focus of this chapter is to test whether the efficient portfolios created with variances and co-variance estimates using the multivariate GARCH models will have superior *ex post* performance over the traditional approach of estimating the same using a moving or rolling window of time. Two sets of efficient portfolios are created – one using the rolling window of time and other using the multivariate GARCH models and *ex post* returns of these portfolios are calculated for periods of one, three and six months. The *ex post* returns of these two sets of efficient portfolios are then compared to see if there is any statistically significant difference between the two.

The rest of the chapter is organized as follows. Section 14.2 describes the empirical methodology and the sources and details of data. The results of the tests are detailed in section 14.3 and the results are discussed in section 14.4.

14.2 EMPIRICAL METHODOLOGY AND DATA

In portfolio optimization models, the objective is to maximize the return and minimize the risk of the portfolio. The expected return of a portfolio is the weighted average of the returns of individual securities in the portfolio and the weights are the proportion of each of the securities in the portfolio and can be expressed as follows:

$$\bar{R}_P = \sum_{i=1}^{N} X_i \bar{R}_i \tag{14.1}$$

where X_i is the weight of the ith security in the portfolio and $\overline{R}_i$ is the expected return of that asset.

The standard deviation of a portfolio can be expressed as:

$$\sigma_P^2 = \sum_{i=1}^{N} X_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{\substack{k=1 \\ k \neq i}}^{N} X_i X_k \sigma_{i,k} \tag{14.2}$$

where σ^2s are the variances and $\sigma_{i,k}$ is the covariance between the two securities i and k.

The standard method of optimization is to find a set of portfolios, which will give the maximum return for a given level of risk. This set of portfolios are called the efficient set of portfolios and based on their individual risk preferences investors can choose a specific portfolio from this set of optimal portfolios.

Mathematically the optimization problem can be stated as follows:

$$Min\ \sigma_P^2 = \sum_{i=1}^{N} X_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{\substack{k=1 \\ k \neq i}}^{N} X_i X_k \sigma_{i,k} \tag{14.3}$$

Subject to the following constraint:

$$\sum_{i=1}^{N} X_i = 1 \tag{14.4}$$

Portfolios can be created with or without short-selling constraints. In this chapter the portfolios are constructed with short selling constraints, which require the following additional constraint:

$$0 \leq X_i < 1 \tag{14.5}$$

In this chapter I use two different approaches to estimate the expected returns, variances and co-variances using the historic data. The first method is the commonly used rolling estimator, where the unconditional means, variances and co-variances are estimated using a rolling window of fixed N observations over a sample period T. The unconditional mean return and variance of a security i is estimated as:

$$\overline{R}_i = \frac{1}{N}\sum_{t=1}^{N} R_{it} \tag{14.6}$$

$$\sigma_i^2 = \frac{1}{N-1}\sum_{t=1}^{N}\left(R_{it} - \bar{R_i}\right)^2 \tag{14.7}$$

The co-variance between the returns of two securities i and k are estimated as follows:

$$\sigma_{i,k} = \frac{1}{N-1}\sum_{t=1}^{N}\left(R_{it} - \overline{R}_i\right)\left(R_{kt} - \overline{R}_k\right) \tag{14.8}$$

One of the main problems with such rolling estimators is that it does not capture the time-varying nature of means, variances and co-variances. To capture the time varying nature of variances and co-variances, the second method of estimation uses the Dynamic Conditional Correlation (DCC) model of Engle (2002). The conditional correlation between two random variable r_1 and r_2 that have mean zero can be written as:

$$\rho_{12,t} = \frac{E_{t-1}(r_{1,t}r_{2,t})}{\sqrt{E_{t-1}(r_{1,t}^2)E_{t-1}(r_{2,t}^2)}} \tag{14.9}$$

Let $h_{i,t} = E_{t-1}(r_{i,t}^2)$ *and* $r_{i,t} = \sqrt{h_{i,t}\varepsilon_{i,t}}$ *for* $i = 1, 2$, where $\varepsilon_{i,t}$ is a standardized disturbance that has zero mean and variance of one.

Substituting the above into equation (14.1) we get:

$$\rho_{12,t} = \frac{E_{t-1}(\varepsilon_{1,t}\varepsilon_{2,t})}{\sqrt{E_{t-1}(\varepsilon_{1,t}^2)E_{t-1}(\varepsilon_{2,t}^2)}} = E_{t-1}(\varepsilon_{1,t}\varepsilon_{2,t}) \tag{14.10}$$

Using a GARCH (1,1) specification, the covariance between the random variables can be written as:

$$q_{12,t} = \overline{\rho}_{12} + \alpha\left(\varepsilon_{1,t-1}\varepsilon_{2,t-1} - \overline{\rho}_{12}\right) + \beta\left(q_{12,t-1} - \overline{\rho}_{12}\right) \tag{14.11}$$

The unconditional expectation of the cross product is $\overline{\rho}_{12}$, while for the variances $\overline{\rho}_{12} = 1$

The correlation estimator is:

$$\rho_{12,t} = \frac{q_{12,t}}{\sqrt{q_{11,t}\, q_{22,t}}} \tag{14.12}$$

This model will be mean-reverting if $\alpha + \beta < 1$. The matrix version of this model can then be written as:

$$Q_t = S(1 - \alpha - \beta) + \alpha(\varepsilon_{t-1}\varepsilon'_{t-1}) + \beta Q_{t-1} \tag{14.13}$$

where S is the unconditional correlation matrix of the disturbance terms and $Q_t = |q_{1,2,t}|$.

The log-likelihood for this estimator can be written as:

$$L = -\frac{1}{2}\sum_{t=1}^{T}\left(n\log(2\pi) + 2\log|D_t| + \log|R_t| + \varepsilon'_t R_t^{-1}\varepsilon_t\right) \tag{14.14}$$

where $D_t = diag\left\{\sqrt{h_{i,t}}\right\}$ and R_t is the time-varying correlation matrix. With these estimates of variances and correlations, the covariance matrix is constructed.

The main purpose of this chapter is to study whether the use of time-varying variances and co-variances in portfolio optimization models will result in better *ex post* results as compared to the traditional rolling estimates. For this purpose portfolios are created using twenty stocks from Dow Jones Industrial Average Index. The time period covered is from January 1995 and December 2004. For consistency the stocks used in this study are the ones that were part of the Dow Jones Index for the entire period of study, except for Microsoft, which was included in the list in 1999. The weekly returns for each of these stocks are obtained from Bloomberg.

A window of five years is used in estimating the means, variances and co-variances using the rolling estimator. This window is moved by one month for the next five years, creating a total set of 60 separate estimates. These estimates are the inputs used in the portfolio optimization model. With each set of monthly inputs, a set of efficient portfolios is estimated. Each of these efficient portfolios contain one minimum variance portfolio and ten efficient portfolios with increasing levels of risk compared to the minimum variance portfolio.

For the DCC estimators I use the same set of five-year rolling windows, but to capture the time-varying nature of variances and co-variances, the end of the period values of the same is input into the portfolio optimization model. For example, using the DCC model one can estimate 260 variances and correlations for a period of five years. But for estimating the efficient set of portfolios, only the variances and correlations for the last week of the sample period is used. For example, for the time period from 3 January 2000 to 27 December 2004, the variances and correlations used are taken for the last week of the time period, which is 27 December 2004. In this way it is possible to capture the full extent of the time-varying nature of these variables as it existed at the time of construction of the portfolio. To be consistent with the time-varying nature of variances and correlations, the mean returns for the last four weeks of the sample is used as the expected mean for each of the stocks.

Using the above procedure, I am able to get the weights of the individual stocks in each of the efficient portfolio. Using these weights and the actual returns of each of the 20 stocks for periods of one-month, three-months and six-months from the date when the efficient portfolio is created, *ex post* returns of the efficient set of portfolios are calculated for each of the 60 months for which efficient sets are calculated. The performance of efficient portfolios computed using the rolling method and the DCC method are then compared using the following regression equation:

$$R_{j,t} = \alpha + \beta Dummy_{j,t} + \varepsilon_{j,t} \tag{14.15}$$

where $R_{j,t}$ is the pooled returns of all eleven efficient portfolios for a period of sixty months and $Dummy_{j,t}$ is a dummy variable, which is 1 if the portfolio is estimated using the DCC method and 0 if it is estimated using the rolling method. If the regression coefficient β is significant, then it indicates that there is difference in the *ex post* performance of the portfolios estimated using the two different methods. The value of this variable is also the difference between the *ex post* returns of portfolios estimated using the two different methods.

14.3 RESULTS

The descriptive statistics of weekly returns of the 20 stocks in this study are given in Table 14.1.

The time period in this study covers the tech bubble of the latter half of the 1990s as well as the dramatic events of 9/11 and the subsequent downturn in the stockmarkets. The average returns for all the 20 stocks are positive for the entire period, with Microsoft having the highest and General Motors the lowest weekly returns. There is also considerable variation in standard deviation of the returns of these 20 stocks, with a low of 0.030 for Exxon Mobil and a high of 0.053 for Coca-Cola. Fourteen out of the 20 stocks had negative skewness, which is an indication that during this period these stocks had more crashes than booms. Kurtosis measures the heaviness of tails, and ten of the stocks had a measure greater than three, which is an indication that these stock returns had fatter tails than that for a normal distribution. Finally, the Jarque-Bera test strongly rejects the normality assumption for the returns of all 20 stocks in the sample.

Table 14.1 Descriptive statistics of weekly returns from 9 January 1995 to 27 December 2004

Name	Mean	Std. dev.	Skewness	Kurtosis	Jarque-Bera
3M Co.	0.00269	0.035925	0.128619	2.899320	183.5649
Alcoa Inc.	0.00234	0.049083	0.092774	1.559698	53.4535
American Express	0.00370	0.046580	−0.689641	4.191052	421.7922
Boeing Co.	0.00178	0.050970	−1.294453	9.671405	2171.8343
Caterpillar Inc.	0.00281	0.046521	0.072897	1.022872	23.1296
E.I. du Pont	0.00165	0.041369	−0.275150	1.255070	40.6906
Exxon Mobil	0.00287	0.029678	−0.363687	2.217943	118.0474
General Electric	0.00320	0.039681	−0.321620	3.536954	280.0157
General Motors	0.00102	0.043620	−0.349797	3.129738	222.8350
Honeywell	0.00176	0.052903	−1.477493	10.179735	2434.4439
IBM	0.00332	0.048704	0.233023	2.151064	104.9592
JP Morgan	0.00231	0.052093	0.007047	1.218793	32.1892
McDonald's	0.00176	0.040196	−0.096122	1.434563	45.3901
Merck & Co.	0.00162	0.043450	−0.899514	4.107817	435.7310
Microsoft	0.00400	0.050139	−0.229235	1.463085	50.9342
SBC Comm.	0.00134	0.043267	0.175905	2.952786	191.5921
Coca-Cola	0.00239	0.053019	−0.746449	7.363402	1223.0493
P & G	0.00283	0.041425	−3.865948	41.841616	39227.5663
United Technology	0.00390	0.043373	−2.294469	21.525604	10495.5498
Walt Disney	0.00131	0.047635	−0.533730	4.322953	429.5935

The average standard deviations of the 20 stocks estimated using the two methods for the five-year period from January 2000 to December 2004 is plotted in Figure 14.1. With the rolling method, the standard deviations of individual stock returns are calculated using equation (14.7). The first observation in the plot is the average of standard deviations estimated for the time period from 9 January 1995 to 27 December 1997 using the weekly returns. From that point onwards, each observation is for a period that is moved ahead by one week. For example, the second observation is for the period from 16 January 1995 to 3 January 2000. In this way weekly observations are created using the rolling window of five years worth of weekly data.

The dynamic volatility estimates in the plot also covered the period from 3 January 2000 to 27 December 2000 and are computed using a GARCH(1,1) model. The difference between the two methods is that the rolling model uses past data to estimate the standard deviations and the GARCH models use only the data within the time period to estimate the standard deviations.

The plots indicate how the GARCH model is able to capture the changes in volatility associated with the 9/11 incident, whereas the rolling method tends to smooth out the volatility.

With 20 stocks in the portfolio, there are 180 covariance estimates for each time period. The average of these correlations calculated using the rolling model and the DCC model are plotted in Figure 14.2. The rolling

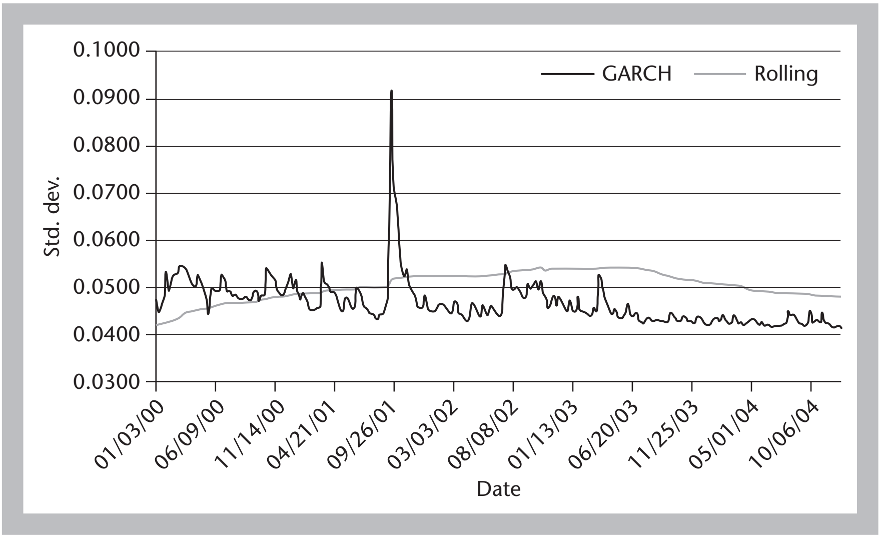


Figure 14.1 Average volatility-average weekly standard deviations from 3 January 2000 to 27 December 2004

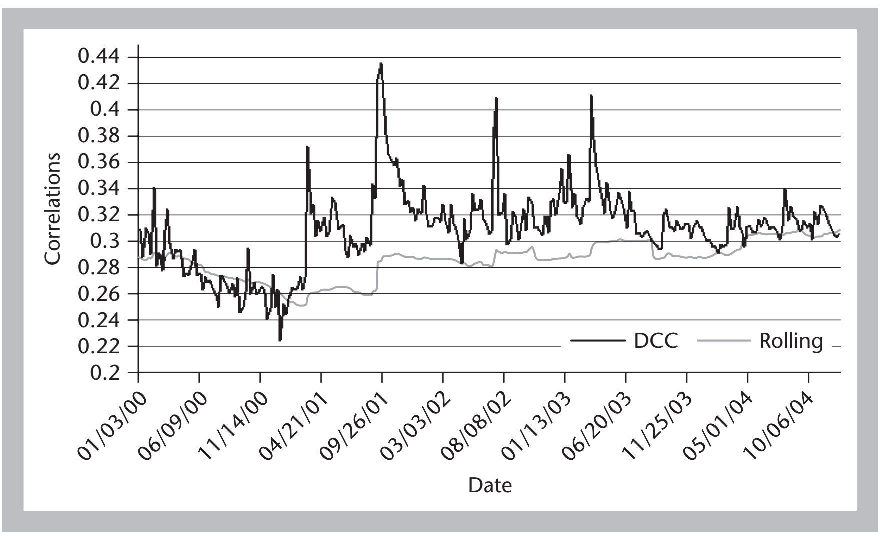


Figure 14.2 Average weekly correlations from 3 January 2000 to 27 December 2004

correlations are calculated using equation (14.8) and, in a similar fashion as the standard deviations, by using a rolling window of five years. The other set of correlations are estimated using the DCC method described in section 14.2.

As can be seen in this plot, the rolling correlation estimates tend to smooth out the short-term changes in correlations. There is a dramatic increase in the correlations after 9/11, which is much more pronounced with the DCC model than the rolling correlations model. The rolling correlations model and the DCC model are very similar at the beginning and towards the end of the period, where the fluctuations in the average correlations are relatively small. On the other hand for the period after 9/11, the average correlations estimated using DCC method had much higher variability than the rolling model.

The plots of efficient set of portfolios estimated using the DCC model is given in Figure 14.3. This plot contains 36 efficient portfolio sets for each of the months from January 2002 to December 2004. The purpose of this graph is to illustrate how the efficient frontier changes with the changes in expected returns, variances and co-variances.

It is clear from the plots of efficient frontiers that the frontier moved around quite a bit during the 30 months. In most of the instances, the standard deviations of the minimum variance portfolios are very close to each other, but the expected returns are substantially different in most of the cases. It may also be observed that in some instances entire sets of efficient portfolios have negative returns. This is due to the fact that I use the four-week average returns in my portfolio optimization model, and in a few instances all the stocks in the portfolio has negative returns.

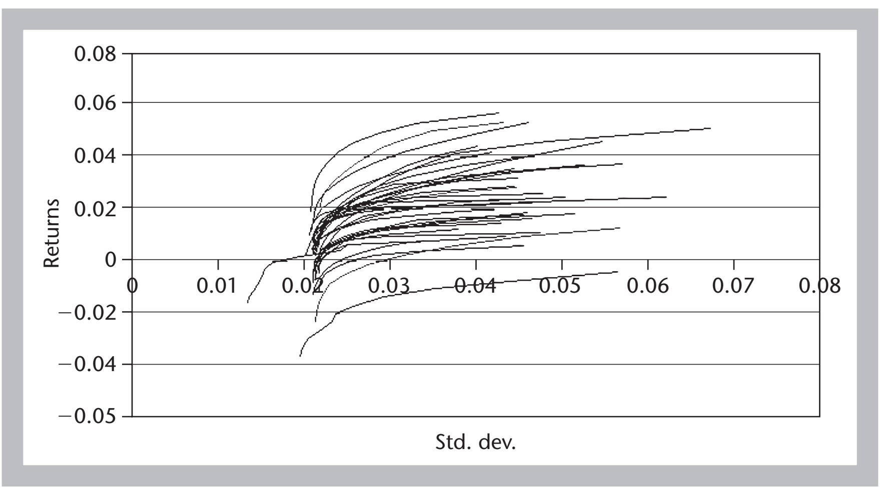


Figure 14.3 Efficient portfolios for the months January 2003 to December 2004

Table 14.2 Descriptive statistics of *ex post* returns of efficient portfolios

Model	Mean	Std. dev.	Skewness	Kurtosis	Jarque-Bera
			All portfolios		
			One-month returns		
Rolling	−0.0044	0.066813	−0.514897	2.3146	179.4399
DCC	0.0033	0.070526	0.012346	2.87996	233.5973
			Three-month returns		
Rolling	−0.0032	0.097966	−0.844897	0.92505	83.4231
DCC	0.0044	0.100200	0.649588	3.00677	299.9517
			Six-month returns		
Rolling	0.0048	0.127424	−0.753842	0.89383	85.8892
DCC	0.0077	0.131530	0.038403	2.02565	114.8855
			[Low-risk portfolios]		
			One-month returns		
Rolling	0.0001	0.058357	0.315829	1.28023	25.8994
DCC	0.0045	0.060822	0.351980	1.85805	50.1717
			Three-month returns		
Rolling	0.0057	0.083142	−0.328805	0.81550	13.9494
DCC	0.0037	0.089233	0.186751	2.63235	89.8324
			Six-month returns		
Rolling	0.0157	0.1084	−0.485141	0.60520	16.6189
DCC	0.0113	0.111627	−0.289139	0.07091	4.3136
			[High-risk portfolios]		
			One-month returns		
Rolling	−0.0081	0.072990	−0.537141	2.32654	100.1449
DCC	0.0023	0.077762	0.325195	2.85595	130.8367
			Three-month returns		
Rolling	−0.0106	0.108346	−0.784015	0.51280	41.5057
DCC	0.0049	0.108615	0.847932	2.88417	170.7150
			Six-month returns		
Rolling	−0.0122	0.140195	−0.745651	0.55011	38.5307
DCC	0.0048	0.146148	0.184092	2.23437	78.2019

The summary statistics of *ex post* returns of efficient portfolios created using the rolling model and DCC model are given in Table 14.2. For each of the 60 months, one minimum variance portfolio and ten efficient portfolios are created and the *ex post* returns of each of these portfolios are calculated for periods of one month, three months and six months. Furthermore, these

Table 14.3 OLS regression output for *ex post* returns against the DCC dummy

Period	α (t-stat)	β (t-stat)	Adj. R^2 (F-stat)	Obs.
		[All portfolios]		
One month	−0.0044	0.0077	0.0024	1340
	(1.6540)***	(2.0513)**	(4.2080)**	
Three months	−0.0032	0.0076	0.0007	1340
	(0.8353)	(1.4128)	(1.9961)	
Six months	0.0005	0.0072	0.0001	1340
	(0.0962)	(1.0264)	(1.0536)	
		[Low-risk portfolios]		
One month	0.0001	0.0045	0.0001	610
	(0.0197)	(0.9282)	(0.8616)	
Three months	0.0057	−0.0018	0.0001	610
	(1.1507)	(0.2605)	(0.0679)	
Six months	0.0157	−0.0044	0.0001	610
	(0.0962)	(1.0264)	(1.0536)	
		[High risk portfolios]		
One month	−0.0081	0.0104	0.0034	732
	(2.0543)**	(1.8604)***	(3.4612)***	
Three months	−0.0106	0.01553	0.0037	732
	(1.8682)***	(1.9365)***	(3.7498)***	
Six months	−0.0122	0.01696	0.0021	732
	(1.6261)***	(1.6320)***	(3.3672)***	

* Significant at 1%; ** Significant at 5%; *** Significant at 10%.

portfolios are divided into two groups based on the standard deviations of the efficient portfolios. For each month, the sample is divided into a set of low-risk portfolios comprising of five of the lowest variance portfolios, and another set of high-risk portfolios comprising of six portfolios with the highest risk.

From the results it is clear that except for two out of the nine cases, the *ex post* returns of efficient portfolios created using the DCC model are higher than those of the portfolios created using the rolling model. This point is further proved in the regression analysis that follows.

The results of the regressions using equation (14.15) are given in Table 14.3. Pooled *ex post* returns of efficient portfolios are regressed against the dummy variable that has a value of one for those portfolios for which the inputs were computed using the DCC model. Three sets of regressions are made, one for the total sample, one for the low-risk portfolios and one for the high-risk portfolios.

With the complete sample of all portfolios, the DCC model portfolios has significantly higher *ex post* returns than those of the rolling model. With the low-risk portfolios the DCC model do not have a statistically significant difference in returns as compared to the portfolios created using the rolling model. On the other hand, the DCC model is clearly superior to the rolling model for high-risk portfolios. In this case, the DCC model has returns statistically significant from that of the rolling model for all the three periods for which the comparison is made.

14.4 CONCLUSION

Even though the mean-variance optimization models have been around for more than 50 years, practical uses of these models have been limited for two reasons. Initially the model was not widely used due to the lack of widespread availability of the computational power required for both the estimation of variances and correlations, as well as the running of the optimization model itself. With the advent of faster computers, this problem has been considerably reduced in the past 20 years. The second and more serious limitation of the model is the way the inputs into the model are estimated. Until recently, computationally efficient multivariate models were not available for estimating the co-variances between asset returns. With the introduction of various multivariate GARCH models, this problem is somewhat mitigated. This chapter has used one such model for estimating the co-variances to see whether portfolios created using these inputs exhibit superior performance over those created with traditional estimates of co-variances. The results indicate that the use of time-varying variances and co-variances enhances the *ex post* performance of the efficient set of portfolios.

REFERENCES

Best, M.J. and Grauer, R.R. (1991) "On the Sensitivity of Mean–Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results", *Review of Financial Studies*, 4(2): 315–42.

Bollerslev, T. (1990) "Modeling the Coherence in Short Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model", *Review of Economics and Statistics*, 72(3): 498–505.

Chopra, V.K. and Ziemba, W.T. (1993) "The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice", *Journal of Portfolio Management*, 19(2): 6–11.

Engle, R. (2002) "Dynamic Conditional Correlation: A Simple Class of Multivariate GARCH Models", *Journal of Business and Economic Statistics*, 20(3): 339–50.

Fletcher, J. and Hiller, J. (2001) "An Examination of Resampled Portfolio Efficiency", *Financial Analysts Journal*, 57(5): 66–74.

Forbes, K.J. and Rigbon, R. (2002) "No Contagion, Only Interdependence: Measuring Stock Market Comovements", *Journal of Finance*, 57(5): 2223–61.

Goetzmann, W.N., Li, L. and Rouwenhorst, K.G. (2005) "Long-Term Global Market Correlations". *Journal of Business*, 78(1): 1–38.

Jobson, J.D. and Korkie, B.M. (1981). "Putting Markowitz Theory to Work", *Journal of Portfolio Management*, 7(4): 70–4.

Jorion, P. (1986) "Bayes–Stein Estimation for Portfolio Analysis", *Journal of Financial and Quantitative Analysis*, 21(3): 279–92.

Jorion, P. (1991) "Bayesian and CAPM Estimators of the Means: Implications for Portfolio Selection", *Journal of Banking and Finance*, 15(3): 717–27.

Kroner, K.F. and Ng, V.K. (1998) "Modeling Asymmetric Comovements of Assets Returns", *Review of Financial Studies*, 11(4): 817–44.

Markowitz, H.M. (1952) "Portfolio Selection", *Journal of Finance*, 7(1): 77–91.

Michaud, R.O. (1989) "The Markowitz Optimization Enigma: Is Optimized Optimal?", *Financial Analysts Journal*, 45(1): 31–42.

CHAPTER 15

The Derivation of the NPV Probability Distribution of Risky Investments with Autoregressive Cash Flows

Jean-Paul Paquin, Annick Lambert and Alain Charbonneau*

15.1 INTRODUCTION

Frederick Hillier's (1963) seminal paper was probably the first to propose the use of probabilistic information to assess risk in the process of capital budgeting. However, such an approach to investment decision was short-lived when Sharpe published his 1964 paper, supplemented by Lintner's (1965) and Mossin's (1966) articles, thus setting the conceptual ground for what was to become the modern capital asset pricing model (CAPM). Even Hertz's (1964) simulation methodology and Wagle's (1967) statistical analysis of risk in capital investment projects did not fare better. In fact, all the

* I would like to thank Dr F.-E. Racicot for his useful comments and encouragement.

probabilistic approaches to risky investment decisions were swept away by the Sharpe–Lintner–Mossin CAPM revolution as it became the creed of modern financial theory. Such a result was unavoidable given that, under the capital asset pricing theory, the dispersion (as well as the higher moments) in the probability distribution of future cash flows became an irrelevant statistic. Systematic risk, as calculated by the beta, became the only relevant measure of risk.

However, probabilistic information has crept back into the process of risky investment evaluation when various authors have drawn a close parallel between risky capital investments and financial options, thus giving rise to real option theory (Brennan and Schwartz, 1985; Copeland and Antikarov, 2001; Cox, Ross and Rubinstein, 1979; Dixit and Pindyck, 1994; Ingersoll, Jr. and Ross, 1992; Trigeorgis, 1993). However, notwithstanding the CAPM orthodoxy, how can one justify the reintroduction of total risk considerations in the investment decision when systematic risk should be the only relevant datum?

In pursuance of the Hillier probabilistic approach, this chapter deals with capital investment decisions for which total risk is relevant to the firm. In the first part we revisit and summarize some of the main research results that have lead some authors to come to the conclusion that under financial distress total risk matters to the firm. Such a conclusion is reached after drawing a clear distinction between the concept of systematic risk as proposed by academics under the Modigliani–Miller capital budgeting normative paradigm and the concept of total effective risk as implicitly used by managers under a positive probabilistic paradigm. Such a distinction will entail the use of two totally different measures of risk: (a) risk measured by the mean volatility of rates of return around a market index of central tendency, and (b) risk measured by the lower-tail of a NPV probability distribution. Section 15.2 discusses the systematic risk and the perfect economy, while section 15.3 deals with the total risk and the real economy.

Section 15.4 will address the question of the project NPV probability distribution; for one cannot estimate the lower-tail of a NPV probability distribution without specific knowledge as to its total probability distribution. In the second part, we derive important results concerning the investment project NPV distribution when the operating cash flows probability distributions are unknown and not independent in probability (for example, are not independent identically distributed (independent identically distributed) random variables, or are not normal independently distributed random variables). Specifically, we will deal with serially correlated discounted net cash flows. We will also demonstrate that, although first-order autocorrelated cash flows do not invalidate the Central Limit Theorem's asymptotic convergence properties to a Normal distribution, the introduction of any discount rate in the NPV equation does so, except for the very special case where net cash flows are normally distributed. Does this result imply that the probabilistic approach, as a general approach to capital investment, is at a dead-end? This is the question which will be addressed in part three of the article.

Section 15.5 will explore the extent to which positive discount rates can invalidate the applicability of the CLT to the derivation of the NPV probability distribution. Such an exploration is carried out in two steps. The first step involves computer simulations to generate NPV probability distributions with first-order and second-order autoregressive cash flows under three probability distributions (normal, uniform, and double-exponential). Section 15.6 consists in testing the statistical difference between the simulated NPV probability distributions and a normal probability distribution. The statistical test will involve a chi-square test and will aim at determining at the 1% level of significance the threshold discount rate over which the CLT is invalidated when applied to the NPV model. Section 15.7 concludes.

15.2 SYSTEMATIC RISK AND THE PERFECT ECONOMY

When the CAPM was introduced in the mid-1960s, it was rapidly adopted by the academic community because of its theoretical elegance, its conceptual contribution, and also because of its mathematical and statistical simplicity. However, many of the model's assumptions did not hold in reality. According to Milton Friedman (1953), this should not be considered as a fundamental flaw for any theory has to be judged, not by the realism of its assumptions, but by extent to which it provides meaningful explanations and valid predictions. An examination of the CAPM's logical consequences against observed reality becomes unavoidable.

A plethoric number of empirical studies have therefore attempted to validate the CAPM as an explanatory and predictive theory. "However, in general, the results offer only limited support of the CAPM" (Levy and Sarnat, 1994: 337). In complete contradiction with the CAPM's most fundamental precept, studies by Miller and Scholes (1972) and Levy (1978) showed that unsystematic risk (measured by the residual variance) turned out to be statistically significant in the determination of securities' required rates of return. Such a result could be attributed to the violation of one of the CAPM's most basic *sine qua non* assumption which states that investors hold diversified portfolios. However, as was shown by Blume, Crocket and Friend (1974), the typical investor holds but a very limited number of securities in his portfolio, less than four stocks on average. This might very well explain the greatest challenge to the CAPM when Fama and French (1992) found no systematic relation between return and risk as measured by beta. Given that investors do not hold diversified market portfolios, it was no surprise for Levy and Sarnat (1994: 339) also to conclude that systematic risk, as measured by beta, contributed significantly less than did total risk to the explanation of securities' required rates of return.

From a different perspective, theoretical considerations from the CAPM would also imply that companies in the same sector should face similar systematic risks and, therefore, should apply similar discount rates when assessing their investment projects. However, many surveys have found high variation in the hurdle rates used within same companies and same industry sectors; this would indicate that hurdle rates are not related directly to the cost of capital as prescribed by the CAPM. For instance, Poterba and Summers (1995) found from a sample of 228 companies (out of all Fortune 1000 companies) that only 12 percent of the hurdle rates total variance could be explained by the industry sector. A simple linear regression of the hurdle rates of firms on their respective beta coefficient revealed that the beta was not statistically significant in explaining variations in the hurdle rates. Their results also indicated that depending on the project, hurdle rates varied substantially within a same company, strategic projects having much lower hurdle rates. According to Jagannathan and Meier (2002), firms generally use much higher hurdle rates than the CAPM prescribed cost of capital. What is rationed is not financial capital, as usually assumed by a MM frictionless and transparent economy, but managerial talent and organisational capital. They demonstrate that when the firm has substantial real options, the project selection decision will be optimal as long as the hurdle rate is sufficiently high.

Finally, Roll's 1980 article casts serious doubts on the very empirical testability of the CAPM. Roll's critique, which has not yet been rebutted, is even more devastating to the CAPM for it brings us back to Popper's Falsifiable Principle (1934) according to which scientific theories should always lead to propositions that are potentially falsifiable by experimental observation.

Nevertheless, until a better theory is proposed, the CAPM will continue to exercise a dominant role in the theory of modern finance. The same can be said about the Modigliani–Miller perfect markets assumptions. Replacing these by more realistic assumptions has resulted in capital investment rules for each possible change in assumptions; each new rule not being sufficiently compelling to justify the replacement of the perfect market assumption. The unique decision criterion obtained under the perfect market assumption thus becomes a simplifying and unifying concept for teaching capital budgeting.

The Modigliani–Miller paradigm with its no default risk assumption might explain why the CAPM proponents have adopted equity return variance as a measure of risk. Again, in view of the fact that under the CAPM no default risk is assumed for any security and that only undiversifiable mean volatility of stock returns is relevant in determining the security's required rate of return, then the only relevant measure of risk is provided by the undiversifiable rate of return variance. Volatility of stock returns and risk became conceptual equivalents for the academic community. However, such a definition of risk runs counter to the everyday and commonly understood dictionary definition of risk. Risk is generally associated with the probability of occurrence of an undesirable event, for example, the probability of loss. One might then very well argue the CAPM has adopted a surrogate measure of risk which does not fit well in a world where the actual risk of default is very real.

15.3 TOTAL RISK AND THE REAL ECONOMY

Under the Modigliani–Miller (MM) paradigm, the economy is completely transparent and frictionless. A transparent economy implies that all economic agents share the same information and that there does not exist informational asymmetries between various stakeholders, such as current shareholders, debtholders, future shareholders, managers, suppliers, employees and customers among others. A frictionless economy means that the economy adjusts instantaneously and costlessly (no transaction or default costs) and that its workings are not hampered by various material, financial, managerial or organizational constraints. Consequently, financing an investment project through debt or equity is irrelevant in a MM economy. This explains why a project's value is the same irrespective of the firm that undertakes it or of its contribution to the firm's total risk. All valuable investment projects can be financed whatever the firm's financial position.

A similar type of reasoning is adopted under the CAPM efficient market hypothesis as shareholders are assumed to be investing directly into a project providing a required rate of return satisfying the one-period efficient market equilibrium conditions. Again, the economy is transparent and frictionless. Such a rule has been proposed and adopted by academics for over 40 years even though Eugene Fama's article (1970) demonstrates that the CAPM formula cannot generally be used for discounting cash flows in a multi-period framework. The CAPM decision rule states that any investment project with a positive expected NPV should be accepted, irrespective of its own volatility or of its contribution to the firm's total risk, for only systematic risk is relevant. We may qualify such a point estimate or certainty equivalent approach as normative to the extent that market conditions under which the investment process is taking place as well as the required rate of return both represent idealised conditions and are not at all descriptive of the actual workings of the real economy. Obviously, the proposed CAPM certainty equivalent evaluation and decision rule aim at determining the market value of an investment project. However, the market to which it is referring is the one of a one-period transparent and frictionless economy abiding by the Modigliani–Miller paradigm. In the very same sense, the

CAPM efficient market completely ignores the fact that the real economy is neither transparent nor frictionless.

Indeed, the legal system establishes a clear distinction as to the roles, the rights and obligations of debtholders, shareholders and managers. Such legal constraints imply informational asymmetries which explain why the financing of risky investment projects, even those profitable, is not easily obtained. The difficulty to finance risky investment projects is even exacerbated under conditions of financial distress as the total risk of the firm rapidly becomes the fundamental and dividing issue between the main stakeholders.

For instance, when the probability of financial distress or bankruptcy of a firm is not trivial, and, consequently, when its equity value is low, then funds provided by shareholders serve essentially to make safer the debtholders' risky outstanding debt, in addition to providing at their own expense the rate of return that the new shareholders will be seeking (Myers, 1977). We may also add that when a firm's probability of bankruptcy is significant, it becomes quite rational for shareholders to increase the total risk of the firm by accepting very risky investment projects that might very well rescue the value of their equity even if this implies increased risks at the expense of debtholders. The shareholders risk to lose little and to gain much for in the worst case scenario shares would become worthless anyway. The shareholders would be actually transferring part of their total risk to the debtholders and thus maximizing the wealth of shareholders instead of the value of the firm (Jensen and Meckling, 1976).

Also, as a firm's probability of financial distress increases, investors might find it evermore difficult, due to asymmetrical information, to distinguish sound projects that might increase the shareholders' value from pet projects that aim essentially at increasing the size of the firm and consequently the powerbase, perquisites consumption, salaries and stock options of top managers. As a consequence of informational asymmetries, valuable projects might be foregone in the process of capital budgeting given the cash shortage experienced by a firm under financial distress (Stulz, 1999).

The proposition according to which managers should be risk neutral and should be using the CAPM certainty equivalent decision rule is therefore not applicable when a firm experiences financial distress. That such a certainty equivalent decision rule has been proposed and used by academics for the last 40 years is understandable given that under the Modigliani–Miller perfect market paradigm it is always feasible to finance any profitable project even when a firm is close to financial distress. It should surprise nobody to learn that the CAPM equilibrium share price equation exposes itself to large values of probability of loss (Laughhunn and Sprecher, 1977). Now, considering that the CAPM assumes no default risk, it is quite logical that such an efficient market would set security prices without regard to the risk of bankruptcy caused by any failure to meet legal debt claims. Under the MM assumption of perfect and costless contracting:

> the problems that crop up when a firm becomes close to financial distress disappear because the firm can always costlessly recapitalize itself so that it is no longer close to financial distress. In the real world, such costless recapitalization is a dream. As a result, total risk matters and has to be taken into account when a firm evaluates a project. (Stulz, 1999: 9)

When a risky investment project imposes an additional cost on a financially strained firm through an increase of its total risk, the decision makers must quantify the marginal increase in total risk. To take this cost into account, managers have to quantify their total risk and have to understand how a new project might impact the firm's total risk. Being close to action, managers have both *ex ante* and *ex post* information advantage over shareholders and debtholders. They might therefore try to maximize their own welfare (as any typically rational person might do) at the expense of shareholders or debtholders. However, given appropriate incentives (this is what stock options aim at), managers will take decisions to the shareholders' advantage. However, when projects go astray, shareholders will hold project managers responsible for the failed project and will certainly not think about blaming the economy's systematic risk for its failure, the more so when the firm is in financial distress. The probability of loss then becomes important information, not as a criterion but as a constraint, in the selection and management of investment projects. Contrary to the concept of systematic risk which is drawn from a normative paradigm, the concept of total risk is derived from a positive probabilistic paradigm and aims at assessing the effective probability of loss. Therefore, it is just comes as a logical consequence that the hurdle rate that should be used to assess investment projects in a positive probabilistic paradigm should not be the CAPM prescribed cost of capital but the effective weighted marginal cost of capital of the firm.

The cost of total risk depends, among other things (a) on how the project is incorporated or organized, and (b) on how the firm is financed. The conventional capital budgeting practice is to include the project within the firm. Such a practice may not always be efficient considering that the credit risk supported by creditors is related to the firm's total risk, rather than just the project's risk. Given that creditors have claims against the entire firm, this implies the obligation to assess the firm's total portfolio of projects' and operations' risks, a costlier operation than assessing the risk of a single project (Shaw and Thakor, 1987). Furthermore, incorporating the project within the firm creates an asset substitution moral hazard whereby cash flows can be diverted from safe projects to riskier ones at the creditors' expense. Unless covenants prevent such substitution, creditors would recognize such a moral hazard and adjust the marginal cost of capital accordingly thereby impacting the total risk of the firm. On the other hand, organizing the project as a distinctive legal entity prevents such a substitution but generates its own types of risk. To the increased overhead costs and underinvestment moral hazard problem, one must consider the increased financial cost generated by the increased financial risk that must be supported by the incorporated project (Flannery, Houston and Venkataraman, 1993).

Contrary to what the MM paradigm asserts, the cost of total risk depends also on how the firm is financed. Debt financing improves the profitability of the firm by confering tax benefits and thus lowering its cost of capital, but makes the probability of financial distress and bankruptcy more likely. A highly levered firm must therefore assess the impact on the firm's total risk. Debt financing has a cost, so has equity financing. Otherwise, as Stulz (1999) remarks, all firms would be all-equity financed with no probability of financial distress. Agency costs and asymmetrical information explains why equity financing is costly since there are few all-equity firms. The cost of equity financing is, at the margin, equal to the cost of total risk (Stulz, 1999). When total risk matters, the appropriate measure of risk is obviously not an equity return volatility index. A firm can increase at no additional cost its total risk when the probability of financial distress is unaffected by a risky project:

> However, any increase in risk that increases the probability of distress is costly and should be accounted for when evaluating the costs and benefits of a project. Because the risk that is costly is the risk associated with large losses, the appropriate measures of risk are lower-tail measures of risk such as Value-at-Risk or Cash-flow-at-Risk rather than measures such as volatility of stock returns or volatility of cash-flows. (Stulz, 1999: 9)

In other words, one needs to know the probability distribution of risky investment projects.

15.4 THE NPV PROBABILITY DISTRIBUTION AND THE CLT: THEORETICAL RESULTS

Hillier (1963) invoked the Central Limit Theorem (CLT) to explain why the NPV probability distribution should be Normal. His conclusion rests on the argument that when the discounted cash flows are

> mutually independent random variables, with finite means and variances, which are either identically distributed or uniformly bounded, then (by the Lindeberg Theorem) the Central Limit Theorem will hold and the sum of these random variables will be approximately normal if n is large. If this holds, the probability distribution of the measures of the merit of an investment will be approximately normal, regardless of whether the X_j random variables are normal or not. (Hillier, 1963: 446)

However, Hillier (1964) would later on modify such a statement by pointing out that since the net present value equation is not the direct sum of random variables, but rather a sum of discounted cash flows, then the shape of the NPV distribution would be dominated by the early cash flows, the more so the higher the discount rate. Wagle (1967) would correctly conclude, without providing any mathematical proof:

> Thus even if we had independently distributed cash flows continuing forever, the variance of the present value of the first n cash flows would remain finite as $n \to \infty$, and in this case it is known that the distribution of the present value will not tend to normality unless each of the net cash flows is normally distributed. (Wagle, 1967: 18)

It is true that most versions of the CLT apply to a direct sum of independent random variables. However, as Wagle correctly argues, the fact that the NPV is the sum of discounted random cash flows does invalidate the CLT asymptotic convergence of the NPV probability distribution towards a Normal distribution. As shown in Appendix 1, the CLT does not apply strictly to the NPV probability distribution whenever the discount rate differs from zero; unless, of course, the probability distribution of cash flows is Normal.

As for the assumption of probability independence between net cash flows, a certain number of authors (Hoeffding and Robbins, 1948) have extended the CLT to the case of dependent random variables. However, the conditions under which these theorems are stated require conditional distributions, are subjected to very restrictive conditions, or involve special conditions which are difficult to comply with or to assess, most of all in the context of cash flow analysis. We prove in Appendix 2 that probability independence is not a necessary condition for obtaining asymptotic convergence towards a Normal distribution. More specifically, we show that the sum of equally-weighted first-order autoregressive cash flows converges toward a Normal probability distribution. Still, in the case where one would be dealing with discounted and first-order autoregressive cash flows, the NPV probability distribution would not strictly comply with the CLT. Considering that in a strict sense the discount rate, however small, ultimately invalidates the CLT, must we conclude, for practical purposes, that the CLT should never be used? Before exploring such a matter by simulation, let us consider (Appendix 1) the logarithm of the characteristic function in terms of its cumulants:

$$\Psi_{\frac{\sum \alpha_t \varepsilon_t}{\sqrt{\sum \alpha_t^2}}} = \sum_{t=1}^{n} \log \varphi_\varepsilon \left(\frac{\alpha_t h}{\sqrt{\sum \alpha_t^2}} \right) = -\frac{h^2}{2} - \frac{i}{3!} \sum_{t=1}^{n} \left(\frac{\alpha_t}{\sqrt{\sum \alpha_t^2}} \right)^3 h^3 K_3$$
$$+ \frac{1}{4!} \sum_{t=1}^{n} \left(\frac{\alpha_t}{\sqrt{\sum \alpha_t^2}} \right)^4 h^4 K_4 + \ldots$$

Table 15.1 Discount rates and the first term factor of the cumulants of the NPV probability distribution

k_c	K_3	K_4	K_5
0.01	0.00276	0.000388	0.000034
0.05	0.02834	0.008643	0.002635
0.10	0.07230	0.03012	0.01254
0.15	0.12042	0.05946	0.02936
0.20	0.16890	0.15277	0.05160
0.25	0.21600	0.12960	0.07776
0.30	0.26088	0.16669	0.10651
0.35	0.30310	0.20367	0.13682
0.40	0.34278	0.23990	0.16789
0.45	0.37972	0.27496	0.19916
0.50	0.41408	0.30860	0.23004

We show in Appendix 1 that the limit of the first term in the expansion of the factor multiplying each cumulant is given by:

$$\lim_{n\to\infty} \frac{\alpha_1^2}{\sum_{t=1}^{n} \alpha_t^2} = 1 - \frac{1}{(1+k_c)^2} \neq 0 \qquad \text{whenever} \quad k_c \neq 0$$

When $k_c = 0$, then all the cumulants of an order higher than 3 are multiplied by a weight of 0, thus ensuring the asymptotic convergence of the NPV probability distribution towards the Normal probability distribution. So what happens as to the effectiveness of the CLT when starting from 0 the discount rate k_c is increased progressively? Table 15.1 gives, for increasing values of the discount rate k_c, the limit value of the first and the largest term serving as weight for cumulants of order 3, 4 and 5.

From Table 15.1, it is obvious that the importance of the various cumulants decreases as their order increases. Therefore, there is no need to consider higher order cumulants. On the other hand, the importance of the weights of the various cumulants increases as the discount rate is increased. Also, the relative difference between the cumulant factors decreases as the discount rate is increased. In other words, higher-order cumulants acquire relative importance as the discount rate is increased. For low values of k_c, from 1% to 10%, we could be justified in assuming that the cumulants of order higher than 2 might not hamper the effectiveness of the CLT's asymptotic properties concerning the NPV probability distribution. For higher values of k_c, that is rates over 30%, it would seem quite plausible to assume that cumulants of order higher than 2 might invalidate the asymptotic properties of the NPV probability distribution.

15.5 THE NPV PROBABILITY DISTRIBUTION AND THE CLT: SIMULATION MODELS AND STATISTICAL TESTS

Our fundamental assumption is to the effect that as long as the discount rate does not exceed a threshold value (to be determined statistically), then the CLT applies to the NPV, for example, the NPV probability distribution does not differ significantly from a Normal distribution. To test such an assumption we have resorted to a simulation experiment using three probability distributions: (a) a uniform distribution, (b) a double exponential distribution, and finally (c) a normal distribution. The uniform probability distribution was chosen because it represents the case where uncertainty is at its maximum (for example, maximum entropy). This distribution is symmetrical and represents any extreme case for which decision-makers have very limited information. Thus, this case would apply when probability distributions are relatively symmetrical and are bell-shaped. The double exponential has the feature of having a thick tail. This might reveal particularly important and instructive considering that risk analysis, in the context of financial distress, focuses on the lower-tail of a probability distribution. However, we are quite aware that the double exponential distribution does provide a realistic description for net cash flows. As for the normal probability distribution, it was used to ensure, in accordance with our demonstration in Appendix 2 and contrary to what many authors have stated, the validity of the CLT even in situations with highly correlated cash flows.

The density of the uniform probability distribution is defined in the following fashion:

$$f_U(\tilde{u}_t) = \begin{cases} 1/2 & \text{for } \tilde{u}_t \in [-1, +1] \\ \\ 0 & \text{otherwise} \end{cases} \qquad \text{with } E_U(\tilde{u}) = 0 \text{ and } \sigma_U(\tilde{u}) = \sqrt{3}/3$$

The double exponential distribution has the following density:

$$f_{DE}(\tilde{u}_t) = \frac{1}{2}\, e^{-|u|} \quad \text{with } E_{DE}(\tilde{u}) = 0 \text{ and } \sigma_{DE}(\tilde{u}) = \sqrt{2}$$

The normal probability distribution is given the standardized form:

$$f_N(\tilde{u}_t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \quad \text{with } E_N(\tilde{u}) = 0 \text{ and } \sigma_N(\tilde{u}) = 1$$

To generate these random variables, we have used the University of Waterloo's Maple.8 simulation software. Simulations of 5,000 runs (NS: number of simulations) were carried out respectively on first-order and second-order autoregressive processes for increasing values of the NPV discount rate k_c. The first-order autoregressive process is defined by the following model:

$$\tilde{\varepsilon}_t = \rho\, \tilde{\varepsilon}_{t-1} + \tilde{u}_t \quad \text{for}: \ 0 \le \rho < 1$$

while the second-order autoregressive process is defined by:

$$\tilde{\varepsilon}_t = \rho_1 \tilde{\varepsilon}_{t-1} + \rho_2 \tilde{\varepsilon}_{t-2} + \tilde{u}_t \quad \text{for: } \begin{aligned} \rho_1 + \rho_2 &< 1 \\ \rho_1 - \rho_2 &< 1 \\ |\rho_2| &< 1 \end{aligned}$$

Thus, these two models comply with the conditions required for stochastic processes to be stationary in mean and in variance (Kendall, 1976; Nelson, 1973). All initial conditions of the simulation experiment were set at: $\tilde{\varepsilon}_1 = \tilde{u}_1$.

The chi-square distribution was used to test all the simulated NPV results for any significant statistical difference between the simulated distribution and a standardized Normal distribution. For the null hypothesis to be accepted, the differences between the theoretical and observed results must be attributable to sampling variability at the designated level of significance. The Normal probability distribution had been subdivided symmetrically into 8 classes with nearly equal probability.

The null hypothesis H_0 is the following: The simulated NPV probability distribution is Normal. The null hypothesis is to be rejected at the 1% level of significance for a chi-square distribution with 7 degrees of liberty when the calculated chi-square values are greater than 18.48. Otherwise, the null hypothesis is not rejected. The calculated chi-square statistic is given by:

$$\chi_c^2 = \sum_{i=1}^{r=8} \frac{(n_i - N\pi_i)^2}{N\pi_i}$$

where N = total number of simulation runs; π_i = theoretical probability from the standardized Normal distribution; n_i = number of simulation observations in class i; and r = total number of classes ($r = 8$).

15.6 THE NPV PROBABILITY DISTRIBUTION AND THE CLT: SIMULATION RESULTS

Let us consider the case of normally distributed cash flows. Table 15.2 summarizes the chi-square statistical test results for cash flows governed by a first-order autoregressive process extending over a 10-year period.

These results make it quite clear that highly correlated cash flows do not invalidate the normality of the NPV probability distribution and consequently the effectiveness of the CLT. Naturally, using a high discount rate does not invalidate the CLT. However, at the 5% level of significance, we would have found six cases for which the null hypothesis would have been rejected. This is quite normal, for this represents 6.66 percent of the total number of simulation trials, a percentage in agreement with such a level of significance.

Table 15.2 Calculated Chi-square table normal distribution $\tilde{\varepsilon} = \rho\tilde{\varepsilon}_{t-1} + \tilde{u}_t$

k_c|rho	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.00	5.0	2.7	7.9	5.5	3.6	1.4	3.6	10.9	5.7	6.4
0.05	7.3	14.7	9.0	7.3	5.7	0.6	8.6	3.7	3.9	3.9
0.10	4.8	3.9	6.9	2.3	3.3	12.2	17.2	15.9	11.7	14.2
0.15	4.8	11.1	3.0	10.2	5.1	11.8	4.1	3.3	7.1	1.7
0.20	2.9	4.2	7.8	2.5	2.0	3.1	3.2	4.8	3.7	7.3
0.25	7.9	4.3	3.9	4.6	5.1	1.5	8.5	4.8	1.7	9.9
0.30	3.2	2.0	15.7	4.6	5.2	3.5	2.2	0.9	1.9	11.0
0.35	4.1	4.5	4.0	3.6	8.9	7.6	14.0	5.9	4.1	8.3
0.40	3.0	13.5	15.1	2.6	8.1	2.5	1.8	11.3	9.2	4.4
***n* = 10**		NS = 5,000			$\chi^2_7(\alpha = 0.01) = 18.48$			$\chi^2_7(\alpha = 0.05) = 14.07$		

Table 15.3 Calculated Chi-square table uniform distribution $\tilde{\varepsilon}_t = \rho\tilde{\varepsilon}_{t-1} + \tilde{u}_t$

k_c|rho	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.00	1.4	12.1	8.7	12.6	4.2	3.5	7.8	10.3	8.1	5.4
0.05	5.1	8.6	4.1	3.1	5.8	7.8	5.6	8.3	7.6	9.9
0.10	7.8	15.6	1.1	5.8	6.5	4.2	5.4	4.1	3.9	13.5
0.15	9.9	6.2	7.3	9.2	8.4	5.4	0.8	6.8	9.6	4.9
0.20	7.2	6.4	7.4	7.2	**18.6(*)**	14.8	12.9	7.0	11.2	11.4
0.25	6.7	11.8	16.9	16.1	8.7	5.8	7.3	10.2	10.0	11.7
0.30	6.1	10.0	15.6	12.9	9.8	14.3	15.9	16.0	14.6	17.5
0.3	11.9	13.0	13.8	13.0	16.1	**20.7(*)**	**24.1(*)**	7.7	**20.6(*)**	13.9
0.40	**24.1(*)**	14.4	18.0	**32.8(*)**	15.4	17.4	14.8	14.5	**24.0(*)**	**20.0(*)**
***n* = 10**		NS = 5,000			$\chi^2_7(\alpha = 0.01) = 18.48$			$\chi^2_7(\alpha = 0.05) = 14.07$		

The following three Tables (15.3, 15.4 and 15.5) illustrate the results for uniformly distributed cash flows over three different time periods: 10, 20 and 40 years for a first-order autoregressive process. The bolded single asterisked (*) values represent cases where the null hypothesis has been rejected at the 1% level of significance.

These results make it quite clear that highly correlated cash flows do not invalidate the effectiveness of the CLT at ensuring the converging of the NPV probability distribution towards a normal distribution. We also observe that the CLT is not invalidated for discount rates lower than 30%, which is still quite high. Only at extremely high discount rates, such as 40%, can we say that the CLT is invalidates and the more so when a 5% level

Table 15.4 Calculated Chi-square table uniform distribution $\tilde{\varepsilon}_t = \rho\tilde{\varepsilon}_{t-1} + \tilde{u}_t$

k_c|rho	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.00	4.7	6.1	2.9	6.8	6.8	16.2	11.1	6.9	3.5	5.7
0.05	2.8	6.5	5.0	7.0	9.5	6.2	6.9	5.4	4.1	9.3
0.10	9.4	7.9	4.0	14.3	5.2	12.8	6.4	**19.0(*)**	5.2	2.2
0.15	8.3	7.7	3.9	7.0	9.1	8.0	9.7	15.6	7.3	2.9
0.20	9.7	10.4	4.1	14.3	6.4	12.1	9.0	6.3	7.0	17.4
0.25	17.4	13.5	5.8	**19.3(*)**	11.1	5.8	12.9	11.6	6.5	13.1
0.30	8.8	16.6	16.9	12.1	4.0	9.8	12.3	10.7	10.7	10.9
0.35	15.9	15.5	15.2	**20.0(*)**	10.9	**20.2(*)**	15.2	9.8	6.1	14.1
0.40	**13.3(*)**	**14.8(*)**	**16.1(*)**	**21.7(*)**	**14.9(*)**	**18.6(*)**	11.9	**13.7(*)**	**19.5(*)**	**24.5(*)**
***n* = 20**	NS = 5,000			$\chi^2_7(\alpha = 0.01) = 18.48$			$\chi^2_7(\alpha = 0.05) = 14.07$			

Table 15.5 Calculated Chi-square table uniform distribution $\tilde{\varepsilon}_t = \rho\tilde{\varepsilon}_{t-1} + \tilde{u}_t$

k_c|rho	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.00	8.2	6.2	4.4	10.8	11.9	6.0	5.2	14.9	5.3	6.0
0.05	19.0(*)	1.8	8.5	2.4	5.2	18.0	5.9	10.4	4.0	3.8
0.10	6.7	8.5	6.4	3.8	4.3	6.9	7.3	7.1	7.4	7.0
0.15	8.6	9.9	5.5	6.6	6.1	6.3	5.1	3.7	8.1	8.4
0.20	5.3	9.4	3.6	10.7	7.7	3.7	11.9	5.9	6.9	9.2
0.25	11.6	11.9	8.5	14.4	6.9	8.0	5.1	16.4	5.1	6.6
0.30	14.4	10.5	7.6	**21.3(*)**	**24.1(*)**	12.4	2.5	7.0	11.0	8.1
0.35	**19.4(*)**	14.7	7.8	14.1	9.3	**19.8(*)**	11.3	8.9	15.6	13.8
0.40	10.3	14.1	**24.1(*)**	11.5	**20.6(*)**	13.9	13.9	17.9	12.2	15.9
***n* = 40**	NS = 5,000			$\chi^2_7(\alpha = 0.01) = 18.48$			$\chi^2_7(\alpha = 0.05) = 14.07$			

of significance is used. However, we see that that increasing the number of years increases the effectiveness of the CLT. The following three tables (Table 15.6, 15.7 and 15.8) provide us the results for double exponential distributed cash flows over three different time periods: 10, 20 and 40 years for a first-order autoregressive process. These results make it quite clear that highly correlated cash flows do not invalidate the effectiveness CLT. Increasing the number of periods, however, improves the effectiveness of the CLT for values of the discount rate k_c lower than 20%. However, over this 20% threshold value, the CLT does not allow convergence of the NPV probability distribution towards a Gaussian distribution. Over the 25% discount

Table 15.6 Calculated Chi-square table double exponential distribution $\tilde{\varepsilon}_t = \rho\tilde{\varepsilon}_{t-1} + \tilde{u}_t$

k_c|rho	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.00	11.7	19.2(*)	3.1	12.8	13.2	4.9	5.5	11.7	**20.4(*)**	17.9
0.05	10.7	11.5	12.0	11.8	18.4	16.5	9.9	15.3	**22.3(*)**	**29.1(*)**
0.10	**23.7(*)**	5.1	**19.8(*)**	13.9	9.5	11.1	13.7	11.9	**19.9(*)**	**20.0(*)**
0.15	9.3	10.5	16.8	10.5	7.2	14.4	**18.5(*)**	15.4	9.5	10.7
0.20	**25.2(*)**	16.1	17.9	**28.7(*)**	**21.1(*)**	**47.4(*)**	17.0	**35.1(*)**	**27.4(*)**	**35.0(*)**
0.25	**24.9(*)**	**18.6(*)**	**25.1(*)**	17.2	17.1	**37.0(*)**	**28.2(*)**	15.7	**21.7(*)**	**29.1(*)**
0.30	**36.4(*)**	**26.4(*)**	**31.8(*)**	**30.3(*)**	**30.0(*)**	**21.2(*)**	**35.5(*)**	**30.6(*)**	**44.1(*)**	**37.8(*)**
0.35	**40.0(*)**	**35.1(*)**	**33.9(*)**	**24.5(*)**	**30.6(*)**	**37.8(*)**	**20.2(*)**	**26.8(*)**	**43.1(*)**	**39.4(*)**
0.40	**51.1(*)**	**34.6(*)**	**45.6(*)**	**28.6(*)**	**42.0(*)**	**58.1(*)**	**49.8(*)**	**45.3(*)**	**61.8(*)**	**47.8(*)**
$n = 10$		NS = 5,000			$\chi^2_7(\alpha = 0.01) = 18.48$			$\chi^2_7(\alpha = 0.05) = 14.07$		

Table 15.7 Calculated Chi-square table double exponential distribution $\tilde{\varepsilon}_t = \rho\tilde{\varepsilon}_{t-1} + \tilde{u}_t$

k_c|rho	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.00	2.9	10.7	9.3	5.3	9.0	5.2	10.9	7.9	10.0	6.8
0.05	**19.3(*)**	8.1	2.5	3.5	9.3	10.3	3.8	4.7	9.4	5.7
0.10	9.9	10.8	16.8	6.4	13.9	15.5	6.6	12.2	6.1	7.8
0.15	13.8	10.4	13.2	13.1	12.9	13.7	15.9	14.2	11.0	11.5
0.20	**19.4(*)**	15.5	17.9	8.2	6.6	6.2	**18.6(*)**	**32.9(*)**	**18.8(*)**	10.
0.25	**42.3(*)**	**29.2(*)**	**28.0(*)**	**30.2(*)**	**34.0(*)**	**29.1(*)**	**28.6(*)**	**27.6(*)**	**38.2(*)**	**33.5(*)**
0.30	**31.4(*)**	**46.5(*)**	**33.3(*)**	**34.5(*)**	**40.7(*)**	**31.9(*)**	**15.9**	**27.6(*)**	**52.2(*)**	17.4
0.35	17.9	**52.4(*)**	**39.2(*)**	**42.1(*)**	**60.4(*)**	**33.7(*)**	**39.1(*)**	**62.0(*)**	**23.2(*)**	**31.9(*)**
0.40	**42.6(*)**	**42.9(*)**	**31.4(*)**	**48.1(*)**	**48.0(*)**	**43.0(*)**	**27.9(*)**	**55.3(*)**	**40.8(*)**	**39.6(*)**
$n = 10$		NS = 5,000			$\chi^2_7(\alpha = 0.01) = 18.48$			$\chi^2_7(\alpha = 0.05) = 14.07$		

rate, the null hypothesis is rejected systematically as cash flows obeying a double exponential probability distribution do not conform to the CLT.

The simulation results, however, differ drastically from the uniform distribution when the discount rate is increased. We observe that as soon as the discount rate crosses the 20% line, then the CLT ceases to ensure convergence of the NPV probability distribution towards a normal distribution.

We therefore come to the conclusion that, up to a certain point, a thick-tailed distribution like the double exponential would limit the effectiveness of the CLT in ensuring the normality of the NPV distribution. In such a case,

Table 15.8 Calculated Chi-square table double exponential distribution $\tilde{\varepsilon}_t = \rho\tilde{\varepsilon}_{t-1} + \tilde{u}_t$

k_c|rho	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.00	5.1	3.2	7.9	10.4	4.2	4.1	4.6	5.0	3.2	5.3
0.05	5.3	6.0	5.8	2.4	6.6	2.3	9.5	5.5	5.0	5.7
0.10	6.7	15.7	16.4	13.5	7.2	4.7	11.0	5.9	11.1	5.2
0.15	4.5	**22.4(*)**	15.6	12.5	17.9	**20.4(*)**	**19.1(*)**	**26.3(*)**	13.7	**23.2(*)**
0.20	**19.4(*)**	15.7	**22.4(*)**	16.7	**21.2(*)**	10.5	**33.4(*)**	11.1	**28.3(*)**	12.3
0.25	**20.6(*)**	13.1	**26.5(*)**	**21.8(*)**	**42.0(*)**	**20.0(*)**	**21.0(*)**	**19.0(*)**	**19.5(*)**	**29.6(*)**
0.30	**35.1(*)**	**32.9(*)**	**22.5(*)**	**24.3(*)**	**19.2(*)**	**35.7(*)**	**35.4(*)**	**39.3(*)**	**35.1(*)**	**22.2(*)**
0.35	**34.7(*)**	**47.9(*)**	**24.5(*)**	**42.1(*)**	**20.1(*)**	**30.1(*)**	**42.0(*)**	**38.5(*)**	**25.7(*)**	**23.0(*)**
0.40	**41.4(*)**	**31.3(*)**	**39.4(*)**	**45.7(*)**	**43.5(*)**	**57.6(*)**	**47.8(*)**	**34.3(*)**	**37.4(*)**	**43.8(*)**
***n* = 40**		NS = 5000		$\chi^2_7(\alpha = 0.01) = 18.48$			$\chi^2_7(\alpha = 0.05) = 14.07$			

a 20% discount rate would constitute an upper limit, which, incidentally, is still pretty high. On the other hand, one has to ask oneself if such a thick-tailed distribution provides a realistic description of investment decision problems facing managers. One would suspect that such a distribution is fairly rare considering that they imply highly probable extreme values. Now, it is a well known fact that most investment decisions involve bounded monetary consequences. Decision-makers always have the possibility of opting out of an investment project in order to avoid the extremely negative consequences.

Simulation trials have also been carried out for second-order autoregressive processes. The results do not bring any noticeable difference from those obtained with first-order autoregressive processes under the same three probability distributions. The conclusion remains unchanged.

15.7 CONCLUSION

This chapter has dealt with the evaluation of risky capital investment projects when cash flows are serially dependent and conform either to a first-order or a second-order autoregressive stochastic stationary process. The authors have demonstrated mathematically that the NPV probability distribution does not strictly conform to the CLT asymptotic Normal distribution properties. The only exception occurs when the discount rate is set to zero. Under such conditions, it is also demonstrated that the CLT's limit property is not hampered when cash flows are serially dependant and obey a first-order autoregressive process.

However and as soon as a positive discount rate is introduced into the NPV equation, then the CLT does not apply in a strict mathematical sense. In fact, the higher the investment project discount rate and the less the CLT would be applicable to the NPV probability distribution. The authors explore through simulation runs and statistical testing, using the normal, uniform and double exponential probability distributions, the boundaries limiting the applicability of the CLT in ensuring convergence towards a Normal distribution.

In summary, the results are the following: managers and analysts are justified in invoking the CLT when assigning the normal probability distribution to an investment project probability distribution. As long as the cash flows are bell-shaped or uniformly distributed, the CLT may be invoked however highly serially correlated are the project cash flows and the discount rate. However, those projects whose cash flows have extremely high or low values with high probability may invalidate the CLT whenever the discount rate exceeds 15 to 20 percent; for low discount rates, the CLT would still be effective and reliable.

APPENDIX 1: THE CLT AND THE NPV PROBABILITY DISTRIBUTION

Let P be the present value of net cash flows X_t over a period of n years. Now, let us assume that these net cash flows are random variables $\tilde{X}_t$ with the additional features of being independent in probability and stationary in mean and in variance. The present value P must therefore be considered as a random variable $\tilde{P}$ equal to the weighted sum of the n net random cash flows $\tilde{X}_t$:

$$\tilde{P} = \sum_{t=1}^{n} \tilde{X}_t (1 + k_c)^{-t}$$

where k_c, the cost of capital, is the appropriate discount or hurdle rate. We posit that:

$$\tilde{X}_t = \mu_X + \tilde{\varepsilon}_t \qquad (\mu_X \text{ is a constant or a trend})$$

We further require the cash flow series to be stationary in mean and in variance. The cash flows are therefore expressed in terms of their deviation to such a trend, and without loss in generality, we may write:

$$\tilde{X}_t = \tilde{\varepsilon}_t \qquad \text{for } t = 1, 2, 3, \cdots, n$$

These random error terms are assumed independent in probability and obey the following probabilistic assumptions:

$$E(\tilde{\varepsilon}_t) = 0$$
$$V(\tilde{\varepsilon}_t) = \sigma_u^2 = 1 \quad \text{constant for all } t$$
$$\text{with} \quad Cov(\tilde{\varepsilon}_\tau, \tilde{\varepsilon}_\theta) = 0, \quad \text{for } \tau \neq \theta.$$

To simplify the notation, let us define $\alpha_t = (1+k_c)^{-t}$. Equation (1) becomes:

$$\tilde{P} = \sum_{t=1}^{n} \alpha_t \tilde{\varepsilon}_t$$

We must therefore verify, in accordance with the CLT, whether this weighted average of random error terms converges towards a normal probability distribution. Given our initial assumptions we deduct:

$$E(\tilde{P}) = \sum_{t=1}^{n} \alpha_t E(\tilde{\varepsilon}_t) = 0$$

whereas

$$V(\tilde{P}) = \sum_{t=1}^{n} \alpha_t^2 V(\tilde{\varepsilon}_t) = \sum_{t=1}^{n} \alpha_t^2.$$

Therefore, to verify the CLT we must demonstrate that:

$$\lim_{n\to\infty} \frac{\sum\limits_{t=1}^{n} \alpha_t \tilde{\varepsilon}_t}{\sqrt{\sum\limits_{t=1}^{n} \alpha_t^2}} \to N(0,1)$$

Let $\varphi_{\tilde{X}}(h) = E(e^{ih\tilde{X}}) = 1 + \sum_{t=1}^{\infty} \frac{(ih)^t}{t!} \mu_t$ be the characteristic function of any random variable $\tilde{X}$. Given that the $\tilde{\varepsilon}_t$'s are independent in probability, we may write the characteristic function of their weighted sum in term of their argument as:

$$\varphi_{\frac{\sum\limits_{t=1}^{n} \alpha_t \tilde{\varepsilon}_t}{\sqrt{\sum\limits_{t=1}^{n} \alpha_t^2}}}(h) = E\left(e^{\frac{ih\sum \alpha_t \tilde{\varepsilon}_t}{\sqrt{\sum \alpha_t^2}}}\right) = \prod_{t=1}^{n} e^{\frac{ih\alpha_t \tilde{\varepsilon}_t}{\sqrt{\sum \alpha_t^2}}} = \prod_{t=1}^{n} \varphi_{\tilde{\varepsilon}}\left(\frac{\alpha_t h}{\sqrt{\sum \alpha_t^2}}\right)$$

Let us take the logarithm of the characteristic function in term of its arguments and thus define the Ψ function:

$$\begin{aligned}
\Psi_{\frac{\sum \alpha_t \tilde{\varepsilon}_t}{\sqrt{\sum \alpha_t^2}}} &= \sum_{t=1}^{n} \log \varphi_{\tilde{\varepsilon}_t}\left(\frac{\alpha_t h}{\sqrt{\sum \alpha_t^2}}\right) \\
&= \sum_{t=1}^{n} \log \left[1 + i\frac{\alpha_t}{\sqrt{\sum \alpha_t^2}} h\mu_1 - \frac{1}{2}\left(\frac{\alpha_t}{\sqrt{\sum \alpha_t^2}}\right)^2 h^2 \mu_2 \right. \\
&\qquad \left. - \frac{i}{3!}\left(\frac{\alpha_t}{\sqrt{\sum \alpha_t^2}}\right)^3 h^3 \mu_3 + \frac{1}{4!}\left(\frac{\alpha_t}{\sqrt{\sum \alpha_t^2}}\right)^4 h^4 \mu_4 + \dots \right]
\end{aligned}$$

Using the development of $\log(1+z)$, given by $\log(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} - \frac{z^4}{4} + \ldots$, we obtain the logarithm of the characteristic function in terms of its cumulants:

$$\Psi_{\frac{\sum \alpha_t \tilde{\varepsilon}_t}{\sqrt{\sum \alpha_t^2}}} = \sum_{t=1}^{n} \log \varphi_{\tilde{\varepsilon}_t}\left(\frac{\alpha_t h}{\sqrt{\sum \alpha_t^2}}\right)$$

$$= \sum_{t=1}^{n}\left[i\frac{\alpha_t}{\sqrt{\sum \alpha_t^2}} hK_1 - \frac{1}{2}\left(\frac{\alpha_t}{\sqrt{\sum \alpha_t^2}}\right)^2 h^2K_2 - \frac{i}{3!}\left(\frac{\alpha_t}{\sqrt{\sum \alpha_t^2}}\right)^3 h^3K_3 + \frac{1}{4!}\left(\frac{\alpha_t}{\sqrt{\sum \alpha_t^2}}\right)^4 h^4K_4 + \ldots \right]$$

where K_j is the cumulant of order j. By assumption, we have set $K_1 = 0$ et $K_2 = 1$ whereas it is obvious that $\sum_{t=1}^{n}\left(\frac{\alpha_t}{\sqrt{\sum \alpha_t^2}}\right)^2 = 1$. Hence:

$$\Psi_{\frac{\sum \alpha_t \tilde{\varepsilon}_t}{\sqrt{\sum \alpha_t^2}}} = \sum_{t=1}^{n} \log \varphi_{\varepsilon}\left(\frac{\alpha_t h}{\sqrt{\sum \alpha_t^2}}\right) = -\frac{h^2}{2} - \frac{i}{3!}\sum_{t=1}^{n}\left(\frac{\alpha_t}{\sqrt{\sum \alpha_t^2}}\right)^3 h^3K_3 + \frac{1}{4!}\sum_{t=1}^{n}\left(\frac{\alpha_t}{\sqrt{\sum \alpha_t^2}}\right)^4 h^4K_4 + \ldots$$

which may be written as:

$$\Psi_{\frac{\sum \alpha_t \tilde{\varepsilon}_t}{\sqrt{\sum \alpha_t^2}}} = \sum_{t=1}^{n} \log \varphi_{\varepsilon}\left(\frac{\alpha_t h}{\sqrt{\sum \alpha_t^2}}\right) = -\frac{h^2}{2} - \frac{i}{3!}\sum_{t=1}^{n}\left(\frac{\alpha^2}{\sum \alpha_t^2}\right)^{3/2} h^3K_3 + \frac{1}{4!}\sum_{t=1}^{n}\left(\frac{\alpha_t^2}{\sum \alpha_t^2}\right)^2 h^4K_4 + \ldots$$

and quite obviously:

$$\lim_{n\to\infty} \frac{\alpha_1^2}{\sum_{t=1}^{n} \alpha_t^2} = \lim_{n\to\infty} \frac{(1+k_c)^{-2}}{\sum_{t=1}^{n}(1+k_c)^{-2t}} = \frac{(1+k_c)^2 - 1}{(1+k_c)^2} = 1 - \frac{1}{(1+k_c)^2} \neq 0$$

whenever $k_c \neq 0$.

Given that the cumulants of order higher than 2 generally are not equal to zero, it then follows that:

$$\lim_{n\to\infty} \Psi_{\frac{\sum \alpha_t \tilde{\varepsilon}_t}{\sqrt{\sum \alpha_t^2}}} = \lim_{n\to\infty} \sum_{t=1}^{n} \log \varphi_{\tilde{\varepsilon}} \left(\frac{\alpha_t h}{\sqrt{\sum \alpha_t^2}} \right)$$

$$= \lim_{n\to\infty} \left[-\frac{h^2}{2} - \frac{i}{3!} \sum_{t=1}^{n} \left(\frac{\alpha_t}{\sqrt{\sum \alpha_t^2}} \right)^3 h^3 K_3 + \frac{1}{4!} \sum_{t=1}^{n} \left(\frac{\alpha_t}{\sqrt{\sum \alpha_t^2}} \right)^4 h^4 K_4 + \ldots \right] \neq -\frac{h^2}{2}$$

However, when the discount rate is set equal to zero then:

$$\lim_{n\to\infty} \frac{\alpha_1^2}{\sum_{t=1}^{n} \alpha_t^2} = \frac{1}{n}$$

and the logarithm of the characteristic function in terms of its cumulants can be written as:

$$\Psi_{\frac{\sum \alpha_t \tilde{\varepsilon}_t}{\sqrt{\sum \alpha_t^2}}} = \sum_{t=1}^{n} \log \varphi_{\tilde{\varepsilon}} \left(\frac{\alpha_t h}{\sum \alpha_t^2} \right) = -\frac{h^2}{2} - \frac{i}{3!} \sum_{t=1}^{n} \left(\frac{1}{\sqrt{n}} \right)^3 h^3 K_3 + \frac{1}{4!} \sum_{t=1}^{n} \left(\frac{1}{\sqrt{n}} \right)^4 h^4 K_4 + \ldots$$

Consequently, its limit value can be written as:

$$\lim_{n\to\infty} \Psi_{\frac{\sum \alpha_t \tilde{\varepsilon}_t}{\sqrt{\sum \alpha_t^2}}} = \lim_{n\to\infty} \sum_{t=1}^{n} \log \varphi_{\tilde{\varepsilon}} \left(\frac{\alpha_t h}{\sum \alpha_t^2} \right) = -\frac{h^2}{2}$$

and therefore $\lim_{n\to\infty} \varphi_{\tilde{\varepsilon}} \left(\frac{\alpha_t h}{\sum \alpha_t^2} \right) = e^{-\frac{h^2}{2}}$, which is the characteristic function of the Normal probability distribution.

APPENDIX 2: THE CLT AND THE FIRST-ORDER AUTOREGRESSIVE PROCESS

We consider a weighted sum of random cash flows $\tilde{X}_t$ such that each variate has an equal weight. We thus define the random mean $\tilde{\bar{X}}$ as the sum of n equally weighted random cash flows $\tilde{X}_t$ as:

$$\tilde{\bar{X}} = \sum_{t=1}^{n} \frac{\tilde{X}_t}{n} \tag{A.1}$$

for which $\tilde{X}_t = \mu_X + \tilde{\varepsilon}_t$, for $t = 1, 2, 3 \ldots, n$.

We require the series to be stationary in mean and in variance which implies that they have a constant mean and that they fluctuate about the mean with a constant variance. Such constraints imply that the generating mechanism of the series remains constant through time. We therefore consider series from which the trend has been removed, or in which it was never present (for example, $\mu_X = 0$). The cash flows are therefore expressed in terms of their deviation to such a trend, and:

$$\tilde{X}_t = \tilde{\varepsilon}_t \qquad \text{for } t = 1, 2, 3, \dots, n \tag{A.2}$$

We furthermore assume that the random terms $\tilde{\varepsilon}_t$ obey a first-order autoregressive process defined by:

$$\tilde{\varepsilon}_t = \rho\tilde{\varepsilon}_{t-1} + \tilde{u}_t \qquad \text{for } t = 2, 3, 4, \dots, n \tag{A.3}$$

Stationarity in variance imposes the following additional condition:

$$0 \leq \rho_{t,t-1} < 1$$

Since the process begins at a specific date, we impose the following initial condition:

$$\tilde{\varepsilon}_1 = \tilde{u}_1$$

Finally, we introduce the following probabilistic assumptions:

$$\begin{aligned} E(\tilde{u}_t) &= 0 & & \\ V(\tilde{u}_t) &= \sigma_u^2 = 1 & & \text{constant for all } t \\ Cov(\tilde{u}_\theta, \tilde{u}_\tau) &= 0 & & \text{for: } \theta \neq \tau \\ Cov(\tilde{u}_t, \tilde{\varepsilon}_{t-1}) &= 0 & & \text{for: } t = 2, 3, 4, \dots, n \end{aligned}$$

Under such conditions, we demonstrate that the density probability function of $\tilde{\bar{X}}$ converges towards a Normal distribution. In other words, first-order autocorrelation between undiscounted cash flows does not invalidate the Central Limit Theorem. To demonstration such an assertion we may write from equation (A.3):

$$\tilde{\varepsilon}_t = \sum_{\tau=0}^{t-1} \rho^\tau \tilde{u}_{t-\tau} \qquad \text{for } t = 1, 2, 3, \dots, n \tag{A.4}$$

and equation (1) may be rewritten as:

$$\tilde{\bar{X}} = \sum_{t=1}^{n} \frac{\tilde{\varepsilon}_t}{n} = \frac{1}{n} \sum_{t=1}^{n} \sum_{\tau=0}^{t-1} \rho^\tau \tilde{u}_{t-\tau}$$

Thus yielding:

$$\tilde{\bar{X}} = \frac{1}{n} \left[\tilde{u}_1 \sum_{\tau=0}^{n-1} \rho^\tau + \tilde{u}_2 \sum_{\tau=0}^{n-2} \rho^\tau + \cdots + \tilde{u}_{n-2} \sum_{\tau=0}^{2} \rho^\tau + \tilde{u}_{n-1} \sum_{\tau=0}^{1} \rho^\tau + \tilde{u}_n \right]$$

for example:

$$\begin{aligned} \tilde{\bar{X}} = \frac{1}{n} \Bigg[& \tilde{u}_1 \left(\frac{1-\rho^n}{1-\rho} \right) + \tilde{u}_2 \left(\frac{1-\rho^{n-1}}{1-\rho} \right) + \tilde{u}_3 \left(\frac{1-\rho^{n-2}}{1-\rho} \right) + \cdots \\ & + \tilde{u}_{n-2} \left(\frac{1-\rho^3}{1-\rho} \right) + \tilde{u}_{n-1} \left(\frac{1-\rho^2}{1-\rho} \right) + \tilde{u}_n \left(\frac{1-\rho}{1-\rho} \right) \Bigg] \end{aligned}$$

and

$$\tilde{\bar{X}} = \frac{1}{n\,(1-\rho)}[\tilde{u}_1(1-\rho^n) + \tilde{u}_2(1-\rho^{n-1}) + \tilde{u}_3(1-\rho^{n-2}) + \cdots$$
$$+ \tilde{u}_{n-2}(1-\rho^3) + \tilde{u}_{n-1}(1-\rho^2) + \tilde{u}_n(1-\rho)]$$

By setting $w_t = 1 - \rho^{n-t+1}$ the mean random variable becomes equal to:

$$\tilde{\bar{X}} = \frac{1}{n\,(1-\rho)} \sum_{t=1}^{n} w_t \tilde{u}_t \tag{A.5}$$

Let us now demonstrate that the weighted sum of random variables $\tilde{u}_t$ obeys the CLT. Given the initial assumptions governing the random $\tilde{u}_t$'s, we obtain:

$$E(\tilde{\bar{X}}) = \frac{1}{n(1-\rho)} \sum_{t=1}^{n} w_t E(\tilde{u}_t) = 0$$

and

$$V(\tilde{\bar{X}}) = \frac{1}{n^2(1-\rho)^2} \sum_{t=1}^{n} w_t^2 V(\tilde{u}_t) = \frac{1}{n^2(1-\rho)^2} \sum_{t=1}^{n} w_t^2$$

The CLT will be established once we show that:

$$\lim_{n\to\infty} \frac{\sum_{t=1}^{n} w_t\, \tilde{u}_t}{\sqrt{\sum_{t=1}^{n} w_t^2}} \to N(0,1)$$

Given that the $\tilde{u}_t$'s are independent in probability, then the characteristic function is equal to:

$$\varphi_{\frac{\sum_{t=1}^{n} w_t \tilde{u}_t}{\sqrt{\sum_{t=1}^{n} w_t^2}}}(h) = E\left(e^{\frac{ih\sum w_t \tilde{u}_t}{\sqrt{\sum w_t^2}}}\right) = \prod_{t=1}^{n} e^{\frac{ihw_t \tilde{u}_t}{\sqrt{\sum w_t^2}}} = \prod_{t=1}^{n} \varphi_{\tilde{u}}\left(\frac{w_t h}{\sqrt{\sum w_t^2}}\right)$$

The logarithm of the characteristic function then becomes:

$$\Psi_{\frac{\sum w_t \tilde{u}_t}{\sqrt{\sum w_t^2}}} = \sum_{t=1}^{n} \log \varphi_{\tilde{u}_t}\left(\frac{w_t h}{\sqrt{\sum w_t^2}}\right)$$
$$= \sum_{t=1}^{n} \log\left[1 + i\frac{w_t}{\sqrt{\sum w_t^2}} h\mu_1 - \frac{1}{2}\left(\frac{w_t}{\sqrt{\sum w_t^2}}\right)^2 h^2\mu_2\right.$$
$$\left. - \frac{i}{3!}\left(\frac{w_t}{\sqrt{\sum w_t^2}}\right)^3 h^3\mu_3 + \frac{1}{4!}\left(\frac{w_t}{\sqrt{\sum w_t^2}}\right)^4 h^4\mu_4 + \ldots\right]$$

We may express the logarithm of the characteristic function in terms of its cumulants:

$$\Psi_{\frac{\sum w_t \tilde{u}_t}{\sqrt{\sum w_t^2}}} = \sum_{t=1}^{n} \log \varphi_{\tilde{\varepsilon}_t}\left(\frac{w_t h}{\sqrt{\sum w_t^2}}\right)$$

$$= \sum_{t=1}^{n} \left[i\frac{w_t}{\sqrt{\sum w_t^2}} h\, K_1 - \frac{1}{2}\left(\frac{w_t}{\sqrt{\sum w_t^2}}\right)^2 h^2 K_2 \right.$$

$$\left. - \frac{i}{3!}\left(\frac{w_t}{\sqrt{\sum w_t^2}}\right)^3 h^3 K_3 + \frac{1}{4!}\left(\frac{\alpha_t}{\sqrt{\sum \alpha_t^2}}\right)^4 h^4 K_4 + \ldots \right]$$

By assumption, we have set $K_1 = 0$ et $K_2 = 1$ whereas $\sum_{t=1}^{n}\left(\frac{w_t}{\sqrt{\sum w_t^2}}\right)^2 = 1$. Therefore:

$$\Psi_{\frac{\sum w_t \tilde{u}_t}{\sqrt{\sum w_t^2}}} = \sum_{t=1}^{n} \log \varphi_{\varepsilon}\left(\frac{w_t h}{\sqrt{\sum w_t^2}}\right) = -\frac{h^2}{2} - \frac{i}{3!}\sum_{t=1}^{n}\left(\frac{w_t}{\sqrt{\sum w_t^2}}\right)^3 h^3 K_3$$

$$+ \frac{1}{4!}\sum_{t=1}^{n}\left(\frac{w_t}{\sqrt{\sum w_t^2}}\right)^4 h^4 K_4 + \ldots$$

The limit value of the logarithm of the characteristic function becomes:

$$\lim_{n\to\infty} \Psi_{\frac{\sum w_t \tilde{u}_t}{\sqrt{\sum w_t^2}}} = \lim_{n\to\infty} \sum_{t=1}^{n} \log \varphi_{\tilde{\varepsilon}}\left(\frac{w_t h}{\sqrt{\sum w_t^2}}\right)$$

$$= \lim_{n\to\infty} \left[-\frac{h^2}{2} - \frac{i}{3!}\sum_{t=1}^{n}\left(\frac{w_t}{\sqrt{\sum w_t^2}}\right)^3 h^3 K_3 \right.$$

$$\left. + \frac{1}{4!}\sum_{t=1}^{n}\left(\frac{w_t}{\sqrt{\sum w_t^2}}\right)^4 h^4 K_4 + \ldots \right]$$

Furthermore, given $0 \le \rho < 1$, it follows that:

$$\lim_{n\to\infty} \frac{w_1^2}{\sum_{t=1}^{n} w_t^2} = \lim_{n\to\infty} \frac{(1-\rho^{n-t+1})}{\sum_{t=1}^{n}(1-\rho^{n-t+1})^2}$$

$$= \lim_{n\to\infty} \frac{(1-\rho^{n-t+i})}{\sum_{t=1}^{n}(1+\rho^{2(n-t+1)} - 2\rho^{n-t+1})}$$

$$= \lim_{n\to\infty} \frac{1}{n} = 0$$

Consequently:

$$\lim_{n\to\infty} \Psi_{\frac{\sum w_t \tilde{u}_t}{\sqrt{\sum w_t^2}}} = \lim_{n\to\infty} \sum_{t=1}^{n} \log \varphi_{\tilde{u}} \left(\frac{w_t h}{\sum w_t^2} \right) = -\frac{h^2}{2}$$

and

$$\lim_{n\to\infty} \varphi_{\tilde{u}} \left(\frac{w_t h}{\sum w_t^2} \right) = e^{-\frac{h^2}{2}}$$

which is the characteristic function of the Normal probability distribution.

REFERENCES

Blume, M., Crocket, J. and Friend, I. (1974) "Stock Ownership in the United States: Characteristics and Trends", *Survey of Current Business*, 54(11): 16–40.

Brennan, M.J. and Schwartz, E.S. (1985) "Evaluating Natural Resource Investments", *Journal of Business*, 58(2): 135–57.

Copeland, T. and Antikarov, V. (2001) *Real Options: A Practitioner's Guide* (London: Texere Publishing).

Cox, J., Ross, S. and Rubinstein, M. (1979) "Option Pricing: A Simplified Approach", *Journal of Financial Economics*, 7(3): 229–64.

Dixit, A.K. and Pindyck, R.S. (1994) "*Investment Under Uncertainty* (Princeton, NJ: Princeton University Press).

Dixit, A.K. and Pindyck, R.S. (1995) "The Options Approach to Capital Investment", *Harvard Business Review*, 73(3): 105–15.

Fama, E.F. (1970) "Multiperiod Consumption-Investment Decisions", *American Economic Review*, 60(1): 163–74.

Fama, E.F. and French, K.R. (1992) "The Cross-Section of Expected Stock Returns", *Journal of Finance*, 47(2): 427–65.

Flannery, M.J., Houston, J.F. and Venkataraman, S. (1993) "Financing Multiple Investment Projects", *Financial Management*, 22: 201–20.

Friedman, M. (1953) "The Methodology of Positive Economics", *Essays in Positive Economics* (Chicago, IL: University of Chicago Press).

Hertz, D.B. (1964) "Risk Analysis in Capital Investment", *Harvard Business Review*, 42(1): 95–106.

Hertz, D.B. (1968) "Investment Policies that Pay Off", *Harvard Business Review*, 46(1): 96–108.

Hillier, F.S. (1963) "The Derivation of Probabilistic Information for the Evaluation of Risky Investments", *Management Science*, 9(4): 443–57.

Hillier, F.S. (1964) "The Evaluation of Risky Interrelated Investment. Technical Report No 73 Contract Non-225(53) (NR-642-002) with the Office of Naval Research, published also in 1969 by North-Holland Publishing Co., Netherlands.

Hoeffding, W. and Robbins, H. (1948) "The Central Limit Theorem for Dependent Random Variables", *Duke Mathematical Journal*, 15(3): 773–80.

Ingersoll, Jr., J. and Ross, S. (1992) "Waiting to Invest: Investment and Uncertainty", *Journal of Business*, 65(1): 1–29.

Jagannathan, R. and Meier, I. (2002) "Do We Need CAPM for Capital Budgeting?", *Financial Management*, 31(4): 55–77.

Jensen, M.C. and Meckling, W.H. (1976) "Theory of the Firm: Managerial Behaviour, Agency Costs and Ownership Structure", *Journal of Financial Economics*, 3(4): 305–60.

Kendall, M. (1976) *Time Series* (New York, NY: Hafner Press).
Laughhunn, D.J. and Sprecher, C.R. (1977) "Probability of Loss and the Capital Asset Pricing Model", *Financial Management*, 6(2): 18–25.
Levy, H. (1978) "Equilibrium in an Imperfect Market: A Constraint on the Number of Securities in the Portfolio", *American Economic Review*, September, 68(4): 643–58.
Levy, H. and Sarnat, M. (1994). *Capital Investment & Financial Decisions* (London: Prentice-Hall).
Lintner, J. (1965) "The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets", *Review of Economic Statistics*, 47(1): 13–37.
McDonald, R. and Siegel, D. (1986) "The Value of Waiting to Invest", *Quarterly Journal of Economics*, November, 101(4): 707–27.
Miller, M.H. and Scholes, M. (1972) "Rate of Return in Relation to Risk: A Re-examination of Some Recent Findings", in M.C. Jensen (ed.), *Studies in the Theory of Capital Markets* (New York, NY: Praeger).
Mossin, J. (1966) "Equilibrium in a Capital Asset Market", *Econometrica*, October, 34(4): 768–83.
Myers, S.C. (1977) "Determinants of Corporate Borrowing", *Journal of Financial Economics*, November, 147–75.
Nelson, C.R. (1973) *Applied Time Series Analysis For Managerial Forecasting* (San Francisco, CA: Holden-Day, Inc.).
Popper, K. (1934) [translated (1959)] *The Logic of Scientific Discovery* (New York, NY: Harper & Row).
Poterba, J.M. and Summers, L.H. (1995) "A CEO Survey of US Companies' Time Horizons and Hurdle Rates", *Sloan Management Review*, 37(1): 43–53.
Roll, R. (1977) "A Critique of the Asset Pricing Theory's Tests", *Journal of Financial Economics*, 4(2): 129–76.
Sharpe, W.F. (1964) "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk", *Journal of Finance*, 19(3): 425–42.
Shaw, S. and Thakor, A.V. (1987) "Optimal Capital Structure and Project Financing", *Journal of Economic Theory*, 44(5): 209–43.
Stulz, R.M. (1999) "What's Wrong with Modern Capital Budgeting?", *Financial Practice and Education*, 9(2): 7–11.
Trigeorgis, L. (1993) "Real Options and Interactions with Financial Flexibility", *Financial Management*, 22(3): 202–24.
Wagle, B. (1967) "A Statistical Analysis of Risk in Investment Projects", *Operational Research Quarterly*, 18(1): 13–33.

CHAPTER 16

Have Volatility Transmission Patterns between the USA and Spain Changed after September 11?

*Helena Chuliá, Francisco J. Climent, Pilar Soriano and Hipòlit Torró**

16.1 INTRODUCTION

On 11 September 2001, the USA experienced its most devastating terrorist attack. This attack had an influence over several economic variables and it obviously affected financial markets. Taking into account the increasing global financial integration, an important question arises: Could recent terrorist attacks have increased even more interrelations between financial markets?

* The authors thank Andrew Patton and Farooq Malik for helpful comments and suggestions. Financial support from CICYT project BEC2003-09607-C04-04, and project GV04A/153 from Generalitat Valenciana and Instituto Valenciano de Investigaciones Económicas (IVIE) is gratefully acknowledged. H. Chuliá benefited from the financial support of a FPU grant from Ministerio de Educación y Ciencia and P. Soriano from the financial support of grant CTBPRB/2003/346 from Generalitat Valenciana.

The main objective of this study is to analyze whether volatility transmission patterns between the US and Spanish stock markets have changed after September 11. In order to do this, we use a multivariate GARCH model and take into account both the asymmetric volatility phenomenon and the non-synchronous trading problem.

An extensive literature has explored volatility transmission across global financial markets. Earlier studies, such as Engel, Ito and Lee (1990), Hamao, Masulis and Ng (1990), Susmel and Engle (1994), Karolyi (1995), Darbar and Deb (1997) and Booth, Martikainen and Tse (1997), use symmetric univariate or multivariate GARCH models. More recent studies introduce asymmetric multivariate GARCH models allowing volatility and covariance to be sensitive to the sign and size of the innovations (see Brooks and Henry, 2000; Ng, 2000; Isakov and Perignon, 2001; Baele, 2005; Cifarelli and Paladino, 2005).

Major global events such as the crisis in the USA in 1987, Mexico in 1994, East Asia in 1997 and Russia in 1998 have received special attention when studying volatility transmission and correlation between countries (Hamao, Masulis and Ng,1990; King and Wadhwani, 1990; Koutmous and Booth, 1995; and Rigobon, 2003). These studies use a sample size that includes one or more crisis and they estimate the model for the pre-crisis and post-crisis periods.

It must be highlighted that most existing studies on spillovers between developed countries focus on the USA, Canada, Japan, the UK, France and Germany.[1] As far as we know, volatility transmission between the USA and Spain has only been studied by Peña (1992), Perez and Torra (1995) and Fernández and Aragó (2003). Peña (1992) uses a symmetric univariate model and the others use multivariate GARCH models. All of them find volatility transmission patterns between the USA and Spain. However, this chapter will be the first to take into account the non-synchronous trading problem and use a sample period that includes the September 11 terrorist attack.

Until now, few studies have examined the effects of the attacks of September 11 on financial markets. Most of them focus on the economy as a whole[2] or on different concrete aspects of the economy. For instance, Poteshman (2005) analyzes whether there was unusual option market activity prior to the terrorist attacks. Ito and Lee (2005) assess the impact of the September 11 attack on US airline demand. Glaser and Weber (2005) focus on how the terrorist attack influenced expected returns and volatility forecasts of individual investors. However, none of them analyzes volatility transmission patterns and how they have been affected by the event. The only paper which analyzes changes in interrelations between stock markets would be Hon, Strauss and Yong (2004), but it tests whether the terrorist attack resulted in an increase in correlation across global financial markets.

When studying volatility transmission between different financial markets, an important fact to take into account is the trading hours in each market. In the case of partially overlapping markets (like Spain and the USA), a jump in prices can be observed in the first market to open when the second one starts trading, reflecting information contained in the opening price. Therefore, this could make volatility increase in this first market. Moreover, as suggested by Hamao, Masulis and Ng (1990), a correlation analysis between partially overlapping markets using close to close returns could produce false spillovers, both in mean and volatility. This is so because it is difficult to separate effects coming from the foreign market from those coming from the own market while it remains closed.

There are several solutions in order to artificially synchronize international markets. First of all, in the case of the USA, information transmission with other markets can be analyzed through American Depositary Receipts (ADRs), which will share trading hours with the US market. The problem is that there are no many ADRs, they are not actively traded and there are microstructure differences between the North American stock market and that from the original country (Wongswan, 2003). Some studies, such as Longin and Solnik (1995) and Ramchand and Susmel (1998), use weekly or monthly data in order to avoid the non-synchronous trading problem. However, the use of low frequency data leads to small samples, which is inefficient for multivariate modeling. On the other hand, some studies, such as Hamao, Masulis and Ng (1990) and Koutmos and Booth (1995), use daily non-synchronous open-to-close and close-to-open returns. Nevertheless, these studies cannot distinguish volatility spillovers from contemporaneous correlations. Finally, Martens and Poon (2001) use 16:00-to-16:00 synchronous stock market series in order to solve this problem. By doing this, they find a bidirectional spillover between the USA and France and between the USA and the UK, contrary to previous studies that only found volatility spillovers from the USA to other countries.

This study innovates with respect the existing literature in two ways. First, we study volatility transmission between the USA and Spain using a recent sample period including the terrorist attack occurred in the USA. As far as we know, September 11 has not yet been included in any paper analyzing volatility transmission in international markets. Second, we use 16:00-to-16:00 synchronous stock market series which reduce the non-synchronous trading problem.

The rest of the chapter is organized as follow. Section 16.2 presents the data and offers some preliminary analysis. Section 16.3 deals with the econometric approach and the asymmetries analysis; section 16.4 presents the empirical results, and, finally, section 16.5 summarizes the main results.

16.2 DATA

The data consist of daily stock market prices recorded at 16:00 Spanish time for the USA (S&P500 index) and Spain (IBEX35 index). Trading at the Spanish

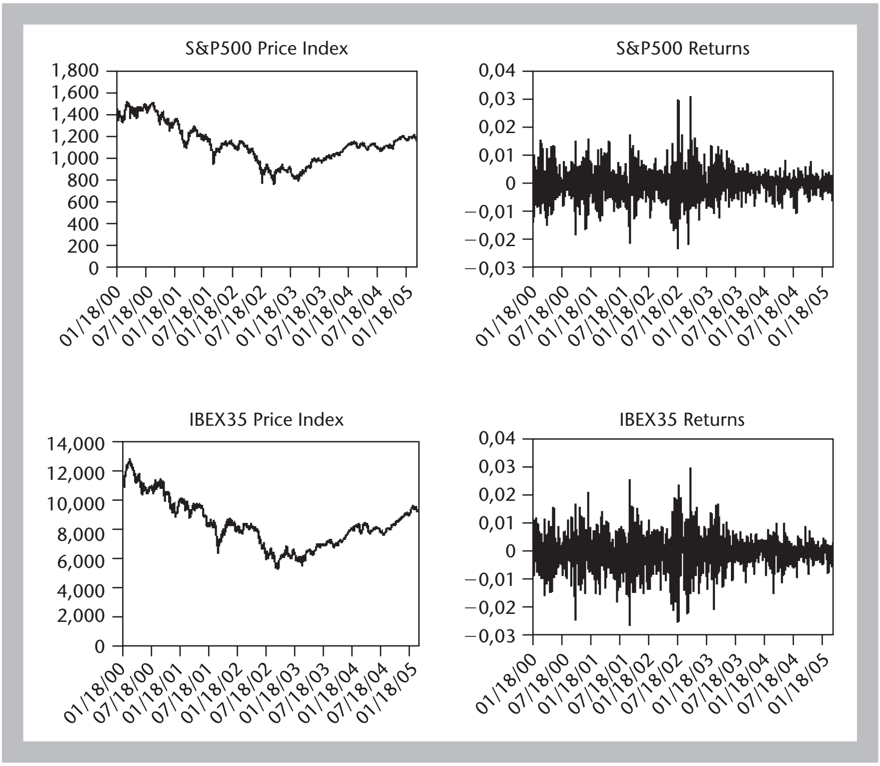


Figure 16.1 Price indexes and returns

Stock Exchange starts at 9:00 and finishes at 17:35 Spanish time. The New York Stock Exchange trades from 9:30 to 16:00 US eastern standard time (15:30 to 22:00 Spanish time). Apart from a few days when these countries change to the summer daylight saving time and again later to the winter time, the overlapping trading hours correspond to Spanish time 15:30 to 17:35. We use stock market prices recorded at 16:00 Spanish time in order to avoid the use of index prices near the open (USA) and close (Spain) of trading.

The data is extracted from Visual Chart Group (www.visualchart.com) for the period 18 January 2000 to 11 April 2005.[3] When there are no common trading days due to holidays in one of the markets, the index values recorded on the previous day are used. Figure 16.1 displays the daily evolution of the indexes and their returns in the studied period.

Table 16.1 presents some summary statistics on the daily returns, which are defined as log differences of index values. The Jarque–Bera test rejects normality of the returns for both countries. This is caused mainly by the excess kurtosis, suggesting that any model for equity returns should accommodate this characteristic of equity returns. The ARCH test reveals that

Table 16.1 Summary statistics

	$R_{1,t}$	*p*-value	$R_{2,t}$	*p*-value
Mean	−0.00015		−0.00013	
Variance	0.00014		0.00019	
Skewness	0.11161	[0.0927]	−0.06255	[0.3461]
Kurtosis	5.97705	[0.0000]	4.90041	[0.0000]
Bera–Jarque	506.536	[0.0000]	206.146	[0.0000]
Q(6)	6.34085	[0.3861]	2.93234	[0.8172]
$Q^2(6)$	214.134	[0.0000]	230.674	[0.0000]
ARCH(6)	126.292	[0.0000]	125.215	[0.0000]
ADF(4)	−1.86827	[0.3476]	−1.72608	[0.4179]
PP(6)	−1.89895	[0.3330]	−1.74016	[0.4107]

Notes: *p*-values displayed as [.]. $R_{1,t}$ and $R_{2,t}$ represent the log-returns of the S&P500 and the IBEX35 indices. The Bera–Jarque statistic tests for the normal distribution hypothesis and has an asymptotic distribution $X^2(2)$. Q(6) and $Q^2(6)$ are Ljung-Box tests for sixth-order serial correlation in the returns and squared returns. ARCH(6) is Engle's test for sixth order ARCH, distributed as $X^2(6)$. The ADF (number of lags) and PP (truncation lag) refer to the Augmented Dickey–Fuller (1981) and Phillips and Perron (1988) unit-root tests. Critical value at 5% significance level of Mackinnon (1991) for the ADF and PP tests (process with intercept but without trend) is −2.86.

returns exhibit conditional heteroskedasticity, while the Ljung–Box test (of sixth order) indicates significant autocorrelation in both markets in squared returns but not in levels. Fat tails and non-normal distributions are common features of financial data. Finally, both the augmented Dickey–Fuller (ADF) and Philips and Perron (PP) tests indicate that both series have a single unit root.

The sample has been divided into two similar-length sub-samples in order to separately analyze the volatility patterns before (18 January 2000–10 September 2001) and after (11 October 2001–11 October 2003) the terrorist attack. There is one month excluded between the pre- and post- sub-samples because during the crisis period other events, such as the attack of the USA on Afghanistan, affected stock market returns making them more unstable. Moreover, US financial markets were closed until 17 September. Finally, it took one month for the S&P500 index to recover to the original level it had before the tragic event occurred.

Table 16.2 displays returns, correlations and volatilities, period by period, for both stock indexes, the S&P500 and the IBEX35. Three facts can be highlighted from these panels. Firstly, in all periods, means equality test can not be rejected. Secondly, the IBEX35 is more volatile than the S&P500 and the Levene's test rejects variance equality in all periods except before September 11. This last result is significant, since it may suggest a change in volatility

Table 16.2 Returns, correlations and volatilities

Panel A: returns and correlations				
Period	μ_1	μ_2	**Mean Test** $H_o{:}\mu_1 = \mu_2$	$\rho_{1,2}$
Total	−0.00015	−0.00013	0.05543	0.6945
Pre-09/11	−0.00069	−0.00087	0.19266	0.6165
Post-09/11	−0.00004	−0.00011	0.07929	0.7332
Mean test $H_o{:}\mu_{pre} = \mu_{post}$	0.73899	0.76731	–	–

Panel B: volatilities			
Period	σ_1	σ_2	**Levene Test** $H_o{:}\sigma_1^2 = \sigma_2^2$
Total	0.01203	0.01399	20.38690*
Pre-09/11	0.01314	0.014449	2.48221
Post-09/11	0.01381	0.01629	12.03559*
Levene test $H_o{:}\sigma_{pre} = \sigma_{post}$	0.01171	3.62934	–

Notes: $\mu_1(\mu_2)$ displays the daily mean return of the S&P500 (IBEX35). $\sigma_1^2(\sigma_2^2)$ displays the daily standard deviation of the S&P500 (IBEX35). Mean test tests the null hypothesis of equality of daily mean returns. The Levene's statistic tests the null hypothesis of equality of daily variances. $\rho_{1,2}$ displays the correlation between both indexes computed from their daily returns in that period. An asterisk (*) denotes a test statistic that exceeds 5% critical value.

Table 16.3 Johansen (1988) test for cointegration

Lags	Null	$\lambda_{trace}(r)$	Critical Value	$\lambda_{max}(r)$	Critical Value
3	r = 0	17.97919	19.96	14.76784	15.67
	r = 1	3.211352	9.24	3.211352	9.24

Notes: The lag length is determined using the AIC criterion. $\lambda_{trace}(r)$ tests the null hypothesis that there are at most r cointegration relationships against the alternative that the number of cointegration vectors is greater than *r*. $\lambda_{max}(r)$ tests the null hypothesis that there are r cointegration relationships against the alternative that the number of cointegration vectors is greater than $r+1$. Critical values are from Osterwald–Lenum (1992).

transmission patterns following the terrorist attack and, therefore, motivates the analysis that will follow. Finally, correlation between both series has increased over time. Table 16.3 shows that both series are not cointegrated, being three the optimal lag length following the AIC criterion.

16.3 THE ECONOMETRIC APPROACH

16.3.1 The model

The econometric model is estimated in a three-step procedure. First, a VAR model is estimated to clean up any autocorrelation behavior. Then, the residuals of the model are orthogonalized. These orthogonalized innovations have the convenient property that they are uncorrelated both across time and across markets. Finally, the orthogonalized innovations will be used as an input to estimate a multivariate asymmetric GARCH model.

Equation (16.1) models the mean equation as a VAR(1) process:

$$\begin{bmatrix} R_{1,t} \\ R_{2,t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix} \begin{bmatrix} R_{1,t-1} \\ R_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix} \tag{16.1}$$

where $R_{1,t}$ and $R_{2,t}$ are USA and Spain returns, respectively, μ_i and d_{ij} for $i=1,2$ are the parameters to be estimated and $u_{1,t}$ and $u_{2,t}$ are the non-orthogonal innovations. The VAR lag has been chosen following the AIC criterion.

The innovations $u_{1,t}$ and $u_{2,t}$ are non-orthogonal because, in general, the covariance matrix $\sum = E(u_t u_t')$ is not diagonal. In order to overcome this problem, in a second step, the non-orthogonal innovations ($u_{1,t}$ and $u_{2,t}$) are orthogonalized ($\varepsilon_{1,t}$ and $\varepsilon_{2,t}$). If we choose any matrix M so that $M^{-1} \sum M'^{-1} = I$, then the new innovations:

$$\varepsilon_t = u_t M^{-1} \tag{16.2}$$

satisfy $\mathrm{E}(\varepsilon_t \varepsilon_t') = I$. These orthogonalized innovations have the convenient property that they are uncorrelated both across time and across equations. Such a matrix M can be any solution of $MM' = \sum$. In this study we have used a structural decomposition of the form suggested by Bernanke (1986) and Sims (1986). In contrast to the Cholesky factorization, this methodology does not embody strong assumptions about the underlying economic structure.

To model the conditional variance-covariance matrix we use an asymmetric version of the BEKK model (see Baba, Engle, Kraft and Kroner, 1989; Engle and Kroner, 1995; and Kroner and Ng, 1998). The compacted form of this model is:

$$H_t = C'C + B'H_{t-1}B + A'\varepsilon_{t-1}\varepsilon_{t-1}'A + G'\eta_{t-1}\eta_{t-1}'G \tag{16.3}$$

where C, A, B and G are matrices of parameters, being C upper-triangular and positive definite and H_t is the conditional variance-covariance matrix in t.

In the bivariate case, the BEKK model is written as follows:

$$\begin{bmatrix} h_{11t} & h_{12t} \\ \cdot & h_{22t} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ 0 & c_{22} \end{bmatrix}' \begin{bmatrix} c_{11} & c_{12} \\ 0 & c_{22} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}' \begin{bmatrix} h_{11,t-1} & h_{12,t-1} \\ \cdot & h_{22,t-1} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}' \begin{bmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \cdot & \varepsilon_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}' \begin{bmatrix} \eta_{1,t-1}^2 & \eta_{1,t-1}\eta_{2,t-1} \\ \cdot & \eta_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \quad (16.4)$$

where $c_{i,j}$,$b_{i,j}$,$a_{i,j}$ and $g_{i,j}$ for all $i, j = 1, 2$ are parameters, $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are the unexpected shock series coming from equation (16.2),$\eta_{1,t} = \max[0, -\varepsilon_{1,t}]$ and $\eta_{2,t} = \max[0, -\varepsilon_{2,t}]$ are the Glosten, Jagannathan and Runkle (1993) dummy series collecting a negative asymmetry from the shocks and, finally, $h_{ij,t}$ for all $i, j = 1, 2$ are the conditional second-moment series.

Equation (16.4) allows for both own-market and cross-market influences in the conditional variance, therefore allowing the analysis of volatility spillovers between both markets. Moreover, the BEKK model guarantees by construction that the variance-covariance matrix will be positive definite.

In equation (16.4), parameters $c_{i,j}$,$b_{i,j}$,$a_{i,j}$ and $g_{i,j}$ for all $i, j = 1, 2$ can not be interpreted individually. Instead, we have to interpret the non-linear functions of the parameters which form the intercept terms and the coefficients of the lagged variances, covariances and error terms. We follow Kearney and Patton (2000) and calculate the expected value and the standard error of those non-linear functions. The expected value of a non-linear function of random variables is calculated as the function of the expected value of the variables, if the estimated variables are unbiased. In order to calculate the standard errors of the function, a first-order Taylor approximation is used. This linearizes the function by using the variance-covariance matrix of the parameters as well as the mean and standard error vectors.

The parameters of the bivariate BEKK system are estimated by maximizing the conditional log-likelihood function:

$$L(\theta) = -\frac{TN}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left(\ln|H_t(\theta)| + \varepsilon_t' H_t^{-1}(\theta)\,\varepsilon_t\right) \quad (16.5)$$

where T is the number of observations, N is the number of variables in the system and θ denotes the vector of all the parameters to be estimated. Numerical maximization techniques were used to maximize this non-linear log likelihood function based on the BFGS algorithm.

In order to estimate the model in equations (16.1), (16.2) and (16.3), it is assumed that the vector of innovations is conditionally normal and a quasi-maximum likelihood method is applied. Bollerslev and Wooldridge (1992) show that the standard errors calculated using this method are robust even when the normality assumption is violated.

16.3.2 Asymmetric volatility impulse response functions (AVIRF)

The Volatility Impulse-Response Function (VIRF), proposed by Lin (1997), is a useful methodology for obtaining information on the second-moment interaction between related markets. The VIRF, and its asymmetric version, measure the impact of an unexpected shock on the predicted volatility. This is:

$$R_{s,3} = \frac{\partial vechE\left[H_{t+s}|\psi_t\right]}{\partial dg(\varepsilon_t \varepsilon_t')} \tag{16.6}$$

where $R_{s,3}$ is a 3×2 matrix, $s=1,2,\ldots$ is the lead indicator for the conditioning expectation operator, H_t is the 2×2 conditional covariance matrix, $\partial dg(\varepsilon_t \varepsilon_t') = (\varepsilon_{1,t}^2,\ \varepsilon_{2,t}^2)'$ and ψ_t is the set of conditioning information. The *vech* operator transforms a symmetric NxN matrix into a vector by stacking each column of the matrix underneath the other and eliminating all supradiagonal elements.

In volatility symmetric structures, it is not necessary to distinguish between positive and negative shocks, but with asymmetric structures the VIRF can change with the sign of the shock. The asymmetric VIRF (AVIRF) for the asymmetric BEKK model is taken from Meneu and Torró (2003) by applying (16.6) to (16.3):

$$R_{s,3}^{+} = \begin{cases} a & s=1 \\ (a+b+1/2g)R_{s-1,3}^{+} & s>1 \end{cases} \tag{16.7}$$

$$R_{s,3}^{-} = \begin{cases} a+g & s=1 \\ (a+b+1/2g)R_{s-1,3}^{-} & s>1 \end{cases} \tag{16.8}$$

where $R_{s,3}^{+}$ ($R_{s,3}^{-}$) represents the VIRF for positive (negative) initial shocks and a, b and g are 3×3 parameter matrices. Moreover, $a = D_N^{+}(A' \otimes A')D_N$, $b = D_N^{+}(B' \otimes B')D_N$, $g = D_N^{+}(G' \otimes G')D_N$, where D_N is a duplication matrix, D_N^{+} is its Moore-Penrose inverse and $\otimes$ denotes the Kronecker product between matrices, that is:

$$D_N = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad D_N^{+} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & ½ & 0 & 0 \\ 0 & 0 & ½ & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{16.9}$$

It is important to note that this impulse response function examines how fast asset prices can incorporate new information. This fact lets us test for the speed of adjustment, analyze the dependence of volatilities across the returns of the S&P500 and the IBEX35 and distinguish between negative and positive shocks.

16.4 EMPIRICAL RESULTS

16.4.1 Model estimation

Table 16.4 displays the estimated VAR-asymmetric BEKK model of equations (16.1) to (16.3). Panels A, B and C show the estimation results for the total period, the pre-September 11 period and the post-September 11 periods respectively. The low *p*-values obtained for most of the parameters shows that the model fits well the data. Table 16.5 shows the standardized residuals analysis. It can be observed that the standardized residuals appear free from serial correlation and heteroskedasticity.

As has been mentioned above, the parameters of Table 16.4 cannot be interpreted individually. Instead, we have to focus on the non-linear functions that form the intercept terms and the coefficients of the lagged variance, covariance and error terms. Panels A, B and C of Table 16.6

Table 16.4 Estimation results for the VAR(1)-asymmetric BEKK model

Panel A: total period		
	$R_{1,t}$	$R_{2,t}$
μ	−0.000152	−0.000118
	(0.64)	(0.75)
$R_{1,t-1}$	0.000906	0.081108
	(0.98)	(0.06)
$R_{2,t-1}$	−0.008487	−0.037128
	(0.79)	(0.32)

$$C=\begin{bmatrix} \underset{(0.00)}{-0.001149} & \underset{(0.29)}{0.000160} \\ & \underset{(0.00)}{-0.000469} \end{bmatrix} \qquad B=\begin{bmatrix} \underset{(0.00)}{0.949919} & \underset{(0.00)}{0.031741} \\ \underset{(0.04)}{-0.024743} & \underset{(0.00)}{0.943548} \end{bmatrix}$$

$$A=\begin{bmatrix} \underset{(0.00)}{-0.155282} & \underset{(0.00)}{0.222199} \\ \underset{(0.00)}{-0.110431} & \underset{(0.09)}{-0.059604} \end{bmatrix} \qquad G=\begin{bmatrix} \underset{(0.00)}{0.317651} & \underset{(0.64)}{0.018773} \\ \underset{(0.06)}{0.033911} & \underset{(0.00)}{0.266847} \end{bmatrix}$$

Continued

Table 16.4 Continued

Panel B: pre-September 11

	$R_{1,t}$	$R_{2,t}$
μ	−0.00705	−0.000866
	(0.26)	(0.21)
$R_{1,t-1}$	0.023479	0.048834
	(0.70)	(0.47)
$R_{2,t-1}$	−0.045542	−0.028893
	(0.41)	(0.63)

$$C = \begin{bmatrix} \underset{(0.00)}{0.002731} & \underset{(0.00)}{0.004694} \\ & \underset{(0.99)}{-0.000000} \end{bmatrix} \qquad B = \begin{bmatrix} \underset{(0.00)}{-0892311} & \underset{(0.17)}{0.038724} \\ \underset{(0.00)}{0.099708} & \underset{(0.00)}{-0.860363} \end{bmatrix}$$

$$A = \begin{bmatrix} \underset{(0.43)}{0.066746} & \underset{(0.00)}{-0.207397} \\ \underset{(0.17)}{0.123347} & \underset{(0.73)}{0.026290} \end{bmatrix} \qquad G = \begin{bmatrix} \underset{(0.00)}{0.445703} & \underset{(0.84)}{-0.015082} \\ \underset{(0.00)}{0.206972} & \underset{(0.00)}{0.274034} \end{bmatrix}$$

Panel C: post-september 11

	$R_{1,t}$	$R_{2,t}$
μ	−0.000091	−0.000191
	(0.87)	(0.78)
$R_{1,t-1}$	−0.002057	0.163352
	(0.97)	(0.03)
$R_{2,t-1}$	−0.006250	−0.095988
	(0.90)	(0.13)

$$C = \begin{bmatrix} \underset{(0.00)}{0.001351} & \underset{(0.00)}{-0.001115} \\ & \underset{(0.99)}{-0.000000} \end{bmatrix} \qquad B = \begin{bmatrix} \underset{(0.00)}{0.954493} & \underset{(0.17)}{0.029803} \\ \underset{(0.18)}{0.004472} & \underset{(0.00)}{0.958434} \end{bmatrix}$$

$$A = \begin{bmatrix} \underset{(0.00)}{-0.137017} & \underset{(0.00)}{0.180273} \\ \underset{(0.00)}{-0.138980} & \underset{(0.12)}{0.044365} \end{bmatrix} \qquad G = \begin{bmatrix} \underset{(0.00)}{0.295883} & \underset{(0.20)}{-0.018435} \\ \underset{(0.19)}{0.029120} & \underset{(0.00)}{0.178761} \end{bmatrix}$$

Notes: This table shows the estimation of the model defined in equations (16.1), (16.2) and (16.3). *p*-values appear in brackets. In the three cases the necessary conditions for the stationarity of the process are satisfied.

display the expected value and the standard errors of these non-linear functions for the total period, the pre-September 11 and the post-September 11 period.

Table 16.5 Summary statistics for the standardized residuals of the model

Panel A: summary statistics for the standardized residuals of the total period				
	$\varepsilon_{1,t}/\sqrt{h_{11,t}}$		$\varepsilon_{2,t}/\sqrt{h_{22,t}}$	
Q(6)	3.94519	[0.68409]	4.40124	[0.62255]
$Q^2(6)$	3.85331	[0.69652]	5.51455	[0.47970]
ARCH(6)	1.25783	[0.97393]	3.26566	[0.77484]
Panel B: summary statistics for the standardized residuals of the pre-September 11 period				
	$\varepsilon_{1,t}/\sqrt{h_{11,t}}$		$\varepsilon_{2,t}/\sqrt{h_{22,t}}$	
Q(6)	1.94140	[0.92501]	3.48928	[0.74540]
$Q^2(6)$	2.57484	[0.86000]	12.4568	[0.05252]
ARCH(6)	1.58439	[0.95369]	7.36116	[0.28873]
Panel C: summary statistics for the standardized residuals of the post-September 11 period				
	$\varepsilon_{1,t}/\sqrt{h_{11,t}}$		$\varepsilon_{2,t}/\sqrt{h_{22,t}}$	
Q(6)	4.19449	[0.65037]	1.18848	[0.97745]
$Q^2(6)$	5.52306	[0.47867]	2.06753	[0.91338]
ARCH(6)	1.01880	[0.98489]	3.65743	[0.72292]

Notes: Q(6) and $Q^2(6)$ are Ljung–Box tests for sixth order serial correlation in the standardized residuals and squared residuals. ARCH(6) is Engle's test for sixth order ARCH, distributed as $\chi^2(6)$. The *p*-value of these tests are displayed as [.].

Panel A indicates that, during the total period, the S&P500 volatility is directly affected by its own volatility ($h_{1,1}$) but not by the IBEX35 volatility ($h_{2,2}$). Our findings suggest that the S&P500 volatility is affected by its own shocks (ε_1^2) and indirectly by the IBEX35 shocks ($\varepsilon_1\varepsilon_2$). Finally, the coefficient for its own asymmetric term (η_1^2) is significant, indicating that the negative shocks on the S&P500 returns affect more its volatility than the positive shocks. Thus, our findings suggest that S&P500 volatility is directly affected by its own events.

The behavior of the IBEX35 volatility does not differ much from that of the S&P500 during the total period. The IBEX35 volatility is affected by its own volatility ($h_{2,2}$), but not by the S&P500 volatility. Interestingly, the IBEX35 volatility is affected by the S&P500 shocks (ε_1^2) but not by its own shocks

Table 16.6 Results of the linearized multivariate BEKK model

Panel A: total period									
S&P500 conditional variance equation									
$h_{11,t} = 1,32 \times 10^{-6}$	$+0,9023\ h_{11,t-1}$	$+0,0603\ h_{12,t-1}$	$+0,0006\ h_{22,t-1}$	$+0,02411\varepsilon^2_{1,t-1}$	$-0,069\varepsilon_{1,t-1}\varepsilon_{2,t-1}$	$+0,0122\varepsilon^2_{2,t-1}$	$+0,1009\eta^2_{1,t-1}$	$+0,0215\eta_{1,t-1}\eta_{2,t-1}$	$+0,0012\eta^2_{2,t-1}$
$2,35 \times 10^{-6}$	0,0044	0,0188	0,0006	0,0060	0,0039	0,0087	0,0123	0,0565	0,0012
(0,5617)	(203,15)	(3,2127)	(1,0312)	(3,9983)	(−17,711)	(1,4079)	(8,1658)	(0,3813)	(0,9457)
IBEX35 conditional variance equation									
$h_{22,t} = 2,46 \times 10^{-7}$	$+0,0010h_{11,t-1}$	$+0,0599h_{12,t-1}$	$+0,8903h_{22,t-1}$	$+0,0494\varepsilon^2_{1,t-1}$	$-0,026\varepsilon_{1,t-1}\varepsilon_{2,t-1}$	$+0,0035\varepsilon^2_{2,t-1}$	$+0,0003\eta^2_{1,t-1}$	$+0,0100\eta_{1,t-1}\eta_{2,t-1}$	$+0,0712\eta^2_{2,t-1}$
$1,01 \times 10^{-7}$	0,0006	0,0183	0,0134	0,0102	0,0142	0,0042	0,0015	0,0210	0,0129
(2,4325)	(1,5978)	(3,2716)	(66,158)	(4,8342)	(−1,8657)	(0,8499)	(0,2324)	(0,4769)	(5,5200)
Panel B: pre-September 11 period									
S&P500 conditional variance equation									
$h_{11,t} = 7,46 \times 10^{-6}$	$+0,7962\ h_{11,t-1}$	$+0,0691\ h_{12,t-1}$	$+0,0099\ h_{22,t-1}$	$+0,0044\varepsilon^2_{1,t-1}$	$-0,0277\varepsilon_{1,t-1}\varepsilon_{2,t-1}$	$+0,0152\varepsilon^2_{2,t-1}$	$+0,1986\eta^2_{1,t-1}$	$+0,1845\eta_{1,t-1}\eta_{2,t-1}$	$+0,0428\eta^2_{2,t-1}$
$2,80 \times 10^{-6}$	0,0261	0,0509	0,0006	0,0115	0,0393	0,0225	0,0506	0,1133	0,0257
(2,6631)	(30,547)	(1,3559)	(1,5813)	(0,3882)	(−0,7039)	(0,6770)	(3,9265)	(1,6279)	(1,6643)
IBEX35 conditional variance equation									
$h_{22,t} = 2,20 \times 10^{-5}$	$+0,0015\ h_{11,t-1}$	$+0,0666\ h_{12,t-1}$	$+0,7402\ h_{22,t-1}$	$+0,0430\varepsilon^2_{1,t-1}$	$-0,0110\varepsilon_{1,t-1}\varepsilon_{2,t-1}$	$+0,0007\varepsilon^2_{2,t-1}$	$+0,0002\eta^2_{1,t-1}$	$+-0,0083\eta_{1,t-1}\eta_{2,t-1}$	$+0,0751\eta^2_{2,t-1}$
$1,50 \times 10^{-5}$	0,0022	0,0474	0,1459	0,0320	0,0329	0,0041	0,0023	0,0432	0,0451
(1,4693)	(0,6793)	(1,4054)	(5,0739)	(1,3429)	(−0,3309)	(0,1675)	(0,0977)	(−0,1912)	(1,6656)

Continued

Table 16.6 Continued

Panel C: post-September 11 period

S&P500 conditional variance equation

$h_{11,t} =$									
$1{,}82 \times 10^{-6}$	$+0{,}9110\ h_{11,t-1}$	$+0{,}0569\ h_{12,t-1}$	$+0{,}0002\ h_{22,t-1}$	$+0{,}0188\varepsilon^2_{1,t-1}$	$-0{,}0494\varepsilon_{1,t-1}\varepsilon_{2,t-1}$	$+0{,}0193\varepsilon^2_{2,t-1}$	$+0{,}0875\eta^2_{1,t-1}$	$+0{,}0172\eta_{1,t-1}\eta_{2,t-1}$	$+0{,}0008\eta^2_{2,t-1}$
$6{,}05 \times 10^{-7}$	0,0018	0,0043	0,00003	0,0036	0,0049	0,0055	0,0042	0,0544	0,0013
(3,0171)	(517,13)	(13,0777)	(0,6622)	(5,2752)	(−10,069)	(3,4925)	(20,966)	(0,3165)	(0,6577)

IBEX35 conditional variance equation

$h_{22,t} =$									
$1{,}24 \times 10^{-6}$	$+0{,}0009\ h_{11,t-1}$	$+0{,}0571\ h_{12,t-1}$	$+0{,}9186\ h_{22,t-1}$	$+0{,}0325\varepsilon^2_{1,t-1}$	$-0{,}01596\varepsilon_{1,t-1}\varepsilon_{2,t-1}$	$+0{,}0019\varepsilon^2_{2,t-1}$	$+0{,}0003\eta^2_{1,t-1}$	$-0{,}0066\eta_{1,t-1}\eta_{2,t-1}$	$+0{,}0319\eta^2_{2,t-1}$
$4{,}82 \times 10^{-7}$	0,0001	0,0043	0,0013	0,0037	0,0099	0,0025	0,0005	0,0049	0,0032
(2,5799)	(6,6162)	(13,1568)	(709,68)	(8,6504)	(1,6033)	(0,7843)	(0,6382)	(−1,3466)	(9,8872)

Notes: h_{11} and h_{22} denote the conditional variance for the S&P500 and IBEX35 return series, respectively. Below the estimated coefficients are the standard errors, with the corresponding *t*-values given in parentheses.

The expected value is obtained taking expectations to the non-linear functions, therefore involving the estimated variance-covariance matrix of the parameters. In order to calculate the standard errors, the function must be linearized using first-order Taylor series expansion. This is sometimes called the "delta method". When a variable Y is a function of a variable X, i.e., $Y = F(X)$, the delta method allows us to obtain approximate formulation of the variance of Y if: (1) Y is differentiable with respect to X and (2) the variance of X is known. Therefore:

$$V(Y) \approx (\Delta Y)^2 \approx \left(\frac{\partial Y}{\partial X}\right)^2 (\Delta X)^2 \approx \left(\frac{\partial Y}{\partial X}\right)^2 V(X)$$

When a variable Y is a function of variables X and Z in the form of $Y = F(X, Z)$, we can obtain approximate formulation of the variance of Y if: (1) Y is differentiable with respect to X and Z and (2) the variance of X and Z and the covariance between X and Z are known. This is:

$$V(Y) \approx \left(\frac{\partial Y}{\partial X}\right)^2 V(X) + \left(\frac{\partial Y}{\partial Z}\right)^2 V(Z) + 2\left(\frac{\partial Y}{\partial X}\right)\left(\frac{\partial Y}{\partial Z}\right) Cov(X, Z)$$

Once the variances are calculated it is straightforward to calculate the standard errors.

(ε_2^2). Moreover, the coefficient for its own asymmetric term (η_2^2) is significant, indicating that the negative shocks on the IBEX35 returns affect more its volatility than the positive shocks.

Panel B indicates that, during the pre-September 11 period, the S&P500 volatility is directly affected by its own volatility ($h_{1,1}$) but not by the IBEX35 volatility. The results show that the S&P500 volatility is neither affected by its own shocks (ε_1^2) nor by the shocks on the IBEX35 (ε_2^2). Thus, our findings suggest that S&P500 volatility is only affected by its own volatility.

The equation for the IBEX35 in Panel B, shows that the index volatility is only affected by its own volatility ($h_{2,2}$). Therefore, higher levels of conditional volatility in the past are associated with higher conditional volatility in the current period. Our results indicate that the IBEX35 volatility is not affected by shocks affecting neither the IBEX35 returns (ε_2^2) nor the S&P500 returns (ε_1^2).

Panel C displays the result for the post-September 11 period. The S&P500 volatility is directly affected by its own volatility ($h_{1,1}$) but not by the IBEX35 volatility ($h_{2,2}$). Our findings suggest that the S&P500 volatility is affected by its own shocks (ε_1^2) and the IBEX35 shocks (ε_2^2). Finally, the coefficient for its own asymmetric term (η_1^2) is significant, indicating that the negative shocks on the S&P500 returns affect more its volatility than the positive shocks.

The IBEX35 volatility is affected by its own volatility ($h_{2,2}$) and by the S&P500 volatility ($h_{1,1}$). Again, the IBEX35 volatility is not affected by its own shocks (ε_2^2) but it is affected by the S&P500 shocks (ε_1^2). Finally, the coefficient for its own asymmetric term (η_2^2) is significant, indicating that the negative shocks on the IBEX35 returns affect more its volatility than the positive shocks.

If we compare the results for the pre-September 11 and post-September 11 periods, it can be observed that, before September 11, both variances (S&P500 and IBEX35) were only affected by their own past volatilities. However, after September 11, the S&P500 volatility is also affected by its own shocks and the IBEX35 shocks. On the other hand, the IBEX35 volatility is also affected by the S&P500 shocks and the S&P500 variance. These results show that, after September 11, there has been an increase in the volatility transmission between the USA and Spain.

16.4.2 Asymmetric volatility impulse response functions (AVIRF)

Figures 16.2 to 16.7 present the AVIRFs computed following Lin (1997) and Meneu and Torró (2003). When unexpected shocks are positive (Figures 16.2, 16.4 and 16.6), graphical analysis shows that in the post-September 11 and total period, there exists a significant volatility spillover from the S&P500 to the IBEX35 (about 5% of the shock, Figures 16.2-C and 16.6-C). Thus, there is no significant volatility spillover from the S&P500 to the IBEX35 before

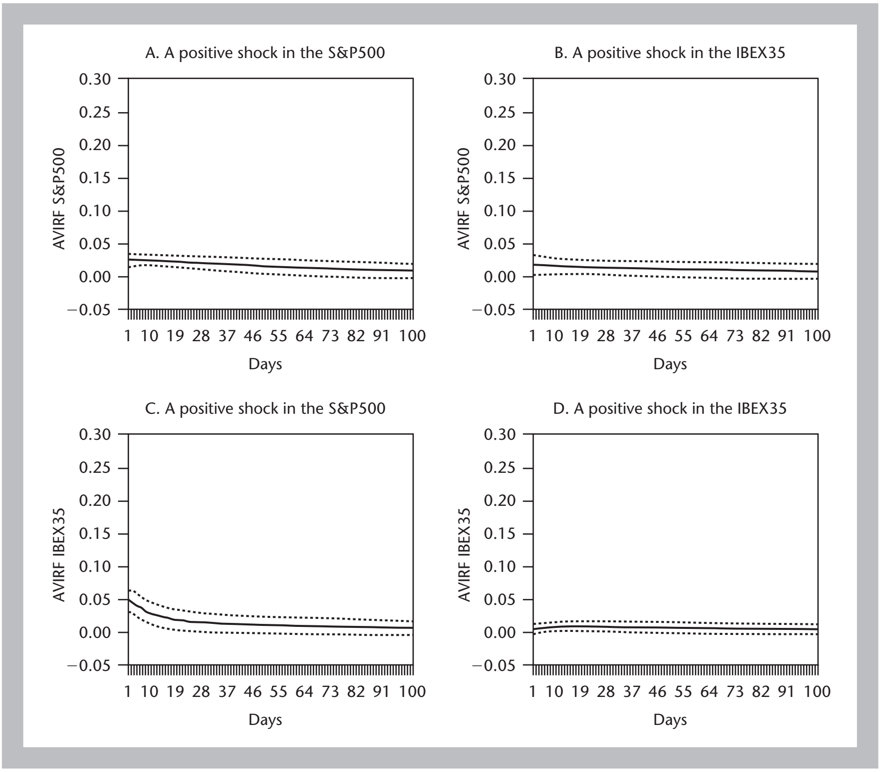


Figure 16.2 AVIRF to positive unexpected shocks from the VAR-asymmetric BEKK. Total period (dashed lines display the 90% confidence interval)

September 11, as suggested by the wide confidence intervals, but it becomes significant after the terrorist attack. However, the reverse is only detected in the post-September 11, although the effect is almost imperceptible (Figures 16.2-B, 16.4-B and 16.6-B).

If unexpected shocks are negative, Figures 16.3, 16.5 and 16.7 show that there is only significant volatility spillover from the S&P500 to the IBEX35. As in the case of a positive shock, it becomes significant only after September 11. Negative shocks in the IBEX35 have an important effect on its own volatility that takes about 40 days to be absorbed when the whole sample is analyzed (Figure 16.3-D). This effect also becomes significant only after September 11 (Figures 16.3-D and 16.7-D) but not before (Figure 16.5-D). Negative shocks in the S&P500 also have an important effect on its own volatility that takes more than 100 days to be absorbed in all analyzed periods (Figures 16.3-A, 16.5-A and 16.7-A).

By comparing positive and negative shocks coming from the IBEX35, we can observe that they have a different impact on its own volatility.

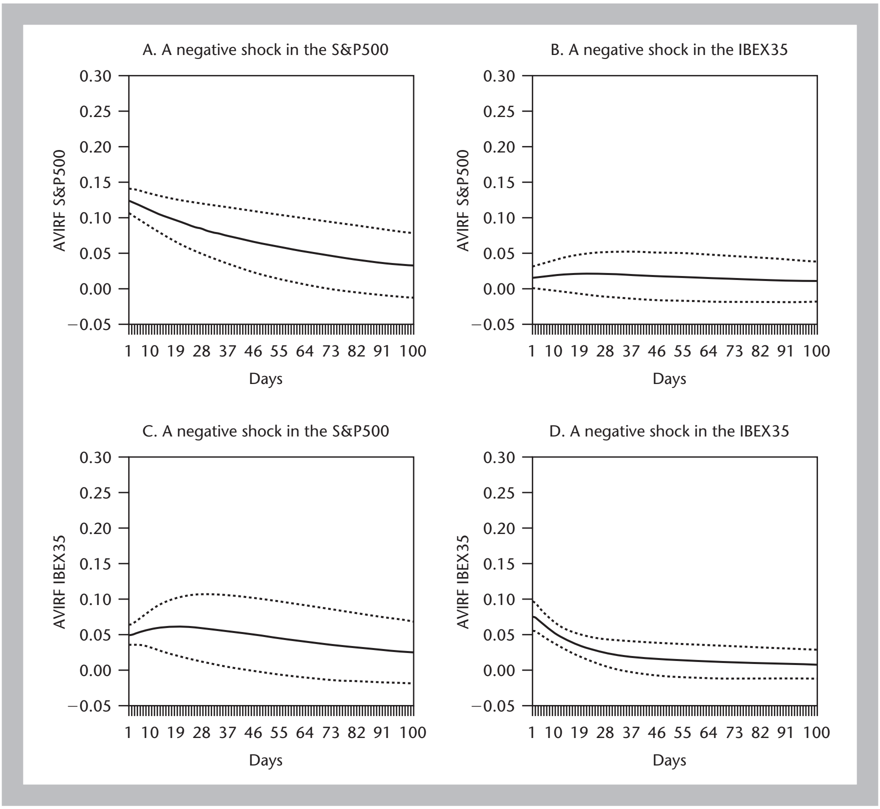


Figure 16.3 AVIRF to negative unexpected shocks from the VAR-asymmetric BEKK. Total period (dashed lines display the 90% confidence interval)

There is no significant impact coming from positive shocks (Figures 16.2-D, 16.4-D and 16.6-D) and, after September 11, negative shocks have a significant effect that takes a very long time to die out due to its persistence. This asymmetric effect can explain why, when estimation results where analyzed in the previous section, we concluded that the IBEX35 was not affected by its own unexpected shocks. This occurs when unexpected shocks are taken as a whole. In the case of the S&P500, there also exists an asymmetric effect, which is observed in all analyzed periods.

The AVIRF figures depend on the estimated coefficients coming from matrix G. In all periods, the figures confirm the existence of an asymmetric effect of positive and negative shocks on own conditional variances (related to coefficients g_{11} and g_{22}). A negative shock has a stronger impact than a positive one. However, this asymmetric effect does not exist in the case of shocks coming from another market (coefficients g_{12} and g_{21}).

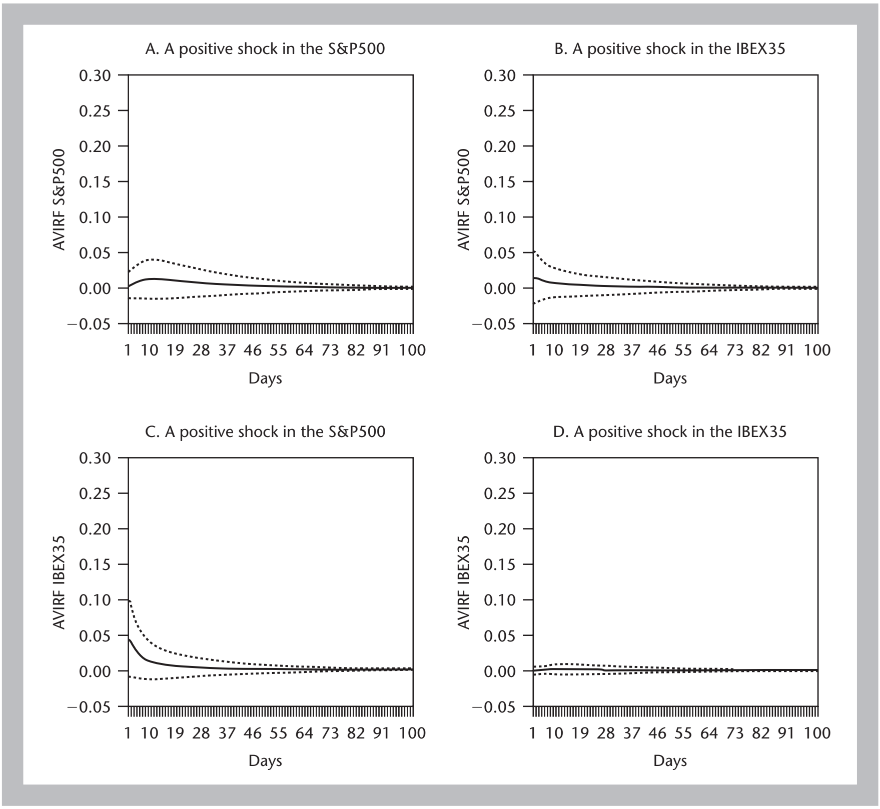


Figure 16.4 AVIRF to positive unexpected shocks from the VAR-asymmetric BEKK. Pre-September 11 period (dashed lines display the 90% confidence interval)

In general, the most appealing results are: (1) conditional variances are more sensitive to negative than to positive shocks; (2) unexpected shocks in S&P500 have more impact on the whole covariance matrix than unexpected shocks in IBEX35; (3) IBEX35 variance is more sensitive to any shock than S&P500 variance; (4) before September 11, there are no significant volatility spillovers in any direction and, after the terrorist attack, there is unidirectional variance causality from the S&P500 to the IBEX35; (4) generally, significant shocks take a long time to die out; and (5) the IBEX35 volatility has an overshooting reaction to a negative shock in the S&P500 when the whole sample is analyzed. This could be due to the high persistence of the IBEX35 or to an overreaction in the Spanish market because, at 16:00, Spanish investors will only have one and a half hours left to react before the market closes at 17:30.

Therefore, these results add evidence in favor of the hypothesis of unidirectional variance causality from the S&P500 to the IBEX35. The AVIRF

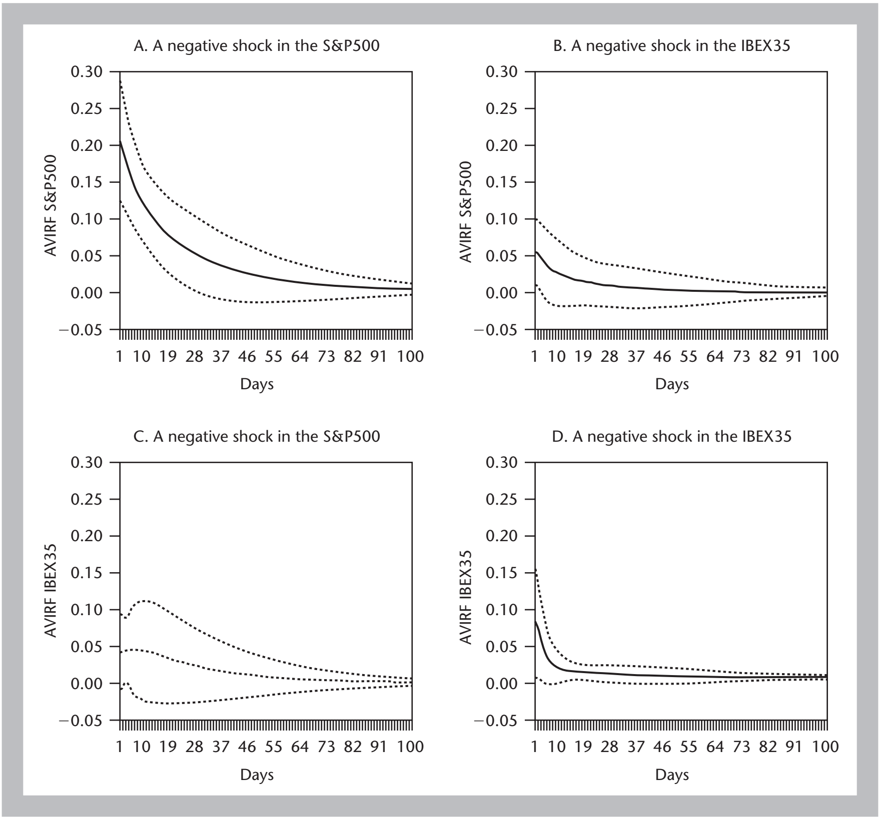


Figure 16.5 AVIRF to negative unexpected shocks from the VAR-asymmetric BEKK. Pre-September 11 period (dashed lines display the 90% confidence interval)

analysis shows that, after September 11, any volatility shock coming from the S&P500 is directly affecting the IBEX35 but the reverse is not true in any period (it exists in the case of positive shocks in the post-September 11 period, but the effect is hardly noticeable). Moreover, a negative shock in the S&P500 is more persistent than a positive shock. Therefore, it can be said that the main source of information comes from negative unexpected returns arising from the S&P500 and it spreads into the Spanish market.

16.5 CONCLUSION

The main objective of this study has been to analyze whether volatility transmission patterns between the US and Spanish stock markets have changed

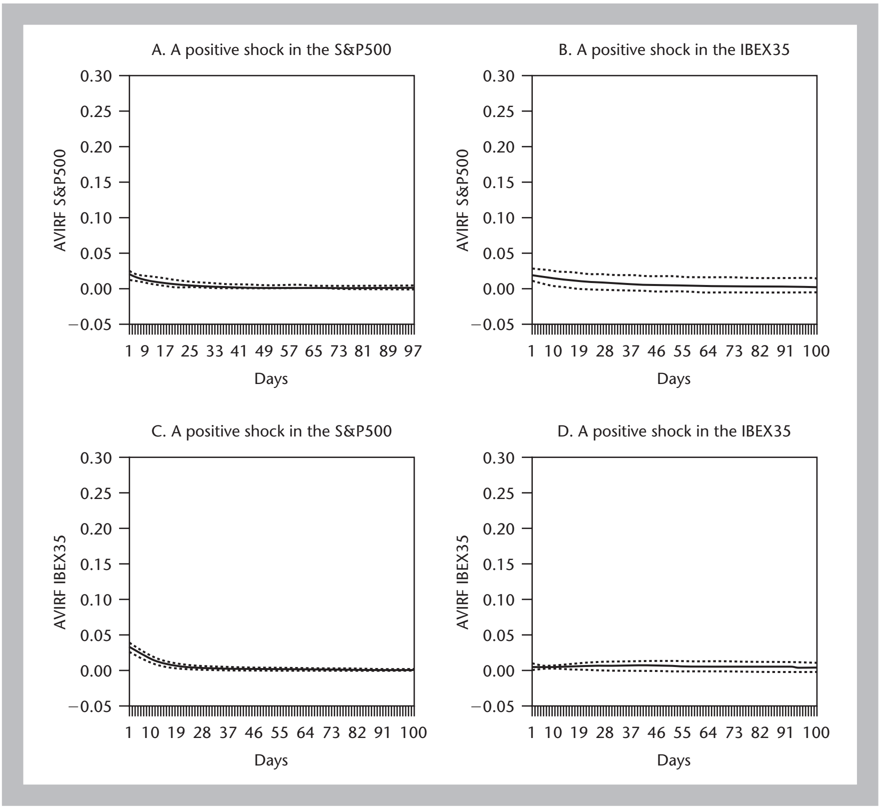


Figure 16.6 AVIRF to positive unexpected shocks from the VAR-asymmetric BEKK. Post-September 11 period (dashed lines display the 90% confidence interval)

after 11 September 2001. In order to do this, we have used a multivariate GARCH model and taken into account both the asymmetric volatility phenomenon and the non-synchronous trading problem. In particular, an asymmetric VAR-BEKK model has been estimated both before and after the terrorist attack, with daily stock market prices recorded at 16:00 Spanish time for the USA (S&P500 index) and Spain (IBEX35 index).

We also present a complementary analysis, the Asymmetric Volatility Impulse Response Functions (AVIRF), which distinguishes effects coming from a positive shock from those coming from a negative shock.

The results confirm that there exist own asymmetric volatility effects in both markets and that volatility transmission from the USA to Spain increased after September 11. Before the event, volatilities in the S&P500 and IBEX35 were only affected by their own past volatilities. However, after September 11, the IBEX35 volatility becomes affected by volatility and shocks

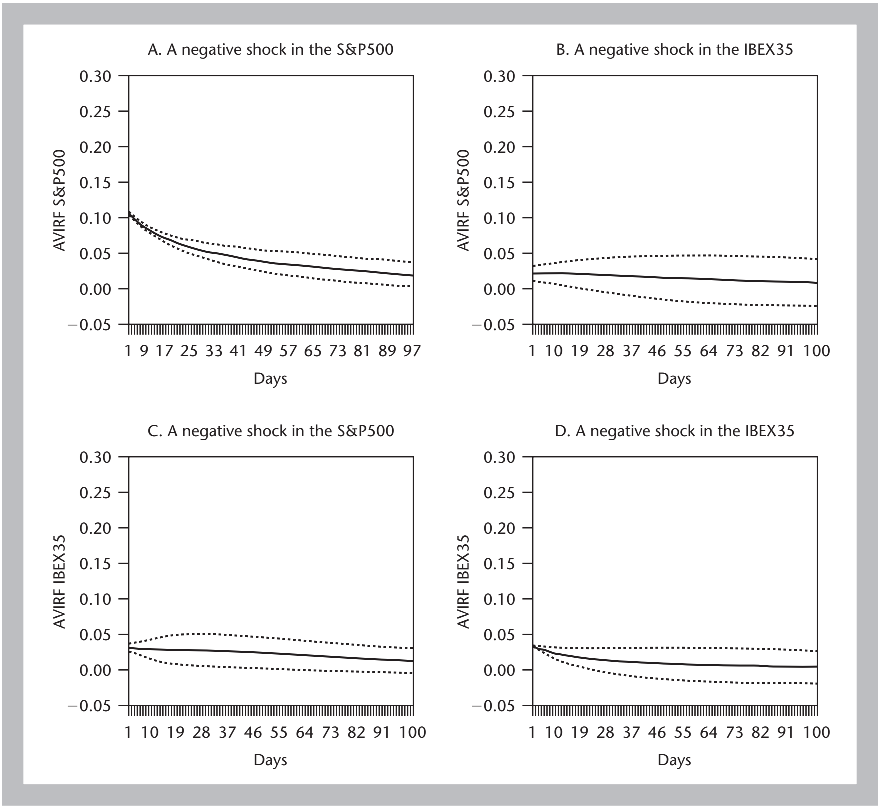


Figure 16.7 AVIRF to negative unexpected shocks from the VAR-asymmetric BEKK. Post-September 11 period (dashed lines display the 90% confidence interval)

in the S&P500, and the S&P500 volatility becomes affected by its own shocks. Therefore, these results show that, after the terrorist attack, there has been an increase in the volatility transmission between the USA and Spain. These results agree with Hon, Strauss and Yong (2004), suggesting that international stock markets respond more closely to US stock market shocks after the crisis than before.

NOTES

1. See Koutmous and Booth (1995), Karolyi (1995), Karolyi and Stulz (1996), Darbar and Deb (1997), Ramchand and Susmel (1998), Brooks and Henry (2000), Longin and Solnik (2001), Martens and Poon (2001) and Bera and Kim (2002).
2. A special issue of the *Economic Policy Review* of the Federal Reserve Bank of New York (2002, vol. 8, no. 2) analyzes general economic consequences of September 11.

A special issue of the *Journal of Risk and Uncertainty* (2003, vol. 26, nos 2/3) deals with the risks of terrorism with a special focus on September 11. A special issue of the *European Journal of Political Economy* (2004, vol. 20, issue 2) deals with the economic consequences of terror.

3. Before 17 January 2000 continuous trading was from 9:30 to 17:00 Spanish Time; therefore, the overlapping trading period was just one hour and a half. Additionally, the beginning of the daylight saving time in October for Spain and the USA coincides. However, the end of this daylight saving time is the first Sunday of April in the USA and the last Sunday of March in Spain. Hence, during the last week of March, the markets overlap only for one hour.

REFERENCES

Baba, Y., Engle, R.F., Kraft, D.F. and Kroner, K.F. (1989) “Multivariate Simultaneous Generalized ARCH”, Mimeo, (San Diego, CA: University of California at San Diego).

Baele, L. (2005) “Volatility Spillover Effects in European Equity Markets”, *Journal of Financial and Quantitative Analysis*, 40(2): 373–401.

Bera, A.K. and Kim, S. (2002) “Testing Constancy of Correlation and Other Specifications of the BGARCH Model with an Application to International Equity Returns”, *Journal of Empirical Finance*, 9(2): 171–95.

Bernanke, B. (1986) “Alternative Explanations of the Money–Income Correlation”, *Carnegie-Rochester Conference Series on Public Policy*, 25: 49–100.

Bollerslev, T. and Wooldridge, J.M. (1992) “Quasi-maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances”, *Econometric Reviews*, 11(2): 143–72.

Booth, G.G., Martikainen, T. and Tse, Y. (1997) “Price and Volatility Spillovers in Scandinavian Stock Markets”, *Journal of Banking and Finance*, 21(5): 811–23.

Brooks, C. and Henry, O.T. (2000) “Linear and Non-linear Transmission of Equity Return Volatility: Evidence from the U.S., Japan and Australia”, *Economic Modelling*, 17(4): 497–513.

Cifarelli, P. and Paladino, G. (2005) “Volatility Linkages Across Three Major Equity Markets: A Financial Arbitrage Approach”, *Journal of International Money and Finance*, 24(3): 413–39.

Darbar, S.M. and Deb, P. (1997) “Co-movements in International Equity Markets”, *Journal of Financial Research*, 20(3): 305–22.

Dickey, D.A. and Fuller, W.A. (1981) “Likelihood Ratios Statistics for Autoregressive Time Series with a Unit Root”, *Econometrica*, 49(4): 1057–72.

Engle, R.F. and Kroner, K.F. (1995) “Multivariate Simultaneous Generalized ARCH”, *Econometric Theory*, 11(1): 122–50.

Engle, R.F., Ito, T. and Ln, W.L. (1990) “Meteor Showers or Heat Waves? Heteroskedastic Intra-daily Volatility in the Foreign Exchange Market”, *Econometrica*, 58(3): 525–42.

Fernández, A. and Aragó, V. (2003) “European Volatility Transmission with Structural Changes in Variance”, Working paper, Universitat Jaume I, Castelló, Spain.

Glaser, M. and Weber, M. (2005) “September 11 and Stock Returns Expectations of Individual Investors”, *Review of Finance*, 9(2): 243–79.

Glosten, L.R., Jagannathan, R. and Runkle, D.E. (1993) “On the Relationship Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks”, *Journal of Finance*, 48(5): 1779–801.

Hamao, Y., Masulis, R.W. and Ng, V. (1990) "Correlations in Price Changes and Volatility Across International Stock Markets", *Review of Financial Studies*, 3(2): 281–307.

Hon, M.T., Strauss, J. and Yong, S. (2004) "Contagion in Financial Markets after September 11: Myth or Reality?", *Journal of Financial Research*, 27(1): 95–114.

Isakov, D. and Perignon, C. (2001) "On the Dynamic Interdependence of International Stock Markets: A Swiss Perspective", *Swiss Journal of Economics and Statistics*, 136(2): 1769–88.

Ito, H. and Lee, D. (2005) "Assessing the Impact of September 11 Terrorist Attacks on U.S. Airline Demand", *Journal of Economics and Business*, 57(1): 75–95.

Johansen, S. (1988) "Statistical Analysis of Cointegration Vectors", *Journal of Economic Dynamics and Control*, 12(2–3): 231–54.

Karolyi, G.A. (1995) "A Multivariate GARCH Model of International Transmissions of Stock Returns and Volatility: The Case of the United States and Canada", *Journal of Business and Economic Statistics*, 13(1): 11–25.

Karolyi, G.A. and Stulz, R.M. (1996) "Why Do Markets Move Together? An Investigation of US–Japan Stock Return Comovements", *Journal of Finance*, 51(3): 951–86.

Kearney, C. and Patton, A.J. (2000) "Multivariate GARCH Modeling of Exchange Rate Volatility Transmission in the European Monetary System", *Financial Review*, 35(1): 29–48.

King, M.A. and Wadhwani, S. (1990) "Transmission of Volatility Between Stock Markets", *Review of Financial Studies*, 3(1): 5–33.

Koutmos, G. and Booth, G. (1995) "Asymmetric Volatility Transmission in International Stock Markets", *Journal of International Money and Finance*, 14(6): 747–62.

Kroner, K.F. and Ng, V.K. (1998) "Modeling Asymmetric Comovements of Asset Returns", *Review of Financial Studies*, 11(4): 817–44.

Lin, W.L. (1997) "Impulse Response Function for Conditional Volatility in GARCH Models", *Journal of Business and Economic Statistics*, 15(1): 15–25.

Longin, F. and Solnik, B. (1995) "Is the International Correlation of Equity Returns Constant: 1960–1990?", *Journal of International Money and Finance*, 14(1): 3–26.

Longin, F. and Solnik, B. (2001) "Extreme Correlation of International Equity Markets", *Journal of Finance*, 56(2): 649–76.

MacKinnon, J.G. (1991) "Critical Values for Cointegration Tests", in R.F. Engle and C.W.J. Granger (eds), *Long-Run Economic Relationships: Readings in Cointegration*, Chapter 13 (Oxford: Oxford University Press).

Martens, M. and Poon, S.H. (2001) "Returns Synchronization and Daily Correlation Dynamics Between International Stock Markets", *Journal of Banking and Finance*, 25(10): 1805–27.

Meneu, V. and Torró, H. (2003) "Asymmetric Covariance in Spot-Futures Markets", *Journal of Futures Markets*, 23(11): 1019–43.

Ng, A. (2000) "Volatility Spillover Effects from Japan and the US to the Pacific-Basin", *Journal of International Money and Finance*, 19(2): 207–33.

Osterwald-Lenum, M. (1992) "A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics", *Oxford Bulletin of Economics and Statistics*, 54(3): 461–72.

Peña, J.I. (1992) "On Meteor Showers in Stock Markets: New York vs Madrid", *Investigaciones Económicas*, 16(2): 225–34.

Perez, J.V. and Torra, S. (1995) "Transmisión Internacional de las Rentabilidades y Volatilidades entre NYSE e IBEX35", *Cuadernos de Economía*, 23(65): 83–101.

Philips, P.C.B. and Perron, P. (1988) "Testing for a Unit Root in Time Series Regression", *Biometrica*, 75(2): 335–46.

Poteshman, A.M. (2005) "Unusual Option Market Activity and the Terrorist Attacks of September 11, 2001", *Journal of Business*.

Ramchand, L. and Susmel, R. (1998) "Volatility and Cross Correlation Across Major Stock Markets", *Journal of Empirical Finance*, 5(4): 397–416.
Rigobon, R. (2003) "On the Measurement of the International Propagation of Shocks: Is the Transmission Stable?", *Journal of International Economics*, 61(2): 261–83.
Sims, C.A. (1986) "Are Forecasting Models Usable for Policy Analysis?", *Federal Reserve Bank of Minneapolis Quarterly Review*, 10(1): 2–16.
Susmel, R. and Engle, R.F. (1994) "Hourly Volatility Spillovers Between International Equity Markets", *Journal of International Money and Finance*, 13(1): 3–25.
Wongswan, J. (2003) "Transmission of Information Across International Equity Markets", *Federal Reserve – International Finance Discussion Papers*, No. 759.

CHAPTER 17

Large and Small Cap Stocks in Europe: Covariance Asymmetry, Volatility Spillovers and Beta Estimates

*Helena Chuliá and Hipòlit Torró**

17.1 INTRODUCTION

Several studies show that small cap returns tend to behave differently from large cap returns (Banz, 1981; Chan and Chen, 1991). This fact suggests that diversifying into small cap stocks might improve portfolio performance. In fact, the main empirical evidence on small cap returns shows that small caps distinguish themselves from large caps due to economic and market related characteristics (for a literature review on this topic see Petrella, 2005).

Moreover, a large number of papers have shown that returns of large capitalization stocks can be used to predict the returns of smaller stocks, but not vice versa (Lo and Mackinlay, 1990). This asymmetry in the predictability of mean returns does not necessarily imply that all the information is

*Financial support from CICYT project BEC2003-09607-C04-04, and project GV04A/153 from Generalitat Valenciana, the Instituto Valenciano de Investigaciones Económicas (IVIE) and an FPU grant from the Ministerio de Educación y Ciencia are gratefully acknowledged.

transmitted from large to small capitalization companies but suggests that there are differences in the dynamics of stocks prices of firms with different market value (Conrad, Gultekin and Kaul, 1991). In addition, Ross (1989) showed that variance changes are directly related to the rate of information flow. Therefore, one way of studying how information spreads between large and small firms take place is studying their volatility spillovers.

Most studies that analyse the effect of news on second moments focus on the US, Australian and Japanese markets. They conclude that volatility surprises for large firms can be used to predict volatility of small firms but not vice versa. Conrad, Gultekin and Kaul (1991), find that shocks to large firm returns are important to the future dynamics of their own volatility as well as the volatility of small firm returns in the US market. Conversely, shocks to small firms have no impact on the behavior of the volatility of large firms. Hendry and Sharma (1999) obtain similar results for the Australian market, and Kroner and Ng (1998) confirm the conclusion of Conrad, Gultekin and Kaul (1991) in a more general context.

When the dynamic relationships between volatility of large and small firms returns are studied, it is necessary to consider asymmetric volatility and covariance. The first one refers to the empirical evidence according to which a negative return shock (unexpected drop in the value of the stock) generates an increase in volatility higher than a positive return shock (unexpected increase in the value of the stock) of the same size (for a literature review on asymmetric volatility see Bekaert and Wu, 2000). Asymmetric covariance refers to the empirical evidence according to which covariance between market and stock returns responds more after negative than after positive market shocks.

In the financial literature, two explanations of the asymmetries in equity markets have been put forward. The first one is based on the leverage effect hypothesis. According to this explanation, a drop in the value of the stock (negative return) increases financial leverage, which makes the stock riskier and increases its volatility (Black, 1976; Christie, 1982). The second explanation is known as the volatility feedback hypothesis. This explanation maintains that the asymmetry in volatility responds to the fact that returns could simply reflect the existence of time-varying risk premiums. If volatility is priced an anticipated increase in volatility raises the required return on equity, leading to an immediate stock price decline (Campbell and Hentschel, 1992; Pindyck, 1984; French, Schewert and Stambaugh, 1987). This hypothesis relies on two basic tenets. Firstly, volatility is persistent and secondly, there exists a positive inter-temporal relation between expected returns and conditional variances.

Consequently, the causality of the asymmetry in equity markets is different. According to the leverage effect, the return shocks lead to changes in conditional volatility; whereas according to the volatility feedback hypothesis return shocks are caused by changes in conditional volatility.

Which effect, leverage effect or volatility feedback effect, is the main determinant of asymmetric volatility, remains an open question. Those studies that focus their analysis on the leverage hypothesis (Christie, 1982 and Schwert, 1989) show that this effect is too small to explain the full asymmetry. On the other hand, authors like Braun *et al*. (1995), Bekaert and Wu (2000) and Wu (2001) find clear evidence in favor of the volatility feedback effect as the main cause of the asymmetric behavior.

Our main contributions to the research in this field are threefold. Firstly, we analyse volatility spillovers between large and small firms in the French, German and British stockmarkets since the existing empirical studies have focused in the American, Japanese and Australian equity markets. In order to do so, a conditional CAPM with an asymmetric multivariate GARCH-in-mean covariance structure is used. Results show that there exist bidirectional volatility spillovers between both types of companies. Secondly, we explore the volatility feedback hypothesis as a possible explanation of asymmetric volatility in stock returns, finding significant evidence for this hypothesis. Finally, the study uncovers that conditional beta coefficient estimates within the used model are insensitive to sign and size asymmetries in the unexpected shock returns but unconditional beta estimates have a significant specification error.

The remainder of the chapter is organized as follows. Section 17.2 formulates the empirical model, while section 17.3 presents the data. Section 17.4 discusses the empirical results; section 17.5 shows an analysis of asymmetries; section 17.6 analyses volatility spillovers between large and small firms; and section 17.7 summarizes the results.

17.2 THE ECONOMETRIC FRAMEWORK

Following Bekaert and Wu (2000), in the present study a conditional version of the CAPM is used to examine the interaction between means and variances. In the assumed version of the conditional CAPM, excess returns of the large cap index is proportional to its conditional variance and excess returns of the small cap index is proportional to the conditional covariance between the small cap and the large cap index returns, being the proportion (constant) the same in both cases: the price of risk. Therefore, the conditional mean equations are defined as:

$$\begin{aligned} r_{1,t} - r^f_{t-1,t} &= \Upsilon\sigma^2_{1,t} + \varepsilon_{1,t} \\ r_{2,t} - r^f_{t-1,t} &= \Upsilon\sigma_{12,t} + \varepsilon_{2,t} \end{aligned} \tag{17.1}$$

where $r_{1,t}$ and $r_{2,t}$ refer to the large and small stock indexes respectively, Υ is the price of risk and $r^f_{t-1,t}$ is the risk-free interest rate known at time $t-1$.

Since the CAPM does not restrict the time variation in second moments, we employ a multivariate GARCH model. The three most widely used models are: (1) the VECH model proposed by Bollerslev *et al.* (1988), (2) the constant correlation model, CCORR, proposed by Bollerslev (1990), and (3) the BEKK model of Engle and Kroner (1995). Each model imposes different restrictions on the conditional covariance and gives very different variance and covariance estimates. More recently, Kroner and Ng (1998) have derived another multivariate GARCH model, the Asymmetric Dynamic Covariance Matrix model, ADC. This model encompasses the above models in the sense that, under certain restrictions, any particular model can be obtained. These authors introduce asymmetries following the Glosten, Jagannathan and Runkel (1993) approach.

The bivariate ADC can be written as:

$$\begin{aligned}\begin{bmatrix}\sigma_{1,t}^2 & \sigma_{12,t}\\ \cdot & \sigma_{2,t}^2\end{bmatrix} &= \begin{bmatrix}\sqrt{\theta_{11,t}} & 0\\ 0 & \sqrt{\theta_{22,t}}\end{bmatrix}\begin{bmatrix}1 & \rho_{12}\\ \rho_{12} & 1\end{bmatrix}\begin{bmatrix}\sqrt{\theta_{11,t}} & 0\\ 0 & \sqrt{\theta_{22,t}}\end{bmatrix}\\ &\quad + \begin{bmatrix}0 & \phi_{12}\\ \phi_{12} & 0\end{bmatrix}\circ\begin{bmatrix}\theta_{11,t} & \theta_{12,t}\\ \cdot & \theta_{22,t}\end{bmatrix}\\ &= \begin{bmatrix}\theta_{11,t} & \phi_{12}\theta_{12,t}+\rho_{12}\sqrt{\theta_{11,t}}\sqrt{\theta_{22,t}}\\ \cdot & \theta_{22,t}\end{bmatrix}\end{aligned} \qquad (17.2)$$

where

$$\begin{aligned}\begin{bmatrix}\theta_{11,t} & \theta_{12,t}\\ \cdot & \theta_{22,t}\end{bmatrix} &= \begin{bmatrix}\omega_{11} & \omega_{12}\\ \cdot & \omega_{22}\end{bmatrix} + \begin{bmatrix}b_{11} & b_{12}\\ b_{21} & b_{22}\end{bmatrix}'\begin{bmatrix}\sigma_{1,t-1}^2 & \sigma_{12,t-1}\\ \cdot & \sigma_{2,t-1}^2\end{bmatrix}\begin{bmatrix}b_{11} & b_{12}\\ b_{21} & b_{22}\end{bmatrix}\\ &\quad + \begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix}'\begin{bmatrix}\varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1}\\ \cdot & \varepsilon_{2,t-1}^2\end{bmatrix}\begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix}\\ &\quad + \begin{bmatrix}g_{11} & g_{12}\\ g_{21} & g_{22}\end{bmatrix}'\begin{bmatrix}\eta_{1,t-1}^2 & \eta_{1,t-1}\eta_{2,t-1}\\ \cdot & \eta_{2,t-1}^2\end{bmatrix}\begin{bmatrix}g_{11} & g_{12}\\ g_{21} & g_{22}\end{bmatrix}\end{aligned}$$

where $\circ$ is the Hadamard product operator (element-by-element matrix multiplication), ω_{ij}, b_{ij}, a_{ij} and g_{ij} for $i,j=1,2$ are parameters, $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are the unexpected shocks series, $\eta_{1,t}=\max[0, -\varepsilon_{1,t}]$ and $\eta_{2,t}=\max[0, -\varepsilon_{2,t}]$ are the Glosten, Jagannathan and Runkel (1993) dummy series collecting a negative asymmetry from the shocks and, finally, $\sigma_{ij,t}$ for all $i, j=1, 2$ are the conditional second moment series. The specification test proposed by Kroner and Ng (1998) is as follow: (1) if $\rho_{12}=b_{12}=b_{21}=a_{12}=a_{21}=g_{12}=g_{21}=0$, a restricted asymmetric VECH is obtained, (2) if $\phi_{12}=b_{12}=b_{21}=a_{12}=a_{21}=g_{12}=g_{21}=0$, the asymmetric CCORR model is derived; (3) if $\phi_{12}=1$ and $\rho_{12}=0$ the asymmetric BEKK model is obtained.

There are two main advantages in simultaneously modeling the conditional mean and variance. First, it is possible to quantify volatility cross effects between both types of firms. Second, it is possible to quantify the influence of the second moments cross effects in the expected returns.

17.3 DATA AND PRELIMINARY ANALYSIS

The data used for the French market come from Euronext, provided by the French Stock Exchange. It consists of daily closing values of the CAC40 index and the MIDCAC index. The data period runs from 2 January 1991 to 25 August 2004. The CAC40 index is calculated from a sample of 40 French stocks listed on the Monthly Settlement market. Component stocks are selected on the basis of their market capitalization and liquidity. The 100 MIDCAC components stocks are selected among the French companies listed on the “Premier Marche” or “Second Marche”, after eliminating: (1) the financial and property companies, (2) the 20% highest ant 20% lowest capitalized companies and (3) issues with a trading-day ratio below 70% (for example, traded on less than 175 of the total 250 trading days in the year).

In the case of the German market, the data has been provided by the German Stock Exchange. It consists of daily closing values of the DAX index and the SDAX index. The data period runs from 2 January 1991 to 25 August 2004. The DAX index is composed by the 30 largest and more actively traded German companies that are listed at the Frankfurt Stock Exchange. On the other hand, the SDAX comprises 50 continuously traded shares of small-sized companies.

The data used for the British market has been provided by the FTSE company. It consists of daily closing values of the FTSE 100 index and SMALL CAP index. The data period runs from 3 March 1993 to 25 August 2004. The FTSE index is composed of the 100 most capitalized British companies that are listed at the London Stock Exchange. On the other hand, the FTSE SMALL CAP index is composed of the next 350 issues that are ranked immediately below the FTSE 100 and FTSE 250. The minimum size of the component stocks of this index is reviewed annually.

Return series are obtained by taking first differences in the log prices of the three markets. The common test of unit roots (Dickey and Fuller, 1981; Philips and Perron, 1988) offered no doubt about this point. The accumulated weekly Treasury bill repo rate of each country is used as the risk free interest rate.

Despite the fact that all the indices are composed by the most liquid stocks traded in the Stock Exchange of each country, weekly frequency is used to overcome possible problems associated with thin trading. In order to transform daily data to weekly frequency, Wednesday closing values, or the previous day if the Wednesday is not a trading day, are used.

The use of market indices, instead of portfolios, provides two advantages to the practitioners of the market. Firstly, they can take signals directly from market indices quotations, therefore, it is not necessary to build portfolios and, secondly, the cost of implementing any potential trading rules can be reduced due to the existence of derivative contracts on the large stock index (Pardo and Torró, 2005).

Figure 17.1 displays the weekly evolution of the indices in the studied period and preliminary data analysis is presented in Tables 17.1 and 17.2.

Table 17.1 displays a summary of the principal statistics for the returns. It can be stated that all indices offer very similar statistics. All series present significant skewness except the CAC40 and the FTSE. Moreover, all present significant kurtosis and the Jarque-Bera statistic indicates that the hypothesis of normality is rejected for all indices. On the other hand, all series present significant autocorrelation and heteroskedasticity. Finally, although equality in means between large and small indices of each country cannot be rejected, the variances equality test is rejected.

Panel (A) in Table 17.2 displays returns, volatilities and correlation coefficients, year by year through the sample period for the French indices. Three facts can be highlighted from this table. First, there are five years (1991, 1992, 1995, 2000 and 2003) in which both indices offer a different sign return but the means equality hypothesis cannot be rejected. Secondly, for every year, except 2000, the CAC40 volatility is larger than the MIDCAC volatility. Another appealing fact is that except in two years (1998 and 2000), the hypothesis of variance equality is rejected. Finally, the correlation between both indices has dropped over time, becoming very small.

Panel (B) in Table 17.2 displays returns, volatilities and correlation coefficients, year by year through the sample period for the German indices. Three facts can be highlighted from this table. First, there are six years (1991, 1992, 1995, 1998, 2000 and 2004) in which both indices offer a different sign return but the means equality hypothesis cannot be rejected. Secondly, for every year, except 1996, the DAX is more volatile than the SDAX. Thirdly, except in three years (1993, 1996 and 2004), the hypothesis of variance equality is rejected. Finally, the correlation between both indices has decreased over time, becoming very small.

Panel (C) in Table 17.2 displays returns, volatilities and correlation coefficients, year by year through the sample period for the British indices. Four facts can be highlighted from this table. First, there are three years (1998, 2000 and 2004) in which both indices offer a different sign but the means equality hypothesis cannot be rejected. Secondly, for every year, except 2000 and 2001, the FTSE volatility is larger than the SMALL CAP volatility. Thirdly, at the beginning of the sample the hypothesis of variance equality is rejected but, at the end of the sample the hypothesis of variance equality cannot be rejected. Finally, the correlation between both indices has dropped over time.

Three facts must be highlighted from the previous analysis. Firstly, from Tables 17.1 and 17.2 it can be accepted that there exist significant differences

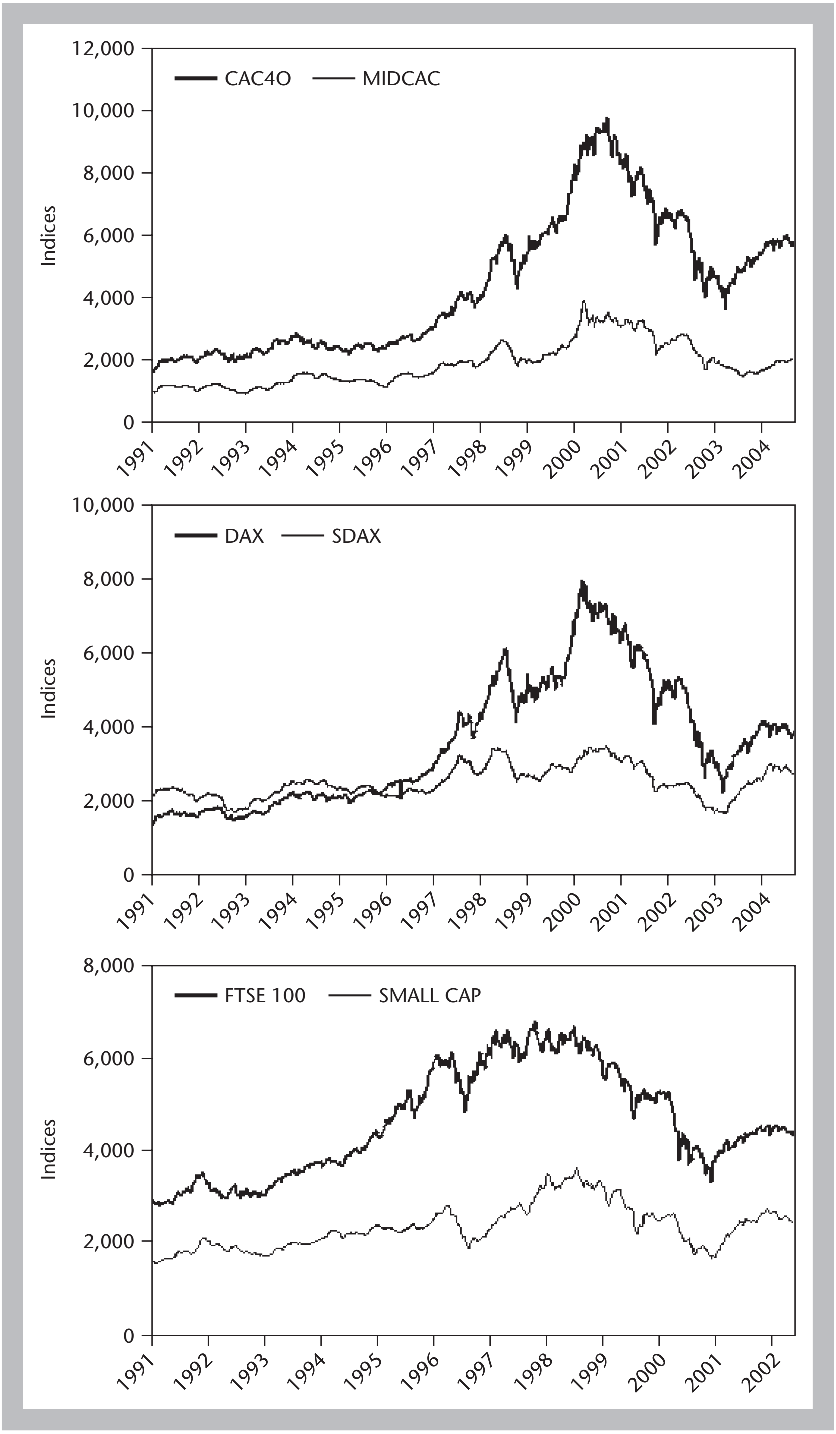


Figure 17.1 Evolution of the stock indices over the studied period

Table 17.1 Summary statistics for the data

Panel A: summary statistics for French indices series				
	Large cap index return		**Small cap index return**	
Mean	0.0018		0.0010	
Mean test	0.2613	[0.609]		
Variance	0.0009		0.0006	
Levene test	55.4551	[0.000]		
Skewness	−0.11225	[0.222]	−0.96712	[0.000]
Kurtosis	2.75102	[0.000]	6.42473	[0.000]
Normality	226.016	[0.000]	1335.55	[0.000]
Q(20)	46.2156	[0.000]	91.9528	[0.000]
Q^2(20)	287.913	[0.000]	140.8845	[0.000]
A(20)	281.982	[0.000]	171.7787	[0.000]
Panel B: summary statistics for German indices series				
	Large cap index return		**Small cap index return**	
Mean	0.0009		−0.0003	
Mean test	0.0407	[0.967]		
Variance	0.0012		0.0003	
Levene test	34.8751	[0.000]		
Skewness	−0.6176	[0.000]	−0.6286	[0.000]
Kurtosis	7.2543	[0.000]	2.1093	[0.000]
Normality	1608.72	[0.000]	179.138	[0.000]
Q(20)	43.3366	[0.002]	103.311	[0.000]
Q^2(20)	135.494	[0.000]	60.2548	[0.000]
A(20)	297.652	[0.000]	85.3739	[0.000]
Panel C: summary statistics for British indices series				
	Large cap index return		**Small cap index return**	
Mean	0.0007		0.0008	
Mean test	0.0787	[0.937]		
Variance	0.0005		0.0003	
Levene test	40.3885	[0.000]		
Skewness	0.14599	[0.147]	−0.93398	[0.000]
Kurtosis	3.61978	[0.000]	4.31934	[0.000]
Normality	325.854	[0.000]	547.1904	[0.000]

Continued

Table 17.1 Continued

Q(20)	42.6930	[0.002]	165.3912	[0.003]
Q^2(20)	196.426	[0.000]	162.9281	[0.000]
A(20)	252.419	[0.000]	216.6783	[0.000]

Notes: Data frequency is weekly. Mean test tests the null hypothesis of means equality and its *p*-value are displayed as [.]. Levene statistic tests the null hypothesis of variances equality and its *p*-value is displayed as [.]. Skewness refers to series skewness coefficient. The asymptotic distribution of the skewness coefficient under the null hypothesis is $N(0,6/T)$, where T is the sample size. The null hypothesis tested is whether that coefficient is equal to zero. Kurtosis refers to the series kurtosis coefficient. The null hypothesis tested is whether that coefficient is equal to zero. The asymptotic distribution of the kurtosis coefficient under the null hypothesis is $N(0,24/T)$, where T is the sample size. Normality refers to the Bera-Jarque statistic test. This statistic tests the normality or non-normality of the series. The Bera-Jarque statistic is calculated as follows, $T\ [S^2/6 + (K-3)^2/24]$, where S is the skewness coefficient and K is the kurtosis coefficient. Under the null hypothesis of normal distribution, the Bera-Jarque statistic has an asymptotic $\chi^2(2)$ distribution. Q(20) and Q^2(20) are Ljung-Box tests for twentieth-order serial correlation in the returns and squared returns. A(20) is Engle (1982) test for twentieth-order ARCH. The *p*-values of these tests are displayed as[.]

in variance between the large cap and small cap indices. Secondly, there are not significant differences in means, although for many years, like 2000, the return was quite different indicating that both markets could be offering different sensitivities to risk factors. Large firms depend on global risk factors, however, risk factors that affect small firms are located basically in their own economy. Thirdly, in all countries the correlation between both indices has dropped. We can interpret this last fact as a segmentation of both markets. Therefore, diversification strategies would be gaining an important role in portfolio management. These results point out that it is important to study more accurately the covariance dynamic between both financial time series.

17.4 RESULTS

This section presents the model estimates. With the ADC model it is possible to quantify volatility spillovers and contrast the volatility feedback effect. In order to estimate the model in equations (17.1) and (17.2), a conditional normal distribution for the innovation vector is assumed and the quasi-maximum likelihood method is applied. Bollerslev and Wooldridge (1992) show that the standard errors calculated by this method are robust even when the normality assumption is violated. Panel (A) of Table 17.3 displays the quasi-maximum likelihood estimates of the CAPM-ADC model for the three countries. Panel (B) displays the Wald test for the restrictions imposed on the ADC model to obtain the encompassed models.

The results for the three countries are quite similar, and three facts can be highlighted from the estimation. Firstly, the price of risk (Y) is positive

Table 17.2 Returns, volatilities and correlations

Panel A: returns, volatilities and correlations of French indices series							
Year	Annualized Returns (%)			Annualized Volatilities (%)			Correlation(3)
	CAC 40	MID CAC	Mean Test(1)	CAC 40	MID CAC	Levene(2) Test	
1991	15.87	−0.54	0.72	17.49	14.47	4.18*	0.78
1992	4.42	−5.78	0.40	19.31	11.78	8.46*	−0.06
1993	27.45	32.38	0.52	17.04	10.47	16.24*	−0.04
1994	−12.55	−3.20	0.38	19.91	11.84	17.70*	−0.06
1995	0.61	−17.12	1.07	14.12	8.66	10.97*	0.55
1996	22.96	33.50	0.58	14.04	11.30	4.47*	0.53
1997	29.06	18.64	0.44	19.00	14.25	4.20*	0.70
1998	29.31	5.58	0.74	25.08	19.95	3.47	0.69
1999	42.45	30.26	0.18	17.91	11.03	10.77*	−0.06
2000	−1.71	21.88	0.30	22.13	31.79	0.75	−0.16
2001	−26.71	−22.41	0.18	25.18	23.03	4.97*	0.29
2002	−36.84	−26.72	0.32	37.48	29.69	8.57*	0.43
2003	8.43	−13.41	0.78	27.04	12.14	12.13*	0.31
2004(4)	6.56	32.32	1.36	13.24	7.69	6.45*	0.06

Panel B: returns, volatilities and correlations of German indices series							
Year	Annualized Returns (%)			Annualized Volatilities (%)			Correlation(3)
	DAX	SDAX	Mean Test(1)	DAX	SDAX	Levene(2) Test	
1991	8.73	−1.76	1.08	16.42	11.00	6.50*	0.72
1992	0.34	−10.25	0.56	16.19	10.18	5.03*	0.7
1993	33.78	33.84	6.00	13.48	10.68	3.18	−0.06
1994	−4.89	−5.82	0.04	18.51	8.60	20.96*	0.62
1995	7.61	−11.35	1.18	14.28	7.29	13.92*	0.53
1996	22.34	8.45	0.41	3.25	8.28	3.05	0.08
1997	39.35	18.28	0.89	20.10	13.07	4.75*	0.71
1998	16.30	−1.85	0.62	24.95	15.45	12.00*	0.71
1999	31.57	6.93	0.86	25.95	12.08	19.36*	0.63
2000	−8.06	4.98	0.47	24.26	12.97	16.40*	0.67
2001	−21.24	−25.48	0.13	30.46	14.19	22.65*	0.63
2002	−0.53	−35.45	0.44	36.94	16.05	20.48*	0.77
2003	15.97	40.96	0.74	33.82	13.06	17.70*	0.18
2004(4)	−13.48	9.09	0.82	17.83	13.95	1.82	0.08

Continued

Table 17.2 Continued

Panel C: returns, volatilities and correlations of British indices series							
Year	Annualized Returns (%)			Annualized Volatilities (%)			Correlation(3)
	FTSE	SMALLCAP	Mean Test(1)	FTSE	SMALLCAP	Levene(2) Test	
1993	20.65	22.06	0.09	11.98	6.81	10.08*	0.63
1994	−11.18	−6.19	0.27	15.29	9.35	15.69*	0.55
1995	17.19	10.27	0.58	10.51	5.26	20.87*	0.60
1996	7.45	7.61	0.25	9.01	6.71	4.79*	−0.07
1997	21.84	6.47	0.93	14.88	7.33	23.26*	0.02
1998	13.59	−11.07	0.94	20.23	16.38	4.49*	0.11
1999	11.04	39.56	1.35	18.44	10.25	18.92*	−0.03
2000	−7.50	2.62	0.25	13.86	14.24	0.002	0.02
2001	−16.72	−21.00	0.15	18.67	20.95	0.07	−0.13
2002	−33.20	−32.59	0.02	28.01	15.79	7.34*	0.19
2003	17.47	34.67	0.74	18.90	13.59	0.08	0.53
2004(4)	0.65	−11.23	0.64	10.76	8.24	0.99	0.56

Notes: Data frequency is weekly.
(1) This column displays the means equality test. Significant coefficients at 95% of confidence level are highlighted with one asterisk (*). (2) This column displays the variances equality test known as Levene test. Significant coefficients at 95% of confidence level are highlighted with one asterisk (*). (3) This column displays the annual correlation between both indices. (4) This row displays the results for the period 7 January to 25 August 2004.

and significant for all countries. Therefore, in these markets the risk is valued. This is a surprising result because most studies find a positive but non-significant relationship between expected return and risk (French, Schwert and Stambaugh, 1997; Campbell and Hentschel, 1992) or a negative and significant relationship (Campbell, 1987; Officer, 1973; Glosten, Jagannatham and Runkle, 1993). Moreover, this result is consistent with the volatility feedback hypothesis. If volatility is priced, an anticipated increase in volatility raises the required return on equity, and therefore there will be observed time-varying risk premiums. Secondly, coefficients g_{11} and g_{22} are significant, showing that in both indices (large cap and small cap), negative asymmetric firm volatility is important for their own dynamic, with the exception of the SDAX index. For this index, the estimated coefficient g_{22} is not significant. Thirdly, coefficients g_{12} and g_{21} are both significant, showing that cross-relationships between negative shocks in both markets are also significant.

The encompassing model restrictions on the ADC were rejected. This result means that the ADC cannot be reduced to any nested model. The estimated values of ϕ_{12} (close but significantly different to 1) and ρ_{12} (close but

Table 17.3 Conditional CAPM ADC-in-mean model estimates and restrictions tests

Estimates of the conditional CAPM ADC-in-mean model in the French market
Panel (A): model estimates

$$\underset{(0.00)}{Y = 2.27}$$

$$C = \begin{bmatrix} \underset{(0.00)}{0.0039} & \underset{(0.00)}{0.0025} \\ & \underset{(0.00)}{0.0018} \end{bmatrix} \qquad B = \begin{bmatrix} \underset{(0.00)}{0.9238} & \underset{(0.00)}{0.0344} \\ \underset{(0.00)}{-0.1168} & \underset{(0.00)}{0.7576} \end{bmatrix}$$

$$A = \begin{bmatrix} \underset{(0.00)}{-0.2231} & \underset{(0.00)}{-0.0577} \\ \underset{(0.00)}{0.3573} & \underset{(0.00)}{0.4894} \end{bmatrix} \qquad G = \begin{bmatrix} \underset{(0.00)}{0.2441} & \underset{(0.00)}{0.0899} \\ \underset{(0.00)}{0.3302} & \underset{(0.00)}{0.5943} \end{bmatrix}$$

$$\phi_{12} = 1.0278$$
$$\rho_{12} = -0.0251$$

Panel (B): testing restrictions for nested models

BEKK	1.95×10^5 (0.00)
VECH	3.80×10^6 (0.00)
CCORR	4.26×10^6 (0.00)

Estimates of the conditional CAPM ADC-in-mean model in the German market
Panel (C): model estimates

$$\underset{(0.00)}{Y = 5.84}$$

$$C = \begin{bmatrix} \underset{(0.00)}{0.0052} & \underset{(0.00)}{0.0019} \\ & \underset{(0.00)}{-0.0019} \end{bmatrix} \qquad B = \begin{bmatrix} \underset{(0.00)}{0.8895} & \underset{(0.01)}{-0.0036} \\ \underset{(0.00)}{-0.1869} & \underset{(0.00)}{0.7948} \end{bmatrix}$$

$$A = \begin{bmatrix} \underset{(0.00)}{0.2895} & \underset{(0.00)}{-0.0526} \\ \underset{(0.00)}{0.2678} & \underset{(0.00)}{0.4590} \end{bmatrix} \qquad G = \begin{bmatrix} \underset{(0.00)}{0.2466} & \underset{(0.02)}{0.0244} \\ \underset{(0.00)}{0.1361} & \underset{(0.88)}{0.0052} \end{bmatrix}$$

$$\phi_{12} = 1.0067$$
$$\rho_{12} = -0.0732$$

Panel (D): testing restrictions for nested models

BEKK	2.96×10^5 (0.00)
VECH	4.62×10^6 (0.00)
CCORR	4.81×10^6 (0.00)

Continued

Table 17.3 Continued

Estimates of the conditional CAPM ADC-in-mean model in the British market
Panel (E): model estimates

$$Y = \underset{(0.00)}{2.75}$$

$$C = \begin{bmatrix} \underset{(0.88)}{0.0048} & \underset{(0.00)}{0.0016} \\ & \underset{(0.07)}{-0.0002} \end{bmatrix} \qquad B = \begin{bmatrix} \underset{(0.00)}{0.7199} & \underset{(0.00)}{0.0355} \\ \underset{(0.00)}{-0.2777} & \underset{(0.00)}{0.7935} \end{bmatrix}$$

$$A = \begin{bmatrix} \underset{(0.00)}{0.4118} & \underset{(0.00)}{0.1649} \\ \underset{(0.00)}{-0.2244} & \underset{(0.00)}{-0.4323} \end{bmatrix} \qquad G = \begin{bmatrix} \underset{(0.00)}{0.1216} & \underset{(0.00)}{0.3958} \\ \underset{(0.00)}{0.6651} & \underset{(0.00)}{0.1287} \end{bmatrix}$$

$$\phi_{12} = 1.0142$$
$$\rho_{12} = -0.0632$$

Panel (F): testing restrictions for nested models

BEKK	2.75×10^5 (0.00)
VECH	4.25×10^6 (0.00)
CCORR	4.37×10^6 (0.00)

Notes: Panels (A), (C) and (E) of this table display the quasi-maximum likelihood estimates of the conditional CAPM ADC-M model defined in equations (1) and (2). *P*-values appear in brackets. In the three cases, the necessary conditions for the stationarity of the process are satisfied.
Panels (B), (D) and (F) display the Wald test for the restrictions imposed on the ADC model to obtain the encompassed models. *p*-values appear in brackets. The specification test proposed by Kroner and Ng (1998) is as follows: (1) If $\rho_{12} = b_{12} = b_{21} = a_{12} = a_{21} = g_{12} = g_{21} = 0$, a restricted asymmetric VECH is obtained; (2) if $\phi_{12} = b_{12} = b_{21} = a_{12} = a_{21} = g_{12} = g_{21} = 0$, the asymmetric CCORR is derived; (3) if $\phi_{12} = 1$ and $\rho_{12} = 0$ the asymmetric BEKK model is obtained.

significantly below zero) reveal that the estimated ADC model has similar properties to the BEKK model, although the encompassing restrictions are clearly rejected.

Table 17.4 displays an analysis of the standardized residuals. It can be observed that, in the large firm index, autocorrelation and heteroskedasticity problems have been successfully amended. Regarding the small cap index, heteroskedasticity disappears but autocorrelation remains. However, it must be highlighted that our main focus is on conditional second moments, and Nelson (1992) shows that misspecification in the conditional mean does not affect the key properties of the second moments.

Figure 17.2 displays the annualized conditional volatility of both indices over the studied period for the three countries. Both volatility series have similar patterns but large stock index volatility is almost always higher than small stock index volatility.

Table 17.4 Summary statistics for the standardized residuals

Panel A: summary statistics for the standardized residuals of France

	$\varepsilon_{1,t}/\sqrt{h_{11,t}}$		$\varepsilon_{2,t}/\sqrt{h_{22,t}}$	
Mean	−0.0146		0.0101	
Variance	0.9866		1.0179	
Skewness	−0.1160	[0.221]	−0.2340	[0.011]
Kurtosis	−0.0311	[0.866]	1.3391	[0.000]
Normality	13.4101	[0.001]	59.6086	[0.000]
Q(20)	29.9128	[0.171]	147.223	[0.000]
Q^2(20)	32.5969	[0.137]	20.8033	[0.408]
A(20)	14.7484	[0.790]	20.5445	[0.424]

Panel B: summary statistics for the standardized residuals of Germany

	$\varepsilon_{1,t}/\sqrt{h_{11,t}}$		$\varepsilon_{2,t}/\sqrt{h_{22,t}}$	
Mean	−0.0819		−0.0714	
Variance	0.9861		0.9996	
Skewness	−1.7001	[0.000]	−0.3606	[0.000]
Kurtosis	11.3915	[0.000]	1.3637	[0.000]
Normality	4192.75	[0.000]	70.5993	[0.000]
Q(20)	30.3931	[0.263]	105.185	[0.000]
Q^2(20)	1.5671	[0.999]	13.4789	[0.855]
A(20)	2.2071	[0.999]	15.8992	[0.722]

Panel C: summary statistics for the standardized residuals of Great Britain

	$\varepsilon_{1,t}/\sqrt{h_{11,t}}$		$\varepsilon_{2,t}/\sqrt{h_{22,t}}$	
Mean	−0.0547		0.0490	
Variance	0.9709		0.9653	
Skewness	−0.2925	[0.003]	−0.2296	[0.022]
Kurtosis	0.4162	[0.039]	0.6244	[0.002]
Normality	12.7144	[0.001]	14.8173	[0.000]
Q(20)	19.9705	[0.459]	163.784	[0.000]
Q^2(20)	16.1603	[0.706]	19.4896	[0.490]
A(20)	21.9096	[0.345]	20.0035	[0.457]

Notes: Skewness refers to series skewness coefficient. The asymptotic distribution of the skewness coefficient under the null hypothesis is $N(0, 6/T)$, where T is the sample size. The null hypothesis tested is whether that coefficient is equal to zero. Kurtosis refers to the series kurtosis coefficient. The null hypothesis tested is whether that coefficient is equal to zero. The asymptotic distribution of the kurtosis coefficient under the null hypothesis is $N(0,24/T)$, where T is the sample size. Normality refers to the Bera-Jarque statistic test. This statistic tests the normality or non-normality of the series. The Bera-Jarque statistic is calculated as follows, $T[S^2/6 + (K-3)^2/24]$, where S is the skewness coefficient and K is the kurtosis coefficient. Under the null hypothesis of normal distribution, the Bera-Jarque statistic has an asymptotic χ^2 (2) distribution. Q(20) and Q^2(20) are Ljung Box tests. The null hypothesis tested is the no existence of twentieth order serial correlation in $\varepsilon_{1,t}$, $\varepsilon_{2,t}$ and $\varepsilon^2_{1,t}$, $\varepsilon^2_{2,t}$ respectively. Finally, A(20) is the Engle (1982) test. The null hypothesis tested is the non-existence of twentieth-order ARCH in the residuals. The p-values of these tests are displayed as [.].

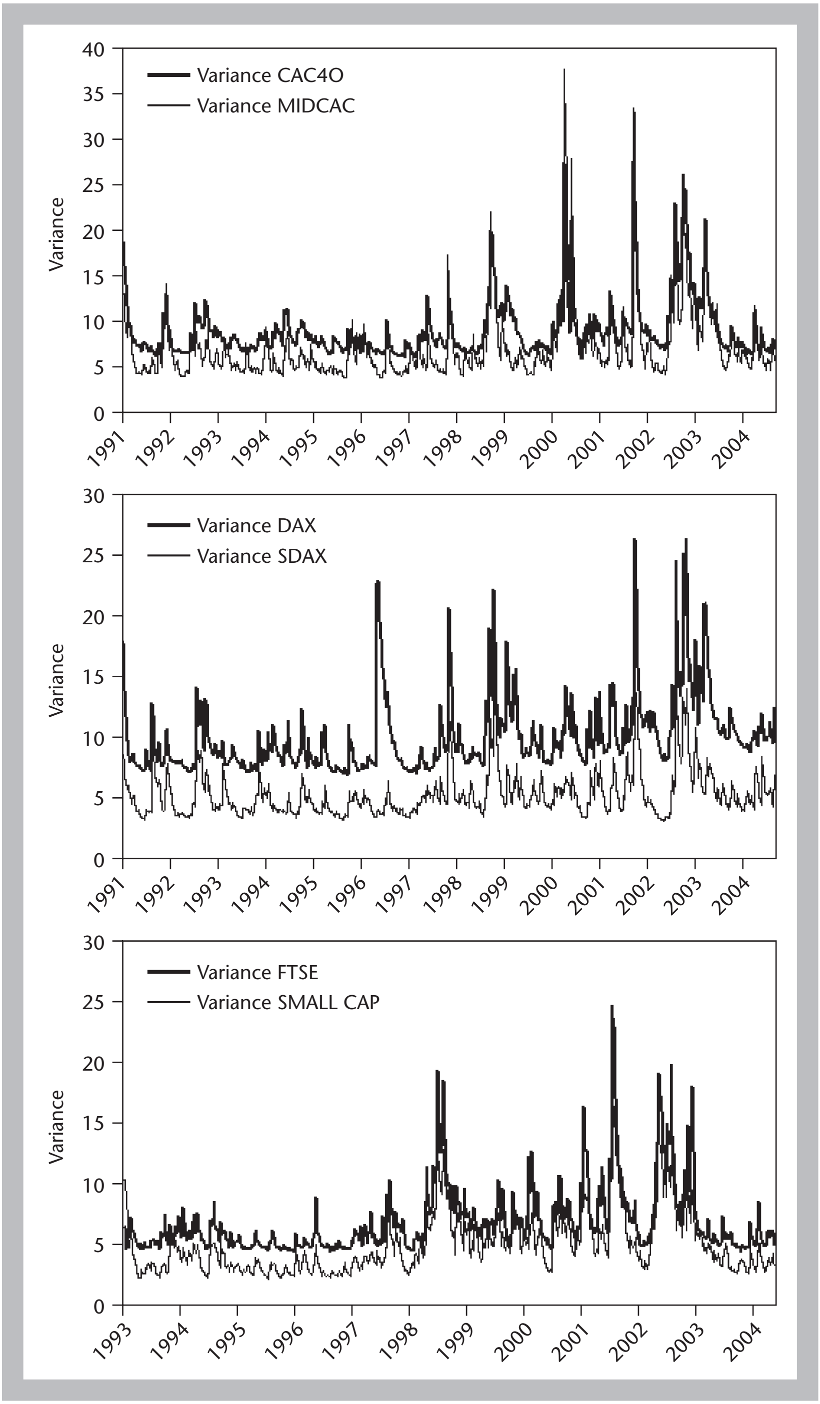


Figure 17.2 Annualized conditional volatility of the stock indices over the studied period

Table 17.5 Unconditional moment estimates of all markets

Unconditional moment estimates			
	France	**Germany**	**Great Britain**
Large cap index variance	9.31	11.73	5.52
Small cap index variance	5.30	3.19	3.04
Covariance	4.38	3.41	0.80
Beta coefficient	0.47	0.29	0.15
Correlation	0.62	0.55	0.20

Notes: The beta coefficient is defined as the quotient between the large cap and small cap indices covariance and the large cap index variance. Variances and covariance are reported in percentage terms.

17.5 ASYMMETRIES ANALYSIS

The asymmetries analysis is carried out in two steps. First, a graphical analysis of news impact surfaces is displayed. Second, the robust conditional moment test of Wooldridge (1990) is conducted before and after estimating the covariance model.

Table 17.5 displays the unconditional moment estimates of all markets in order to facilitate comprehension of the following sections.

17.5.1 News impact surfaces

The news impact surfaces are the multivariate generalization of the news impact curves introduced by Engle and Ng (1993) and Hentschel (1995). The univariate applications plot the conditional variance against the last period's shocks. The multivariate generalization plots the conditional variance and covariance against large- and small-firm shocks from the last period, holding past conditional variances and covariances constant at their unconditional sample levels.

Figure 17.3 shows the news impact surfaces for the conditional second moments and conditional betas obtained from the asymmetric bivariate GARCH specification. Following Engle and Ng (1993) and Kroner and Ng (1998), each surface is represented in the region $\varepsilon_{i,t} = [-5, 5]$ for $i = 1, 2$. The news impact surfaces try to reflect the sensitivity of variances, covariances and betas to the sign and size of the innovations.

Panel (A) of Figure 17.3 displays the news impact surfaces for the French market. It can be seen that the CAC40 variance increases the most when cross-signed shocks take place. The MIDCAC variance surface shows a clear sensitivity to its own negative shocks when positive or negative shocks on

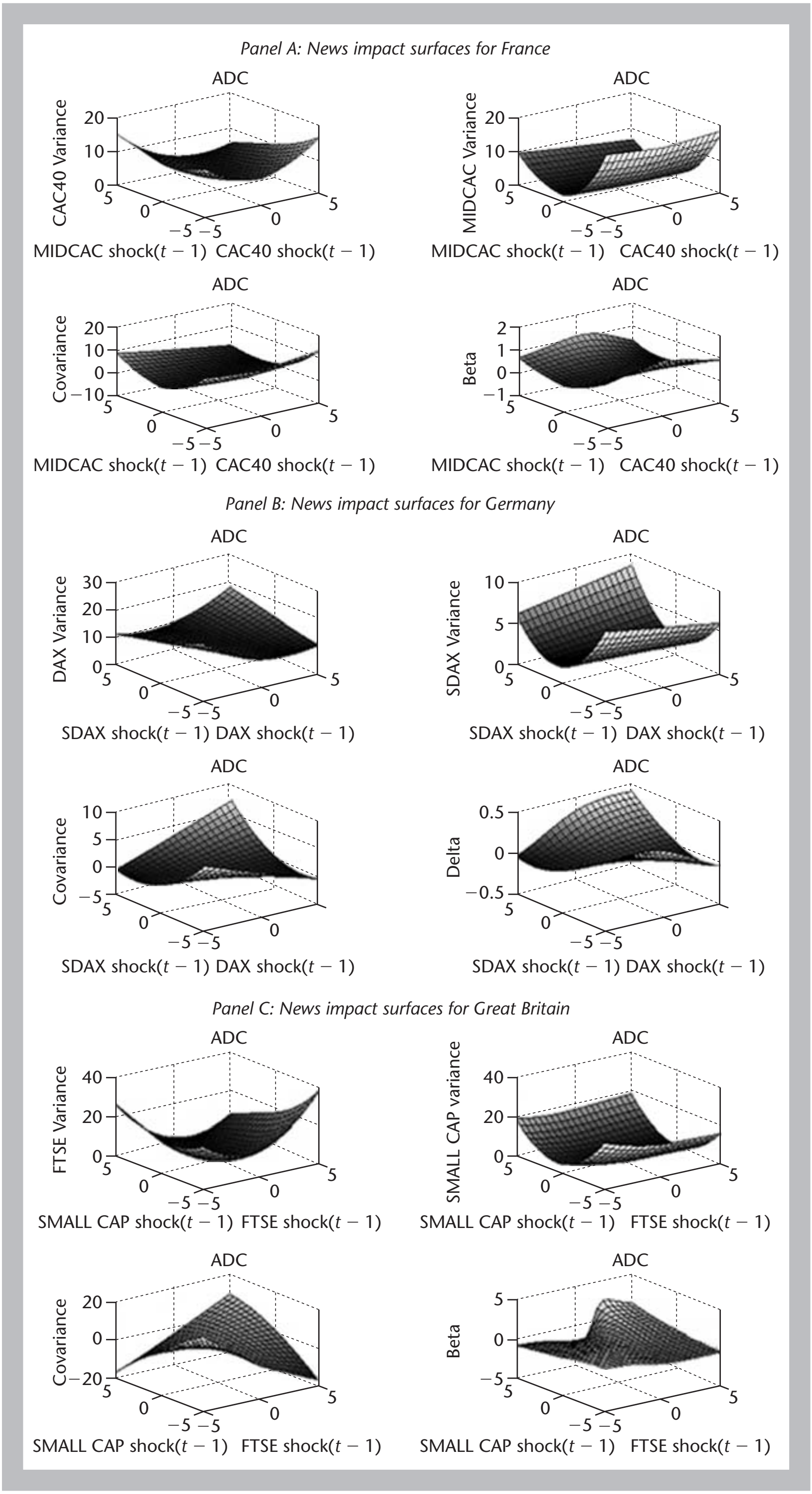


Figure 17.3 News impact surfaces for the ADC model

the CAC40 come together. In addition, the covariance surface is quite flat, increasing as negative shocks in the MIDCAC take larger values. Finally, it can be observed that the beta coefficient surface is quite stable.

Panel (B) of Figure 17.3 displays the news impact surfaces for the German market. It can be appreciated that the DAX variance surface shows a clear asymmetry; variance increases the most when both shocks are of the same sign. In addition, this increase is larger when both shocks are negative. The SDAX variance increases the most when both the large stock index shock and the small stock index shock are of the same sign. Moreover, covariance only increases when both shocks are of the same sign. When shocks are cross-signed, covariance slightly decreases. Finally, it can be appreciated that the beta coefficient has the expected behavior, it increases with the shock size when both socks are of the same sign and decreases when shocks are of different sign.

Panel (C) of Figure 17.3 displays the news impact surfaces for the British market. It can be seen that the FTSE variance increases the most when cross-signed shocks take place or, when both shocks are negative. The SMALL CAP variance increases the most when both shocks are negative. Moreover the covariance surface shows a clear asymmetry; the covariance increases when both shocks are of the same sign whereas decreases when shocks are of different sign. Finally, the beta coefficient surface is very sensitive to extreme positive shocks in the small cap index.

17.5.2 Robust conditional moment test

The robust conditional moment test of Wooldridge (1990) is applied to test how the Glosten *et al.* (1993) modification to the multivariate GARCH model cleans the asymmetries in the conditional covariance matrix. This test enables the identification of possible sources of misspecification in the model and is robust to distributional assumptions. In order to test the validity of a model, a natural approach is to compare the *ex post* cross-product matrix of the residuals ($\sqrt{T}$-consistent estimator) to the estimated covariance matrix. Thus, Kroner and Ng (1998) define a "generalized residual" as $v_{ijt} = \varepsilon_{it}\varepsilon_{jt} - h_{ijt}$ for all $i, j = 1, 2$. If the model is correct, $E_{t-1}(v_{ijt}) = 0$, and therefore v_{ijt} should be uncorrelated with any variable known at time $t-1$. These variables are called misspecification indicators. Kroner and Ng (1998) use three kinds of misspecification indicators. These indicators try to detect misspecification caused by shocks' sign ($I(\varepsilon_{1t} < 0)$ and $I(\varepsilon_{2t} < 0)$), the four quadrant sign combinations ($I(\varepsilon_{1t-1} > 0; \varepsilon_{2t-1} > 0)$, $I(\varepsilon_{1t-1} < 0; \varepsilon_{2t-1} > 0)$, $I(\varepsilon_{1t-1} > 0; \varepsilon_{2t-1} < 0)$, $I(\varepsilon_{1t-1} < 0; \varepsilon_{2t-1} < 0)$) and the misspecification due to the cross effects of shocks' size and sign ($\varepsilon^2_{1t-1} I(\varepsilon_{1t-1} < 0)$, $\varepsilon^2_{1t-1} I(\varepsilon_{2t-1} < 0)$, $\varepsilon^2_{2t-1} I(\varepsilon_{1t-1} < 0)$, $\varepsilon^2_{2t-1}(\varepsilon_{2t-1} < 0)$). $I(.)$ denotes an indicator function that equals one if the argument is true and zero otherwise. Using these misspecification indicators, the robust conditional moment test of Wooldridge (1990) is applied.

Table 17.6 shows the result of the robust conditional moment test. Panels (A), (C) and (E) display the test result when unconditional moment estimates are used. It can be seen that asymmetries are very important, especially in the beta coefficient. Panels (B), (D) and (F) offer the test result when conditional moment estimates are used. After this, no asymmetric pattern remains in the conditional covariance specification. This is an important result because it means that the GARCH specification is gathering all the possible asymmetries in the conditional covariance matrix. This result also guarantees that the analysis of volatility contagion carried out later will be reliable.

Special attention is required for the beta coefficient (following Wooldridge, 1990, a consistent estimator of the beta coefficient is built using the continuous function property on consistent estimators; see Hamilton, 1994: 182). Last column in Table 17.6 shows that the unconditional beta estimate has a significant error specification but the conditional beta estimates within the used model are insensitive to sign and size asymmetries in unexpected shock returns. As beta coefficients are market risk sensitivity measures, it is important to use a conditional model in order to avoid error specification on estimating the beta coefficients.

17.6 VOLATILITY SPILLOVERS

In this section, volatility spillovers between large and small firms are quantified. We differentiate between positive and negative shocks, but we also

Table 17.6 Robust conditional moment tests

Robust Conditional Moment Test in the French Market Panel (A): applied on original returns				
	$\upsilon_{12,t} = r_{1,t}r_{2,t} - \sigma_{12}$	$\upsilon_{1,t} = r^2_{1,t} - \sigma^2_1$	$\upsilon_{2,t} = r^2_{2,t} - \sigma^2_2$	$\upsilon_{beta_t} = r_{1,t}r_{2,t}/r^2_{1,t} - \sigma_{12}/\sigma^2_1$
$I(r_{1,t-1} < 0)$	52.57695***	1.68619	0.94676	317.99998***
$I(r_{2,t-1} < 0)$	44.80857***	4.31226**	0.59401	317.08568***
$I(r_{1,t-1} < 0;\ r_{2,t-1} < 0)$	24.47034***	3.44086*	1.55390	230.99999***
$I(r_{1,t-1} < 0;\ r_{2,t-1} > 0)$	64.00011***	7.41064***	0.84142	87.94500***
$I(r_{1,t-1} > 0;\ r_{2,t-1} < 0)$	30.81728***	0.87521	3.92127**	87.03500***
$I(r_{1,t-1} > 0;\ r_{2,t-1} > 0)$	163.14585***	20.95558***	0.31124	305.02800***
$r^2_{1,t-1}I(r_{1,t-1} < 0)$	1.13863	4.87991**	5.12130**	67.88843***
$r^2_{1,t-1}I(r_{2,t-1} < 0)$	1.43294	6.23543**	5.38658**	72.96868***
$r^2_{2,t-1}I(r_{1,t-1} < 0)$	2.56038	6.59357**	6.08929**	29.55023***
$r^2_{2,t-1}I(r_{2,t-1} < 0)$	2.69359	6.74448***	5.88425**	27.93598***

Continued

Table 17.6 Continued

Panel (B): applied on the residuals of the model estimates

	$\upsilon_{12,t} = \varepsilon_{1,t}\varepsilon_{2,t} - \sigma_{12,t}$	$\upsilon_{1,t} = \varepsilon_{1,t}^2 - \sigma_{1,t}^2$	$\upsilon_{2,t} = \varepsilon_{2,t}^2 - \sigma_{2,t}^2$	$\upsilon_{beta_t} = \varepsilon_{1,t}\varepsilon_{2,t}/\varepsilon_{1,t}^2 - \sigma_{12,t}/\sigma_{1,t}^2$
$I(\varepsilon_{1,t-1} < 0)$	0.10825	0.0791841	0.02020	0.02748
$I(\varepsilon_{2,t-1} < 0)$	1.28730	0.0212508	0.00353	0.21586
$I(\varepsilon_{1,t-1} < 0; \varepsilon_{2,t-1} < 0)$	0.50478	0.445375	0.04633	0.14528
$I(\varepsilon_{1,t-1} < 0; \varepsilon_{2,t-1} > 0)$	0.50645	0.0306266	0.02375	0.59168
$I(\varepsilon_{1,t-1} > 0; \varepsilon_{2,t-1} < 0)$	0.81626	0.101018	0.39831	0.01141
$I(\varepsilon_{1,t-1} > 0; \varepsilon_{2,t-1} > 0)$	0.00228	0.923719	0.01846	0.00738
$\varepsilon_{1,t-1}^2 I(\varepsilon_{1,t-1} < 0)$	1.79865	0.326715	0.04021	1.30713
$\varepsilon_{1,t-1}^2 I(\varepsilon_{2,t-1} < 0)$	1.22818	0.00153	0.44407	1.47350
$\varepsilon_{2,t-1}^2 I(\varepsilon_{1,t-1} < 0)$	0.78609	0.0259686	0.00129	1.56487
$\varepsilon_{2,t-1}^2 I(\varepsilon_{2,t-1} < 0)$	1.79558	0.048851	0.04021	1.77788

Robust Conditional Moment Test in the German Market

Panel (C): applied on original returns

	$\upsilon_{12,t} = r_{1,t}r_{2,t} - \sigma_{12}$	$\upsilon_{1,t} = r_{1,t}^2 - \sigma_1^2$	$\upsilon_{2,t} = r_{2,t}^2 - \sigma_2^2$	$\upsilon_{beta_t} = r_{1,t}r_{2,t}/r_{1,t}^2 - \sigma_{12}/\sigma_1^2$
$I(r_{1,t-1} < 0)$	51.85540***	0.38040	1.46245	308.99998***
$I(r_{2,t-1} < 0)$	48.82691***	0.01518	1.95438	331.99997***
$I(r_{1,t-1} < 0; r_{2,t-1} < 0)$	20.48819***	0.58614	3.76918*	217.99998***
$I(r_{1,t-1} < 0; r_{2,t-1} > 0)$	67.68802***	0.00253	7.80934***	91.00000***
$I(r_{1,t-1} > 0; r_{2,t-1} < 0)$	37.72143***	1.79143	1.06502	113.99999***
$I(r_{1,t-1} > 0; r_{2,t-1} > 0)$	179.36793***	0.01328	2.41438	288.00000***
$r_{1,t-1}^2 I(r_{1,t-1} < 0)$	0.10512	2.18227	3.09202*	33.20338***
$r_{1,t-1}^2 I(r_{2,t-1} < 0)$	0.03183	1.96722	3.73741*	42.87739***
$r_{2,t-1}^2 I(r_{1,t-1} < 0)$	1.08555	1.58957	4.87424**	59.51383***
$r_{2,t-1}^2 I(r_{2,t-1} < 0)$	0.29315	2.26819	5.76897**	58.65094***

Panel (D): applied on the residuals of the model estimates

	$\upsilon_{12,t} = \varepsilon_{1,t}\varepsilon_{2,t} - \sigma_{12,t}$	$\upsilon_{1,t} = \varepsilon_{1,t}^2 - \sigma_{1,t}^2$	$\upsilon_{2,t} = \varepsilon_{2,t}^2 - \sigma_{2,t}^2$	$\upsilon_{beta_t} = \varepsilon_{1,t}\varepsilon_{2,t}/\varepsilon_{1,t}^2 - \sigma_{12,t}/\sigma_{1,t}^2$
$I(\varepsilon_{1,t-1} < 0)$	0.57405	2.22941	0.61996	1.45365
$I(\varepsilon_{2,t-1} < 0)$	0.42169	1.64955	0.61688	1.26071
$I(\varepsilon_{1,t-1} < 0; \varepsilon_{2,t-1} < 0)$	2.17062	0.88730	2.06230	0.88062
$I(\varepsilon_{1,t-1} < 0; \varepsilon_{2,t-1} > 0)$	1.57036	1.90948	1.79316	0.03676
$I(\varepsilon_{1,t-1} > 0; \varepsilon_{2,t-1} < 0)$	2.48587	0.44863	1.66942	1.03544
$I(\varepsilon_{1,t-1} > 0; \varepsilon_{2,t-1} > 0)$	0.04044	1.21521	0.00106	0.71961
$\varepsilon_{1,t-1}^2 I(\varepsilon_{1,t-1} < 0)$	0.88337	0.65406	0.30436	1.09062

Continued

Table 17.6 Continued

$\varepsilon^2_{1,t-1} I(\varepsilon_{2,t-1} < 0)$	0.68327	0.02167	0.30698	0.83677
$\varepsilon^2_{2,t-1} I(\varepsilon_{1,t-1} < 0)$	1.28405	0.01642	1.04079	0.93523
$\varepsilon^2_{2,t-1} I(\varepsilon_{2,t-1} < 0)$	1.29078	0.65406	0.30436	0.98573

Robust Conditional Moment Test in the British Market

Panel (E): applied on original returns

	$\upsilon_{12,t} = r_{1,t}r_{2,t} - \sigma_{12}$	$\upsilon_{1,t} = r^2_{1,t} - \sigma^2_1$	$\upsilon_{2,t} = r^2_{2,t} - \sigma^2_2$	$\upsilon_{beta_t} = r_{1,t}r_{2,t}/r^2_{1,t} - \sigma_{12}/\sigma^2_1$
$I(r_{1,t-1} < 0)$	90.99919***	3.13913*	0.31223	274.00000***
$I(r_{2,t-1} < 0)$	66.41886***	4.32508**	2.61077	254.99999***
$I(r_{1,t-1} < 0;\ r_{2,t-1} < 0)$	35.15208***	5.16067**	1.20278	146.00000***
$I(r_{1,t-1} < 0;\ r_{2,t-1} > 0)$	71.82084***	3.13945*	1.38575	128.00000***
$I(r_{1,t-1} > 0;\ r_{2,t-1} < 0)$	31.81919***	0.19653	1.61586	109.00000***
$I(r_{1,t-1} > 0;\ r_{2,t-1} > 0)$	156.81439***	24.13232***	10.78132***	208.00000***
$r^2_{1,t-1} I(r_{1,t-1} < 0)$	0.93860	3.34471*	2.58045	56.42158***
$r^2_{1,t-1} I(r_{2,t-1} < 0)$	0.84860	3.14928*	2.82037*	49.64218***
$r^2_{2,t-1} I(r_{1,t-1} < 0)$	4.56553**	5.69538**	1.83772	33.92826***
$r^2_{2,t-1} I(r_{2,t-1} < 0)$	4.89998**	6.07446**	2.87220*	33.51451***

Panel (F): applied on the residuals of the model estimates

	$\upsilon_{12,t} = \varepsilon_{1,t}\varepsilon_{2,t} - \sigma_{12,t}$	$\upsilon_{1,t} = \varepsilon^2_{1,t} - \sigma^2_{1,t}$	$\upsilon_{2,t} = \varepsilon^2_{2,t} - \sigma^2_{2,t}$	$\upsilon_{beta_t} = \varepsilon_{1,t}\varepsilon_{2,t}/\varepsilon^2_{1,t} - \sigma_{12,t}/\sigma^2_{1,t}$
$I(\varepsilon_{1,t-1} < 0)$	0.04181	1.49114	0.05898	0.27055
$I(\varepsilon_{2,t-1} < 0)$	1.87779	0.02587	0.81170	0.75620
$I(\varepsilon_{1,t-1} < 0;\ \varepsilon_{2,t-1} < 0)$	0.11286	1.21553	0.32800	1.30884
$I(\varepsilon_{1,t-1} < 0;\ \varepsilon_{2,t-1} > 0)$	0.86349	0.32819	0.16350	0.70133
$I(\varepsilon_{1,t-1} > 0;\ \varepsilon_{2,t-1} < 0)$	1.93700	1.25880	0.07043	0.02128
$I(\varepsilon_{1,t-1} > 0;\ \varepsilon_{2,t-1} > 0)$	1.40476	0.21308	0.39134	0.28494
$\varepsilon^2_{1,t-1} I(\varepsilon_{1,t-1} < 0)$	0.36570	0.00452	0.15896	0.93864
$\varepsilon^2_{1,t-1} I(\varepsilon_{2,t-1} < 0)$	0.60704	0.84934	0.85204	0.92168
$\varepsilon^2_{2,t-1} I(\varepsilon_{1,t-1} < 0)$	0.25875	0.49954	0.16039	0.03524
$\varepsilon^2_{2,t-1} I(\varepsilon_{2,t-1} < 0)$	0.31650	0.00452	0.15896	1.10559

Notes: Panels (A), (C) and (E) give the robust conditional moment test statistic applied on unconditional moment estimates, where σ^2_1, σ^2_2, σ_{12} and beta coefficient are unconditional estimates of the large stock index variance, small stock index variance, its covariance and beta, respectively. Panels (B), (D) and (F) give the robust conditional moment test on the conditional estimates, where $\sigma^2_{1,t}$, $\sigma^2_{2,t}$, $\sigma_{12,t}$ and beta_t coefficient are conditional estimates of the large stock index variance, small stock index variance, its covariance and beta, respectively, obtained from the asymmetric GARCH model. The misspecification indicators are listed in the first column and the remaining columns in each panel give the test statistic computed for the generalized residual calculated as shown in the first row in each panel. $r_{1,t-1}$ and $r_{2,t-1}$ are the return series of the large stock and small stock indexes respectively. $\varepsilon_{1,t-1}$ is the return shock to the large index and $\varepsilon_{2,t-1}$ is the return shock to the small index. The indicator function $I()$ takes the value one if the expression inside the parentheses is satisfied and zero otherwise. All the statistics are distributed as $\chi^2(1)$. Test values highlighted with one (*), two (**) and three (***) asterisks are significant at 90%, 95% and 99%, respectively.

look at the combined effect of typical shocks occurring simultaneously on large cap and small cap indexes. Since the unconditional correlation coefficients between the large cap and the small cap index returns are 0.62, 0.55 and 0.20 for France, Germany and United Kingdom respectively, it can be assumed that, in all markets, shocks of the same sign occur more frequently than shocks of different sign.

As it was mentioned in section 17.1, the volatility feedback hypothesis is one of the theories that try to explain the asymmetric volatility phenomenon. This theory relies on the existence of a positive intertemporal relation between expected returns and conditional variances, therefore, risk premiums are variable throughout the time. In the context of the conditional CAPM that we are considering, the relevant measure of risk for large firms is their variance whereas the relevant risk measure for small firms is their covariance with large firms. Therefore, in this work, the asymmetric volatility in small firms, will be determined by their covariance with large firms.

Table 17.7 shows the incremental effect (annualized and in percentage terms) on variances and covariances produced by a unitary shock in the large and/or small cap indices. Results are quite similar for the three countries. In this table, we can observe volatility and covariance asymmetry and volatility spillovers. Volatility of both indices increase more after negative shocks coming from any market than after positive shocks, with the exception of the SDAX volatility. Positive and negative shocks affecting returns on the SDAX produce the same effect on the SDAX volatility (remind that the estimated coefficient g_{22} was not significant). Covariance also increases more after negative shocks than after positive shocks. It drops in some cases when shocks are of opposite sign or positives. These results are consistent with the existence of time-varying risk premiums and, therefore, with the volatility feedback hypothesis.

If we focus on volatility spillovers, we observe that volatility spillovers between both type of firms, large and small firms, are bidirectional. Both, news coming from large firms and news coming from small firms affect the other market. However, it must be highlighted that news coming from the small cap index increase more the volatility of the large cap index than the opposite, especially, bad news. These volatility spillovers highlight that relevant information for portfolio management comes from the small firms market.

17.7 CONCLUSION

In this chapter we have investigated volatility spillovers between large firms and small firms in the French, German and British Stock Exchanges, taking into account volatility and covariance asymmetry. We use a conditional CAPM with an asymmetric GARCH-in-mean covariance structure that

Table 17.7 Sensitivity of volatility, covariance and risk premium to unitary shocks (1%)

Panel A: sensitivity of volatility, covariance and risk premium of the French indices to unitary shocks (1%)

	CAC40 shock		MIDCAC shock		Common shock			
	+	−	+	−	CAC40(+) MIDCAC(+)	CAC40(−) MIDCAC(−)	CAC40(+) MIDCAC(−)	CAC40(−) MIDCAC(+)
CAC40								
Volatility	1.6088	2.3847	2.5765	3.5083	0.9677	4.2529	4.8152	4.5406
Risk premium	0.0588	0.1291	0.1507	0.2794	0.0213	0.4106	0.5240	0.4659
MIDCAC								
Volatility	0.4161	0.7703	3.5291	5.5516	3.1130	5.8338	5.8252	3.9986
Covariance	0.0067	0.0181	0.0912	0.1934	0.0302	0.2347	0.2671	0.1765
Risk premium	0.0152	0.0412	0.2070	0.4391	0.0686	0.5329	0.6038	0.3990

Panel B: sensitivity of volatility, covariance and risk premium of the German indices to unitary shocks (1%)

	DAX shock		SDAX shock		Common shock			
	+	−	+	−	DAX(+) SDAX(+)	DAX(−) SDAX(−)	DAX(+) SDAX(−)	DAX(−) SDAX(+)
DAX								
Volatility	2.0876	2.7423	1.9311	2.1662	4.0187	4.8751	0.9941	1.7851
Risk premium	0.2545	0.4392	0.2178	0.2740	0.9432	1.3879	0.0577	0.1861
SDAX								
Volatility	0.3793	0.4181	3.3099	3.3101	3.6892	3.6954	2.9311	2.9362
Covariance	0.0074	0.0103	0.0604	0.0610	0.1384	0.1420	−0.0042	−0.0014
Risk premium	0.0432	0.0600	0.3527	0.3562	0.8083	0.8293	−0.0246	−0.0085

Continued

Table 17.7 Continued

Panel C: sensitivity of volatility, covariance and risk premium of the British indices to unitary shocks (1%)								
	FTSE shock		SMALL CAP shock		Common shock			
	+	−	+	−	FTSE(+) SMALL CAP(+)	FTSE(−) SMALL CAP(−)	FTSE(+) SMALL CAP(−)	FTSE(−) SMALL CAP(+)
FTSE								
Volatility	2.9695	3.0963	1.6182	5.0617	1.3514	5.8317	6.6371	4.6710
Risk premium	0.2425	0.2636	0.0720	0.7046	0.0502	0.9352	1.2070	0.5978
SMALL CAP								
Volatility	1.1891	3.0920	3.1174	3.2526	1.9282	4.2454	4.4058	5.1668
Covariance	0.0336	0.0551	0.0480	0.0859	−0.0281	0.1755	0.2421	0.2226
Risk premium	0.0923	0.1516	0.1319	0.2362	−0.0772	0.4827	0.6633	0.6100

Notes: The table shows the incremental effect produced in volatility and covariances by a unitary shock (1%) in the large and/or small index returns. All increases are annualized and in percentage terms. The headline of each column shows the sign of the shock.

accommodates both the sign and the magnitude of return innovations. Moreover, the volatility feedback effect is explored as a possible explanation for asymmetric volatility in stock returns.

We have obtained three important results. First, the estimated price of risk is positive and significant in the three countries. This result indicates that the risk is valued and is consistent with the volatility feedback hypothesis. If volatility is priced, an anticipated increase in volatility raises the required return on equity, leading to an immediate stock price decline. Therefore, time-varying risk premiums will be observed. Moreover, the asymmetric behavior of variances and covariances has been shown; both increase more after negative than after positive shocks.

Second, we find consistent evidence that volatility spillovers between large and small firms are bidirectional in the three countries, as Chuliá and Torró (2006) and Pardo and Torró (2005) find in the Spanish market. Moreover, news coming from the small cap index increases more the volatility of the large cap index than the opposite, especially, bad news. This result adds evidence against the common conclusion according to which volatility spillovers are unidirectional, from large firms to small ones, and shows that news on small firms can also cause volatility in their own returns and in large firm returns.

Finally, the study uncovers that conditional beta coefficient estimates within the model are insensitive to sign and size asymmetries in unexpected shock returns, and that the unconditional beta estimate has a significant specification error. Therefore, for dynamic portfolio management it is necessary to use conditional models in order to avoid specification errors on estimating beta coefficients.

REFERENCES

Banz, R. (1981) "The Relationship Between Return and Market Value of Common Stocks", *Journal of Financial Economics*, 9(1): 3–18.

Bekaert, G. and Wu, G. (2000) "Asymmetric Volatility and Risk in Equity Markets", *The Review of Financial Studies*, 13(1): 1–42.

Black, F. (1976) "Studies of Stock Price Volatility Changes", *Proceedings of the 1976 Meetings of the American Statistical Association*, Business and Economical Statistics Section.

Bollerslev, T. (1990) "Modelling the Coherence in Short-Run Nominal Rates: A Multivariate Generalized ARCH Approach", *Review of Economics and Statistics*, 72(3): 498–505.

Bollerslev, T. and Wooldridge, J.M. (1992) "Quasi-maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances", *Econometric Reviews*, 11(2): 143–72.

Bollerslev, T., Engle, R.F. and Wooldridge, J.M. (1988) "A Capital Asset Pricing Model with Time Varying Covariances", *Journal of Political Economy*, 96(1): 116–31.

Braun, P.A., Nelson, D.B. and Sunier, A.M. (1995) "Good News, Bad News, Volatility and Betas", *The Journal of Finance*, 50(5): 1575–603.

Campbell, J.Y. (1987) "Stock Returns and the Term Structure", *Journal of Financial Economics*, 18(4): 373–99.

Campbell, J.Y. and Henstschel, L. (1992) "No News Is Good News: An Asymmetric Model of Changing Volatility in Stock Returns", *Journal of Financial Economics*, 31(2): 281–318.
Chan, K.C., and Chen, N. (1991) "Structural and Return Characteristics of Small and Large Firms", *Journal of Finance*, 46(4): 1467–84.
Christie, A.A. (1982) "The Stochastic Behaviour of Common Stock Variances-Value, Leverage and Interest Rate Effects", *Journal of Financial Economics*, 10(4): 407–32.
Chuliá H. and Torró, H. (2006) "Asimetrías en Volatilidad, Beta y Contagios entre las Empresas Grandes y Pequeñas Cotizadas en la Bolsa Española", FUNCAS Working Paper, 260.
Conrad, J., Gultekin, M.N. and Kaul, G. (1991) "Asymmetric Predictability of Conditional Variances", *Review of Financial Studies*, 4(4): 597–622.
Dickey, D.A. and Fuller, W.A. (1981) "Likelihood Ratio Statistics for Autorregressive Time Series with a Unit Root", *Econometrica*, 49(4): 1057–72.
Engle, R.F. (1982) "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflation", *Econometrica*, 50(4): 987–1008.
Engle, R.F. and Kroner, K.F. (1995) "Multivariate Simultaneous Generalized ARCH", *Econometric Theory*, 11(1): 122–50.
Engle, R.F. and Ng, V.K. (1993) "Measuring and Testing the Impact of News on Volatility", *Journal of Finance*, 48(5): 1749–78.
French, K.R., Schwert, G.W. and Stambaugh, R.F. (1987) "Expected Stock Returns and Volatility", *Journal of Financial Economics*, 19(1): 3–29.
Glosten, L.R., Jagannathan, R. and Runkel, D.E. (1993) "On the Relation Between the Expected Value and Volatility of Nominal Excess Return on Stocks", *Journal of Finance*, 48(5): 1779–801.
Hamilton, J.D. (1994) *Time Series Analysis* (Princeton, NJ: Princeton University Press).
Hendry, O.L. and Sharma, J. (1999) "Asymmetric Conditional Volatility and Firm Size: Evidence from Australian Equity Portfolios", *Australian Economic Papers*, 38(4): 393–406.
Hentschel, L. (1995) "All in the Family Nesting Symmetric and Asymmetric GARCH Models", *Journal of Financial Economics*, 39(1): 71–104.
Kroner, K.F., and Ng, V.K. (1998) "Modelling Asymmetric Co-movements of Asset Returns", *Review of Financial Studies*, 11(4): 817–44.
Lo, A.W. and Mackinlay, A.C. (1990) "When Are Contrarian Profits Due to Stock Market Overreaction?", *Review of Financial Studies*, 3(2): 175–205.
Officier, R.R. (1973) "The Variability of the Market Factor of the New York Stock Exchange", *Journal of Business*, 46(3): 434–53.
Nelson, D.B. (1992) "Filtering and Forecasting with Misspecified ARCH Models I: Getting the Right Variance with the Wrong Model", *Journal of Econometrics*, 52(1): 347–70.
Pardo, A. and Torró, H. (2005) "Trading with Asymmetric Volatility Spillovers", University of Valencia Working Paper (http://ssrn.com/abstract=427081).
Petrella, G. (2005) "Are Euro Area Small Cap Stocks an Asset Class? Evidence from Mean–Variance Spanning Tests", *European Financial Management*, 11(2): 229–53.
Phillips, P.C.B. and Perron, P. (1988) "Testing for a Unit Root in Time Series Regression", *Biometrica*, 75(2): 335–46.
Pindyck, R.S. (1984) "Risk, Inflation and the Stock Market", *American Economic Review*, 74(3): 334–51.
Ross, S.A. (1989) "Information and Volatility: The No-Arbitrage Martingale Approach to Timing and Resolution Irrelevancy", *Journal of Finance*, 44(1): 1–17.
Schwert, G.W. (1989) "Why Does Stock Market Volatility Change Over Time?", *Journal of Finance*, 44(5): 1115–53.
Wooldridge, J.M. (1990) "A Unified Approach to Robust, Regression-Based Specification Tests", *Econometric Theory*, 6(1): 17–43.
Wu, G. (2001) "The Determinants of Asymmetric Volatility", *Review of Financial Studies*, 14(3): 837–59.

CHAPTER 18

On Model Selection and its Impact on the Hedging of Financial Derivatives

Giuseppe Di Graziano and Stefano Galluccio

18.1 INTRODUCTION

The mathematical theory of derivatives pricing and risk-management is one of the most active fields of research for both academics and practitioners. The celebrated Black–Scholes–Merton (BS) pioneering work paved the way to the development of a general theory of option pricing through the concept of absence of market arbitrage and dynamic replication (Harrison and Pliska, 1981). As is well-known, the simplistic assumptions behind the BS model make it unsuitable to capture and explain the risk borne by complex (exotic) financial derivatives. The need for a departure from the BS paradigm is in fact evident from the analysis of historical time series (Bates, 1996; Pan, 2002; Chernov, Gallant, Ghysels and Tauchen, 2003; and Eraker, Johannes and Polson, 2003, among others), as well as from the observation of the volatility smile phenomenon (Heston, 1993; Dupire, 1994, among others). For these reasons a number of alternative models have been advocated by many authors. Roughly speaking, all dynamic arbitrage-free models aiming at generalizing BS theory can be divided in three main classes, according to the characteristics of the stochastic process driving the dynamics of the underlying assets.

Deterministic or *local volatility* (LV) models assume that the asset follows an Itô diffusion with a volatility function that is a deterministic (non-linear) function of the underlying state variables. Derman and Kani (1994) and, in particular, Dupire (1994) succeeded in showing that, if noise is driven by a single Brownian motion, it is possible to uniquely determine the form of volatility function out of the market prices of traded options. This approach is often referred to as "implied volatility theory".

An alternative, more sophisticated framework consists of assuming that the volatility itself is driven by an independent noise source, usually a Wiener process. This approach, known as *stochastic volatility* (SV) modeling, was introduced and developed by Hull and White (1987), Stein and Stein (1991) and Heston (1993) among others. SV models have recently become very popular in the industry since they provide a simple (yet satisfactory) approach to the modeling of the smile dynamics while LV models are less appealing from this point of view (Andersen and Andreasen, 2000a).

Finally, a generalization of BS can be achieved by relaxing the assumption of diffusive-type dynamics and by assuming the presence of jumps in the asset, in the volatility or in both. From a pure statistical perspective, *jump-diffusion* (JD) models (possibly with the inclusion of stochastic volatility features, or SVJD models) are in excellent agreement with empirical observations, as shown for instance in Eraker, Johannes and Polson (2003). Unfortunately, pricing and hedging in presence of jumps is in general a much harder task to achieve than in pure diffusion models (Duffie, Pan and Singleton, 2000; Föllmer and Schweizer, 1991).

Despite the differences among these three approaches, as some authors have observed (see for instance Schoutens, Simons and Tistaert, 2003), they are all capable of reproducing the observed shape of the implied volatility surface.[1] In fact, one can always parametrically adjust a given model to enforce its unconditional probability distributions to be very "close" to a pre-assigned marginal distribution to the point that the resulting differences would be indistinguishable from a practical point of view. As is well-known, however, two processes with different conditional distributions give rise to totally different sample paths even if their unconditional distributions are the same. Therefore, one might argue that the price of certain path-dependent options and, more importantly, the dynamic risk-management are both heavily affected by model selection. This is indeed the case as the results of this chapter suggest.

This remark is particularly relevant in the context of model implied calibration, a common and well-known market practice. To avoid potential arbitrage opportunities and identify the market price of risk practitioners "calibrate" their models to the market prices of vanilla European options (i.e. the smile) and use the resulting dynamics to evaluate, by arbitrage, more complex derivatives whose prices are not directly available in the market. In doing so, one implicitly assumes that the market contains enough "information" to uniquely identify the model and its associated risk-neutral pricing measure. However, the market prices of European vanilla options do not contain any information about the conditional distributions of the underlying stochastic process. As the above argument suggests, unique model identification is an impossible task to achieve and for any model choice there exists an associated *model risk*. Although many different definitions of model risk are possible (see Cont, 2005, for a review) here we refer to model risk as the residual uncertainty on the price of financial derivatives and on their risk-management once that *all* relevant market information has been properly included in the pricing model (through model calibration).

In this chapter we examine the impact of model choice on the hedging of financial derivatives in a simplified set-up. In particular, we will assume that the true dynamics of the underlying asset (the stock) follows a given stochastic volatility process while market agents, lacking this information, trade according to a different model. In order to guarantee market consistency, we will further require that all available market information be captured by the "wrong" model. This will be achieved by calibrating the model to all option market prices, which we generate using the "right" model. We shall address the hedging error problem both analytically and numerically and show that using the "wrong" model for risk management purposes can generate significant replication errors even for short maturity options and despite the fact that the market smile is almost perfectly matched by trader's model. Needless to say, we do not expect to provide a good representation of real world asset dynamics; our study is meant to gather some intuition about model errors in risk management. In particular, our results suggest that model selection must take into account historical information to obtain a proper representation of the real asset dynamics.

The rest of the chapter is organized as follows. Section 18.2 introduces the model and the notation, while section 18.3 is devoted to the computation of the total hedging error in models with stochastic volatility. Numerical tests are performed in section 18.4. Finally, section 18.5 draws some conclusions and perspectives for future research.

18.2 MODEL AND MATHEMATICAL SETUP

We shall assume that the market consists of a single (non-dividend-paying) underlying asset S_t whose P-dynamics are driven by a two dimensional Brownian motion $B = (B^1, B^2)$ defined on a probability space (Ω, F, P), where $F = (F_t)_{t \geq 0}$ is the natural filtration of B and P the "physical" probability measure. In addition, we postulate the existence of a money market account $\beta_t = \exp(rt)$, where r indicates the constant riskless interest rate.[2] Stochastic volatility models, like the ones considered in this paper, are intrinsically incomplete. It is however possible to "complete" the market by adding a number of tradable (non-redundant) securities, typically liquid vanilla options. In what follows, we will assume that agents are allowed to trade an additional option, say D, on S_t to risk manage their books. The latter assumption, together with the no arbitrage condition and the market price of volatility risk, uniquely determines the risk neutral probability measure P^* under which discounted asset prices are martingales. For convenience we will use β_t as numéraire. The corresponding Wiener process under P^* will be denoted by $W = (W^1, W^2)$. We shall further assume that the covariance process is given by $\langle W^1, W^2 \rangle = \rho t$ for a given choice of the constant correlation coefficient ρ.

In order to introduce the concept of hedging error, we define two different dynamics for S_t under P^*, specifically:

$$\begin{aligned} dS_t &= rS_t dt + \gamma(t, S, v) dW_t^1 \\ dv_t &= \varphi(t, v) dt + \varphi(t, v) dW_t^2 \end{aligned} \tag{18.1}$$

and

$$\begin{aligned} dS_t &= rS_t dt + \Gamma(t, S, v) dW_t^1 \\ dv_t &= \Phi(t, v) dt + \Theta(t, v) dW_t^2 \end{aligned} \tag{18.2}$$

We will also assume that the generic functions γ, φ, ϑ, Γ, Φ, Θ satisfy the usual conditions of existence and unicity to the solutions of the above SDE's (see for instance Jacod and Shiryaev (1987)). In our set-up, equation (18.1) represents the real (unknown to the trader) dynamics of S_t, while equation (18.2) corresponds to the model used by traders as a proxy of the "true" market dynamics. We also assume that the correlation between Wiener noises in the "wrong" model (18.2) is $\rho' \neq \rho$. Our goal is to estimate the hedging error that traders incur when using model (18.2) to risk-manage a vanilla European option written on S_t.

It should be noted that our modeling setup is fairly generic as many well-known models can be recovered by appropriately specifying the functions $\gamma, \varphi, \vartheta, \Gamma, \Phi, \Theta$. For instance, the following special cases are well-known:

1 If $\gamma(t, S, v) = vS_t$, $\varphi = \vartheta = 0$ we recover the Black–Scholes-model.

2 If $\gamma(t, S, v) = \sqrt{v_t} S_t$, $\varphi_t = a(b - v_t)$ and $\vartheta = \alpha\sqrt{v_t}$ for given constants a, b, α we recover the Heston (1993) stochastic volatility model (in this case v_t is the square of the volatility, i.e. the instantaneous variance).

3 If $\gamma(t, S, v) = v_t S_t$, $\varphi_t = a(b - v_t)$ and $\vartheta = \alpha$ for given constants a, b, α we recover the Stein–Stein (1991) stochastic volatility model.

It is in principle possible to extend the results presented in section 18.3 to more general processes (in particular to include the presence of jumps in the state variables) but the mathematical setup needed is more involved and it will be addressed in a separate paper (Di Graziano and Galluccio, 2005).

18.3 ANALYTICAL EXPRESSION OF THE TOTAL HEDGING ERROR

Consider the following situation: a trader sells at time t a European call option struck at K expiring at $T > t$, with payoff $C(S_T) = (S_T - K)^+$. We will first assume that the agent believes the BS model to be correct. In order to hedge his exposure, he will form a dynamic portfolio consisting of $\Delta_{t'}$ shares, where $\Delta_{t'} = \frac{\partial C}{\partial S}|_{t=t'}$ is the option "Delta", and invest an amount $\psi_{t'} = \frac{C_{t'} - \Delta_{t'} S_{t'}}{\beta_t}$ in the money market account for $t \leq t' \leq T$. If the "true" market dynamics were lognormal with constant coefficient as postulated in the BS theory, the agent would be able to perfectly replicate the option price and the hedging error would be identically zero (in a frictionless market). In general, the replication error incurred by using the simple delta hedging strategy does not vanish since, in particular, asset prices do not follow lognormal processes.

In two separate works Carr and Madan (1997) and, independently, El Karoui, Jeanblanc and Shreve (1997) proved that if the asset follows a continuous diffusion process (with possibly time-dependent, adapted coefficients) then the simple delta hedging strategy based on the BS model would result in a total error:

$$Y_{t,T} = \int_t^T \beta_{T-s} \frac{S_s^2}{2} \frac{\partial^2 C_s}{\partial S^2} (\sigma_{BS}^2 - \sigma(s)^2) ds \tag{18.3}$$

at time T, where σ_{BS} is the BS volatility corresponding to the price at which the trader bought the option and $\sigma(t)$ is the actual (or "realized") volatility process. The above fundamental formula indicates that the total hedging error is a function of:

1 The difference of the squares of the volatilities. In particular the trader realizes a gain if the realized volatility path is below the BS volatility and to lose money in the opposite case.

2 The option's Gamma $\frac{\partial^2 C_S}{\partial S^2}$. *Ceteris paribus*, the larger the Gamma the more pronounced is the hedging error.

Although this result is strikingly simple and intuitive, the hedging strategy is too simple to be realistic. For this reason we provide here a different formula that postulates a "true" dynamics for the underlying asset as in equation (18.1) and a hedging model at the trader disposal as in equation (18.2). Moreover, we will also allow market agents to hedge their volatility exposure. For the sake of simplicity we will limit ourselves to provide the error formula and indicate a sketch of its proof. The interested reader will find all details (in a more general setting) in Di Graziano and Galluccio (2005).

The main idea of the proof is the following. At any time $t \le T$, build a portfolio V_t consisting of an amount $\psi_t^{(1)}$ of stocks, $\psi_t^{(2)}$ units in the money market account and a quantity $\psi_t^{(3)}$ of another option D_t (which is used to complete the market). At inception, the trader observes a set of market option prices and calibrates what he believes to be the right model (18.2) to those prices. We assume, for the sake of simplicity, that all relevant option prices are well reproduced by the model (18.2). Assume that the option price function associated to the trader's "wrong" model is given by $C_t = C(S_t, v_t, t)$. He will then choose coefficients $\psi_t^{(i)}$ in V_t so as to make the process Y_t:

$$Y_t = \psi_t^{(1)} S_t + \psi_t^{(2)} \beta_t + \psi_t^{(3)} D_t - C_t$$

vanish at any time when S_t and v_t satisfy equation (18.2). In particular, the amount of stocks and options held in the portfolio will be given by:

$$\psi_t^{(1)} = \frac{\partial C}{\partial S} - \frac{\partial C}{\partial v}\left(\frac{\partial D}{\partial v}\right)^{-1}\frac{\partial D}{\partial S}, \qquad \psi_t^{(3)} = \frac{\partial C}{\partial v}\left(\frac{\partial D}{\partial v}\right)^{-1} \tag{18.4}$$

respectively. Note that the hedging ratios involve only price sensitivities calculated by using the "wrong" model. If the real asset dynamics where given by (18.2), then Y_t would be identically zero. In general this is not true since equation (18.1) (as opposed to equation (18.2)) describes the "true" asset dynamics. By applying Ito's formula toY_t, by using the correct asset dynamics (18.1) for S_t and v_t, and by remembering that option prices generated with model (18.1) are matched by model (18.2), we finally arrive at the following expression of the total hedging error (Di Graziano and Galluccio, 2005):

$$\begin{aligned} Y_{t,T} = &-\int_t^T \beta_{T-s}\frac{\partial C_S}{\partial v}\left(\frac{\partial D_S}{\partial v}\right)^{-1}\left[\frac{1}{2}(\Gamma_s^2-\gamma_s^2)\frac{\partial^2 D_s}{\partial S^2}+\frac{1}{2}(\Theta_s^2-\vartheta_s^2)\frac{\partial^2 D_s}{\partial v^2}\right.\\ &\left.+(\rho'\Gamma_s\Theta_s-\rho\gamma_s\vartheta_s)\frac{\partial^2 D_s}{\partial S\partial v}\right]ds\\ &+\int_t^T \beta_{T-s}\left[\frac{1}{2}(\Gamma_s^2-\gamma_s^2)\frac{\partial^2 C_s}{\partial S^2}+\frac{1}{2}(\Theta_s^2-\vartheta_s^2)\frac{\partial^2 C_s}{\partial v^2}\right.\\ &\left.+(\rho'\Gamma_s\Theta_s-\rho\gamma_s\vartheta_s)\frac{\partial^2 C_s}{\partial S\partial v}\right]ds \end{aligned} \tag{18.5}$$

This formula is the extension of equation (18.3) in a set-up given by equations (18.1) and (18.2).

To shed some light on our result, we first notice that the total hedging error is made of two contributions: one integral is associated to the option to hedge C, the other to the additional claim D in the replicating portfolio. Each of them is in turn composed of three additive terms that can be identified as follows. One term is proportional to the Gamma of the option, i.e. the second derivative of the claim respect to the asset, $\frac{\partial^2 C_s}{\partial S^2}$. This is analogous to the term in equation (18.3) but, differently from that simplified case, our result shows that the actual Gamma-induced error depends on the difference of the squares of the volatility functions $\Gamma(t, S, v)$ and $\gamma(t, S, v)$. In particular, different functional assumptions result in completely different hedging errors. A second term is proportional to the option's "Volga", that is to the second derivative with respect to volatility, $\frac{\partial^2 C_s}{\partial v^2}$. Finally, the third term is proportional to the option's "Vanna", that is the mixed second derivative with respect to asset and volatility, $\frac{\partial^2 C_s}{\partial S \partial v}$. Each of these terms multiplies a factor that, roughly speaking, depends on the difference between "true" and "false" model. Thus, any such term can take both positive and negative values. This last observation is crucial in the applications (section 18.4) since it clearly indicates that if the characteristics Γ, Φ, Θ of the wrong model are biased respect to those of the real one γ, φ, ϑ then the total hedging error assumes largely positive or negative values.

As an important remark, we notice that the first integrand in equation (18.5) is proportional to the ratio of the option's Vega $\frac{\partial C_s}{\partial v}\left(\frac{\partial D_s}{\partial v}\right)^{-1}$. Interestingly, this provides a useful hint on how to optimally select D: one should better choose a claim whose Vega is close to that of the option to hedge C so that $Y_{t,T}$ in equation (18.5) is not systematically biased.

As shown in Schoutens, Simons and Tistaert (2003) and in section 18.4, the fact that the model is made consistent with the relevant set of market option prices is not enough to identify the characteristics of the process and therefore a residual hedging error is always present. In the next section we perform (based on equation (18.5)) a number of empirical tests to quantitatively assess the replication error in two benchmark scenarios.

18.4 NUMERICAL RESULTS

In order to give a quantitative estimate of the hedging error in practical situations we perform two independent empirical tests. In the first we aim at measuring the impact of *model misspecification* on the replicating strategy, while the latter is aimed at estimating the impact of *parameters misspecification*. We say that a model is misspecified if the functional form of the characteristics Γ, Φ, Θ is different from that of the true set γ, φ, ϑ. A typical example is that of a trader that uses the BS model to hedge a contingent claim written on an asset that follows a stochastic volatility process. Similarly, we say that model parameters are misspecified if the functional form of the characteristics Γ, Φ, Θ is correct but the numerical coefficients are different from the corresponding ones in γ, φ, ϑ. The latter situation arises when, for instance, the trader has the right intuition and uses a stochastic volatility model but he assumes that the Brownian motions are uncorrelated. Interestingly, a number of papers have recently appeared where the assumption of zero correlation is essentially justified as it makes the model analytically tractable (Andersen and Andreasen, 2000b; Piterbarg, 2003). Our simple example below demonstrates that, even if a model with no correlation is made consistent with the smile, the residual hedging error can be significant if the actual dynamics is driven by correlated Brownian motions.

We remind the reader that from a mathematical point of view there is no difference between the two kinds of model risk, as correctly pointed out by Cont (2005). However, our empirical tests show that the errors originating from model misspecification are generally much larger and then, potentially, much more dangerous.

To fix the ideas, we specialize equation (18.1) to the following case $\gamma(t, S_t, v_t) = S_t^x \sqrt{v_t}$, $\varphi(t, S_t, v_t) = a(b - v_t)$, $\vartheta(t, S_t, v_t) = \alpha\sqrt{v_t}$ with $x \geq 0$. This model corresponds to a Heston model with CEV-type local volatility and can be analytically handled if $\rho = 0$ (Di Graziano and Galluccio, 2005). In terms of its market explicative power our framework is qualitatively similar to the celebrated SABR model (Hagan, Kumar, Lesniewski and Woodward, 2002) but has the advantage, over SABR, of possessing a mean-reverting instantaneous variance process, which is beneficial in the applications (Galluccio and Le Cam, 2005). In any case, we point out that the qualitative picture emerging from our empirical tests is not affected by the fine details of the model specification equation (18.1).

As anticipated, equation (18.1) will be our benchmark "model"; that is, it will be assumed to represent the actual asset dynamics. We thus generate (by Monte-Carlo simulation) a set of benchmark call option prices using equation (18.1) at different strikes that are meant to approximate to a good degree of accuracy the actual smile observed in the market (for example the S&P index).[3]

In our first study we postulate that the trader believes that BS is the "right" model and risk-manages his hedging portfolio accordingly. He then sells at $t = 0$ a 1 year call struck at $K = 0.975$ when $S = 1$ and enters into a standard delta-hedging self-financing strategy. We simulate the actual path followed by the asset from equation (18.1) and, at expiry $T = 1$, we record the hedging error $Y_{t,T}$. In Figure 18.1 we show the empirical probability distribution of the random variable $Y_{t,T}$.

Two things are worth noticing. First, the average total error is positive. This is mainly a consequence of the particular choice of the parameters used to generate the "true" smile. A different choice of parameters (consistent with the price of the call) can lead to the opposite bias. Second, the result suggests that if one risk manages the portfolio using a simple BS delta hedging argument, then there is a non-negligible probability to get a total error that is a significant fraction of the total profit from the option.

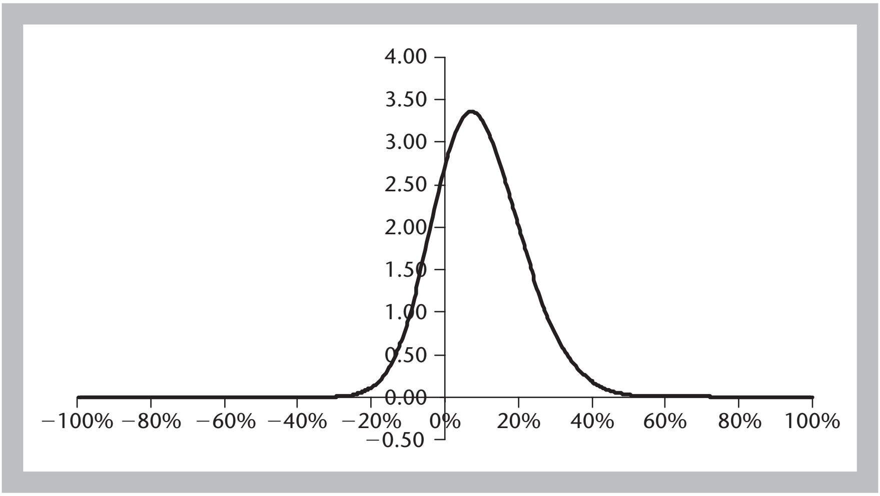


Figure 18.1 Hedging error probability density function in the B–S case. The "real" asset dynamics is given by equation (18.1). All errors are expressed as a percentage cost of the option

Table 18.1 Evaluation of the hedging errors

	Model parameters					
	r	a	b	α	x	ρ
True model	0.02	0.508	0.151	0.4	0.8	−0.2
Trader's model	0.02	0.534	0.75	0.201	0.65	0

Notes: The table shows the values of the parameters used to evaluate the hedging errors. "True" model parameters have been selected to provide a realistic shape of the smile. Trader's model parameters are obtained by minimizing the difference between the smiles generated by the two models. Calibration errors are within the typical implied volatility bid/ask spread of 1%.

This example clearly demonstrates that if the trader's model is a bad representation of reality, the hedging error can be very large.

As a second test, we now assume that the trader uses a stochastic volatility model to hedge the contingent claim which is structurally similar to the "real" one but he has a bad assessment of some parameters. In this case, equation (18.5) holds. In particular he believes that correlation is zero; that is $\rho' = 0$, while in reality $\rho \neq 0$. In our example, we fixed $\rho = -20\%$ and to ensure that all market information is captured we calibrate model (18.2) to the smile generated by equation (18.1). In this procedure we determine the unknown parameters by minimizing the squared difference of option prices between the two models for different values of the strike, following a standard practice. The resulting errors are within the typical bid–ask spread and amount to a fraction of a percent in log-normal units (Table 18.1).

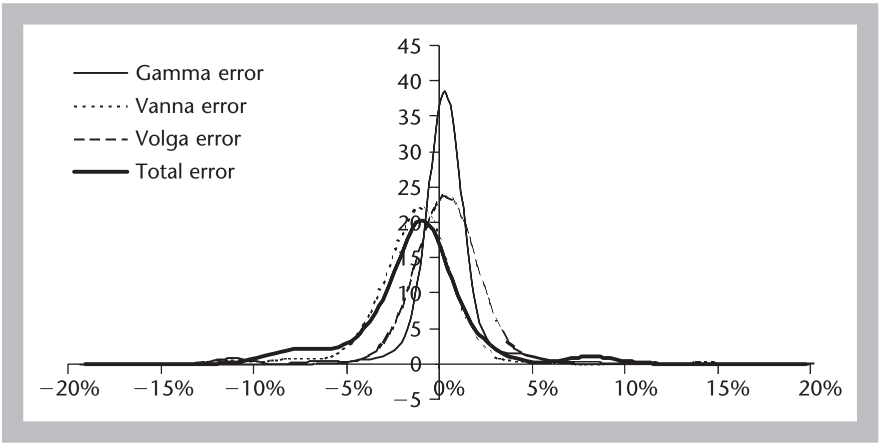


Figure 18.2 The hedging error probability density function obtained by assuming that the additional claim has the same expiry (1 year) as the option to hedge but different strike ($K = 1$). The trader uses a stochastic volatility model. All errors are expressed as a percentage of the initial cost of the option

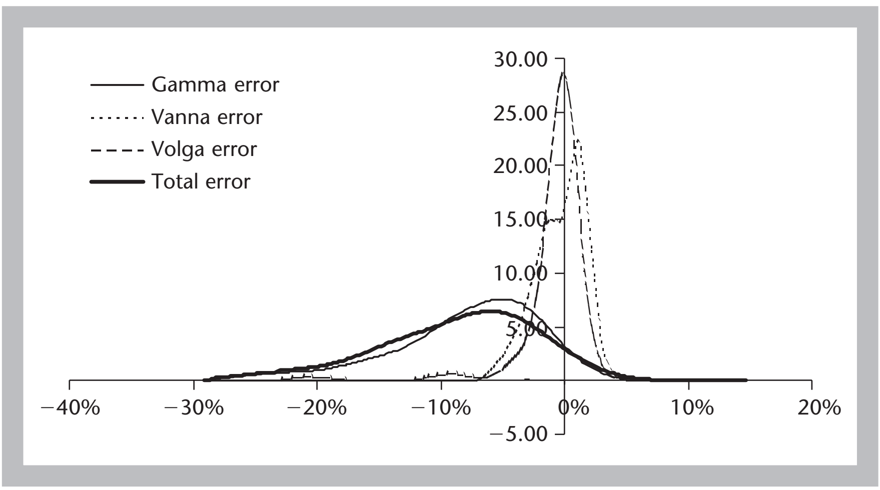


Figure 18.3 The hedging error probability density function obtained by assuming that the additional claim has the same strike ($K = 0.975$) as the option to hedge but different expiry (2 years). The trader uses a stochastic volatility model. All errors are expressed as percentage of the initial cost of the option

We study two different cases. First we assume that we choose for D a liquid at-the-money option that has the same expiry as the option to hedge (i.e. a one-year call). In Figure 18.2 we show the hedging error coming from the Gamma, the Volga and the Vanna terms separately, as well as the total error.

We immediately notice that large errors are now less likely when compared to the BS delta hedging strategy. In addition, the largest contribution to the total error comes from the two Vanna terms since they are the most affected by the wrong assumption on correlation.

As a second example, we consider the situation where D, like C, is an option struck at $K = 0.975$ but it expires one year later, i.e. $T = 2$. Results are shown in Figure 18.3. We now see that the largest contribution to the total error comes from the Gamma terms since the two options have very different Gamma in this case. However, all errors are still much smaller than in the simple BS case.

CONCLUSION

In this chapter we have examined the errors arising from hedging a contingent claim written on a tradable asset that follows a generic stochastic volatility model when traders have a bad assessment of the real dynamics. We provide a general formula for the total hedging error extending some known results to the case of state variables driven by stochastic volatility processes. We have numerically shown that errors due to a bad representation of the whole dynamics are significantly larger than those arising from just a bad estimation of model parameters, in general. However, even if the trader uses a model that is formally equivalent to the true one, errors due to parameters misspecification can still be quite large. This in particular should generate some concern when hedging is performed with a model that assumes no correlation between Wiener noises, a framework that has recently gained some favor in the market due to its mathematical tractability.

NOTES

1. It must be noticed, however, that in general one needs to artificially assume strong time-dependency of model parameters to achieve a good model "calibration", i.e. to ensure that market vs. model errors are within the bid–ask spread (Galluccio and Le Cam, 2005). Despite this, at least theoretically it is possible to well-approximate any smile shape by properly adjusting the postulated dynamics if coefficients are allowed to take arbitrary values.
2. We will assume throughout the paper that interest rates are deterministic. Extending the present approach to include the effect of stochastic rates is possible but results are essentially unaffected by this choice in the range of options expiry we consider.
3. We recall that we do not address here the question of whether equation (18.1) is the correct representation of reality, thus we do not need to calibrate our "benchmark" model to the S&P market. Instead, *we assume* that equation (18.1) is the true model and study the replication error induced by taking equation (18.2) as a good approximation of the market, described by equation (18.1).

REFERENCES

Andersen, L. and Andreasen, J. (2000a) "Jump-Diffusion Processes: Volatility Smile Fitting and Numerical Methods for Option Pricing", *Review of Derivatives Research*, 4(3): 231–62.

Andersen, L. and Andreasen, J. (2000b) "Volatility Skews and Extensions of the LIBOR Market Model", *Applied Mathematical Finance*, 7(1): 1–32.

Bates, D. (1996) "Jumps and Stochastic Volatility: Exchange Rates Processes Implicit in Deutsche Mark Options", *Review of Financial Studies*, 9(1): 69–107.

Carr, P. and Madan, D. (1997) "Towards a Theory of Volatility Trading", in R. Jarrow (ed.), *Volatility*, pp. 417–27.

Chernov, M., Gallant, R., Ghysels, E. and Tauchen, G. (2003) "Alternative Models for Stock Price Dynamics", *Journal of Econometrics*, 116(1): 225–57.

Cont, R. (2005) "Model Uncertainty and Its Impact on the Pricing of Derivative Instruments", *Mathematical Finance*, forthcoming.

Derman, E. and Kani, I. (1994) "Riding on a Smile", *Risk Magazine*, February: 32–9.

Di Graziano, G. and Galluccio, S. (2005) "Evaluating Hedging Errors for General Processes and Applications", Working Paper, BNP Paribas.

Duffie, D., Pan, J. and Singleton, K. (2000) "Transform Analysis and Asset Pricing for Affine Jump-Diffusions", *Econometrica*, 68(6): 1343–76.

Dupire, B. (1994) "Pricing with a Smile", *Risk Magazine*, January: 18–20.

Galluccio, S. and Le Cam, Y. (2005) "Implied Calibration of Stochastic Volatility Jump-Diffusion Models", Working Paper, BNP Paribas.

El Karoui, N., Jeanblanc, M. and Shreve, S. (1997) "Robustness of Black and Scholes Formula", *Mathematical Finance*, 8(2): 93–126.

Eraker, B., Johannes, M. and Polson, M. (2003) "The Impact of Jumps in Volatility and Returns", *Journal of Finance*, 58(3): 1269–300.

Föllmer, H. and Schweizer, M. (1991) "Hedging of Contingent Claims Under Incomplete Information", in M.H.A. Davis and R.J. Elliott (eds), *Applied Stochastic Analysis*, pp. 389–414 (New York, NY: Gordon & Breach).

Hagan, P., Lesniewski, A., Kumar, D. and Woodward, D. (2002) "Managing Smile Risk", *Wilmott Magazine*, September: 84–108.

Harrison, J.M. and Pliska, S.R. (1981) "Martingales and Stochastic Integrals in the Theory of Continuous Trading", *Stochastic Processes and Their Applications*, 11(3): 215–60.

Heston, S. (1993) "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options", *Review of Financial Studies*, 6(2): 327–43.

Hull, J. and White, A. (1987) "The Pricing of Options on Assets with Stochastic Volatility", *Journal of Finance*, 42(2): 281–300.

Jacod, J. and Shiryaev, A.N. (2003) *Limit Theorems for Stochastic Processes*, 2nd edn (Berlin: Springer-Verlag).

Pan, J. (2002) "The Jump Risk Premia Implicit in Options: Evidence from an Integrated Time Series Study", *Journal of Financial Econometrics*, 63(1): 3–50.

Piterbarg, V. (2003) "A Stochastic Volatility Forward LIBOR Model with a Term Structure of Volatility Smiles", Working paper, Bank of America.

Schoutens, W., Simons, E. and Tistaert, J. (2004) "A Perfect Calibration! Now What?", *Wilmott Magazine*, March: 85–98.

Stein, E.M. and Stein, J.C. (1991) "Stock Price Distributions with Stochastic Volatility: An Analytic Approach", *Review of Financial Studies*, 4(4): 727–52.

Index