McGRAW-HILL FINANCE & INVESTING

THE RISK MODELING EVALUATION HANDBOOK

RETHINKING FINANCIAL RISK MANAGEMENT METHODOLOGIES IN THE GLOBAL CAPITAL MARKETS

EDITED BY

GREG N. GREGORIOU, CHRISTIAN HOPPE, AND CARSTEN S. WEHN

Advance Praise for
The Risk Modeling Evaluation Handbook

A book like this helps reduce the chance of a future breakdown in risk management.

—**Campbell R. Harvey,** Professor
The Fuqua School of Business, Duke University

Inadequate valuation and risk management models have played their part in triggering the recent economic turmoil felt around the world. Model risk is thus becoming recognized by risk managers and financial engineers as an important source of additional risk. This timely book, written by experts in the field, will surely help them to measure and manage this risk effectively.

—**Fabrice Douglas Rouah, Ph.D.,** Vice President
Enterprise Risk Management

The Risk Modeling Evaluation Handbook provides a very timely and extremely useful guide to the subtle and often difficult issues involved in model risk—a subject which is only now gaining the prominence it should always have had. Risk practitioners will find it an invaluable guide.

—Kevin Dowd, Professor of Financial Risk Management
Nottingham University Business School

This book collects authorative papers on a timely and important topic written by academics and practitioners. Especially the latter combination makes this book readable to a wide audience, and it should lead to many new insights.

—Philip Hans Franses, Professor of Econometrics and Dean
Erasmus School of Economics, Erasmus University Rotterdam

This invaluable handbook has been edited by experts, with topical contributions on modeling risk, equity and fixed income investments, superannuation funds, asset returns, volatility, option pricing, credit derivatives, equity derivatives, valuation models, expected shortfall, value at risk, operational risk, economic capital, public debt management, financial crises, and political risk. The excellent chapters have been written by leading academics and practitioners, and should prove to be of great value to investment finance and credit risk modelers in a wide range of disciplines related to portfolio risk management, risk modeling in finance, international money and finance, country risk, and macroeconomics.

—Michael McAleer, FASSA, FIEMSS, Professor of Quantitative Finance, Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam; Research Fellow, Tinbergen Institute; Distinguished Chair and Professor, Department of Applied Economics, National Chung Hsing University

This book gives an up-to-date, comprehensive overview of the latest developments in the field of model risk, using state-of-the-art quantitative techniques.

—Ben Tims, Assistant Professor of Finance
Erasmus School of Management, Erasmus University Rotterdam

[T]he previous years have shown that too many capital market experts have blindly trusted their models. This comprehensive compendium addresses all the relevant aspects of model risks which helps practitioners to mitigate the probability of future financial crisis.

—Ottmar Schneck, Professor
European School of Business, Reutlingen

THE RISK MODELING EVALUATION HANDBOOK

This page intentionally left blank

THE RISK MODELING EVALUATION HANDBOOK

Rethinking Financial Risk Management Methodologies in the Global Capital Markets

GREG N. GREGORIOU
CHRISTIAN HOPPE
CARSTEN S. WEHN

EDITORS

McGraw Hill

New York Chicago San Francisco
Lisbon London Madrid Mexico City
Milan New Delhi San Juan Seoul
Singapore Sydney Toronto

The McGraw-Hill Companies

ISBN: 978-0-07-166371-7

MHID: 0-07-166371-1

The material in this eBook also appears in the print version of this title: ISBN: 978-0-07-166370-0, MHID: 0-07-166370-3.

The editors wish to thank the anonymous academic referees who gave generously of their time in reviewing each manuscript and recommending which papers should be included in this book.

CONTENTS

This page intentionally left blank

ABOUT THE EDITORS

Greg N. Gregoriou has published 34 books, over 50 refereed publications in peer-reviewed journals, and 22 book chapters since his arrival at SUNY (Plattsburgh) in August 2003. Professor Gregoriou's books have been published by John Wiley & Sons, McGraw-Hill, Elsevier-Butterworth/Heinemann, Taylor & Francis/Chapman-Hall/CRC Press, Palgrave-Macmillan, and Risk/Euromoney books. His articles have appeared in many journals, including the *Journal of Portfolio Management*, *Journal of Futures Markets*, *European Journal of Operational Research*, *and Annals of Operations Research*, and *Computers and Operations Research*, Professor Gregoriou is coeditor and an editorial board member for the *Journal of Derivatives and Hedge Funds*, as well as an editorial board member for the *Journal of Wealth Management*, the *Journal of Risk Management in Financial Institutions*, and the *Brazilian Business Review*. A native of Montreal, Professor Gregoriou obtained his joint Ph.D. at the University of Quebec at Montreal in finance, which merges the resources of Montreal's major universities (McGill University, Concordia University, and HEC-Montreal). Professor Gregoriou's interests focus on hedge funds, funds of hedge funds, and managed futures. He is also a member of the Curriculum Committee of the Chartered Alternative Investment Analyst Association (CAIA).

Christian Hoppe works as head of portfolio transactions in credit portfolio management in the corporate banking division of Commerzbank AG Frankfurt. His main focus is on structured credit transactions to actively manage the corporate credit portfolio. He is also cofounder and CEO of the Anleihen Finder GmbH in Frankfurt, Germany, an information platform for mezzanine and debt capital. Prior to this he was credit portfolio manager at Dresdner Kleinwort, the investment bank arm of Dresdner Bank AG in Frankfurt. He started his career as a Business and Financial Controller for Dresdner Bank in Frankfurt and was responsible for the corporate client business in Germany. He completed his economics degree at the University of Essen-Duisburg in 2003. While writing his master's thesis, Christian worked in the Institutional Research Department of Benchmark Alternative Strategies GmbH in Frankfurt. Christian is the coauthor of several articles as well as books, author of the German book

Derivate auf Alternative Investments—Konstruktion und Bewertungsmöglichkeiten, published by Gabler, and coeditor of *The Handbook of Credit Portfolio Management*, published by McGraw-Hill.

Carsten S. Wehn is head of market risk control at DekaBank, Frankfurt, where he is responsible for the measurement of market and liquidity risk of the bank and the development of risk methods and models as well as the validation of the adequacy of the respective risk models. Before joining DekaBank, he worked at Deutsche Bundesbank as an auditor and audit supervisor for regulatory examinations of banks' quantitative models for risk measurement and management. He holds a Ph.D. in mathematics and gives lectures at universities. He regularly publishes in well-known industrial magazines as well as in books, mainly about quantitative aspects of risk modelling.

ABOUT THE CONTRIBUTORS

Pauline Barrieu is a reader in the statistics department at the London School of Economics. She holds two Ph.D.s, both with highest honors—in applied mathematics from Paris VI and in finance from the HEC Graduate Business School (France). For the latter, she was awarded the Prix Actuariat 2003 for the best Actuarial Ph.D. in Europe. In June 2006, one of her research papers, "Inf-Convolution of Risk Measures and Optimal Risk Transfer," written with Nicole El Karoui, received the award for the best research paper in quantitative finance by the Europlace Institute of Finance. Her research interests are mainly on the study of problems at the interface between finance and insurance, in particular the design of illiquid products and the securitization of nonfinancial risks. She also works on quantitative methods for risk measurement and robust decision taking, with applications in finance and environmental economics.

Franziska Becker is an academic assistant at the Institute of Finance at the University of Braunschweig.

Emawtee Bissoondoyal-Bheenick is a lecturer in the Department of Accounting and Finance at Monash University in Melbourne, Australia. She has published papers in *Emerging Markets Review*, *Global Finance Journal*, *Quantitative Finance*, and *Applied Financial Economic*s.

Carluccio Bianchi is full professor of economic policy at the Faculty of Economics at the University of Pavia, where he also teaches applied macroeconomics and business cycle analysis. From 1996–2006 he was director of the doctoral program in economics at the University of Pavia. He also lectures about advanced macroeconomics at the ENI Corporate University in Milan and in the Master Program in Cooperation and Development at the University of Pavia. Since November 2006 he is the Dean of the Faculty of Economics at the University of Pavia.

His research interest mainly concerns theoretical and applied macroeconomics and economic policy issues. On these subjects he has published many articles in edited books and in national and international journals.

Laurent Bodson is a Ph.D. candidate in finance and a FNRS research fellow at the HEC Management School of the University of Liège in Belgium.

His areas of expertise include portfolio and risk management, as both a practitioner and a researcher. He also specializes in investment analysis, derivatives, style analysis, stock market price behavior, and integration of higher order moments.

Wolfgang Breuer has been a full professor of finance since March 2000 at RWTH Aachen University, Germany's leading technical university. From October 1995 to February 2000 he was a full professor of finance at the University of Bonn. He earned his Ph.D. in February 1993 and his habilitation degree in July 1995, both at the University of Cologne. After his diploma in 1989 he worked one year in Frankfurt as a consultant at McKinsey & Co., Inc., before he continued his academic career. Professor Breuer has written some dozen books, 60 book chapters, and 40 peer-reviewed journal articles (among others, in the *Journal of Banking and Finance*, the *Journal of Futures Markets*, the *Journal of Institutional and Theoretical Economics*, and the *European Journal of Finance*), comprising a great variety of topics in the field of finance. His current research interests focus on portfolio management, international financial management, and corporate finance.

Robert D. Brooks is a professor in the Department of Econometrics and Business Statistics at Monash University in Melbourne, Australia. He has published a number of papers on empirical finance including articles in the *Journal of Banking and Finance*, the *Journal of International Money Finance*, and *Emerging Markets Review*.

Maria Elena De Giuli is associate professor of mathematical methods for economic and actuarial sciences at the University of Pavia, where she currently teaches mathematical methods for economics and business, quantitative finance, and mathematical finance.

She is a member of the Association for Applied Mathematics in the Social Sciences (AMASES).

Her research topics include financial markets in general, corporate finance, mathematical models and empirical analysis of financial markets, and derivatives pricing. She is currently working on classical and Bayesian models for portfolio analysis and risk management.

She has taught courses on mathematical methods for the Ph.D. in Mathematics for the Analysis of Financial Markets at the University of Milano-Bicocca and for the Master in Methods for Management of Complex Systems at the Istituto Universitario e Studi Superiori in Pavia, Italy.

Bernd Engelmann is a founder and partner of Quanteam AG, a quantitative finance and IT boutique in Frankfurt that focuses on the development of tailor-made front office and risk management solutions for the financial industry. Prior to that, he worked in the research department of Deutsche Bundesbank, where his main research was on the construction and validation of statistical rating models. He has published several articles on asset pricing and risk model validation. He holds a diploma in mathematics and a Ph.D. in finance from the University of Vienna.

Oliver Everling is CEO of Rating Evidence GmbH, a research firm based in Germany. Rating Evidence is involved in the rating and the evaluation of given ratings of enterprises and other businesses. Rating Evidence is an association of the Everling Advisory Services. Since 1998, Everling Advisory Services advises activities, meetings, and publications about rating questions. Among other assignments, Everling is EFFAS Substitute to the CESR Credit Rating Agency Expert Group CWG, Chairman of the International Organization for Standardization (ISO) Working Group "Rating Services," and guest professor at the CUEB Capital University of Economics and Business in Beijing, China.

Cumhur Ekinci is an assistant professor at Istanbul Technical University (ITU). He holds a bachelor's degree in economics from Bogazici University, a master's degree in finance from the University of Paris I Pantheon-Sorbonne, and a Ph.D. in finance from the University of Aix-Marseille III. Professor Ekinci worked in a school trading room at CNAM in Paris and teaches financial markets, investment, corporate finance, and accounting at ITU, CNAM, University of Aix-Marseille II, and ENPC Engineering School. His research topics include market microstructure, high-frequency data, competition among market venues, hedge funds business, and algorithmic trading.

Dean Fantazzini is an associate professor in econometrics and finance at the Moscow School of Economics–Moscow State University. He graduated with honors from the Department of Economics at the University of Bologna (Italy) in 1999. He obtained the Master in Financial and Insurance Investments at the Department of Statistics–University of Bologna (Italy) in 2000 and a Ph.D. in economics in 2006 at the Department of Economics and Quantitative Methods, University of Pavia (Italy). Prior to joining the Moscow School of Economics, he was research fellow at the Chair for Economics and Econometrics, University of Konstanz (Germany) and at the Department of Statistics and Applied Economics, University of Pavia

(Italy). A specialist in time series analysis, financial econometrics, multivariate dependence in finance and economics, Professor Fantazzini has to his credit more than 20 publications, including three monographs.

Franziska Feilke works as an assistant to Marc Gürtler at the Department of Finance at the Technical University of Braunschweig, Germany.

Viviana Fernandez holds bachelor's and master's degrees in economics from Pontificia Universidad Catolica de Chile, and a Ph.D. in economics from the University of California at Berkeley. She is currently an associate professor at the Mining Center of the College of Engineering at Pontificia Universidad Catolica de Chile, and an external research associate of the INFINITI Group, Trinity College Dublin. She has published in *The Review of Economics and Statistics*, *Studies of Nonlinear Dynamics and Econometrics*, *Energy Economics*, the *Journal of Financial Intermediation*, the *Journal of Futures Markets*, *Physica A*, and *Quantitative Finance*, among others. She is currently an associate editor of *The International Review of Financial Analysis* (Elsevier).

Giampaolo Gabbi is a full professor of financial markets and risk management at the University of Siena, Italy, and a senior faculty member of SDA Bocconi School of Management, where he coordinates executive courses on financial forecasting and risk management. At the time of his contribution to this book, he is visiting professor at the City University in London, where he teaches financial regulation. He coordinates the McS program in finance at the University of Siena and also heads the financial areas of the masters in economics program at the same university. Professor Gabbi holds a Ph.D. in banking and finance. He has published many books and articles in refereed journals, including *Managerial Finance*, the *European Journal of Finance*, and the *Journal of Economic Dynamics and Control*.

Hayette Gatfaoui holds a Ph.D. in economics from the University of Paris, Panthéon-Sorbonne. Holding also a MSc in applied mathematics and a MSc in money, banking, and finance, Professor Gatfaoui taught both economics and financial theory during five years at the University of Paris, Panthéon-Sorbonne. She is currently a tenured associate professor at the Rouen School of Management, where she lectures about capital market finance, derivative products, mathematics for finance, and quantitative methods. She is also advisor for firms about risk management and performance measurement topics.

Matthias Gehrke holds a professorship in business administration, especially costing and controlling, at the FOM University of Applied Sciences in

Essen, Germany. In 1989 he received a Ph.D. in chemistry from the Johann-Wolfgang-Goethe-University in Frankfurt, Germany. Shortly after receiving his Ph.D., Professor Gehrke founded a company that develops and distributes medical equipment. He was the managing director of this company until 2004 and grew it into a worldwide organization focusing on certain niche markets in the medical area. After this, he started as a self-employed consultant for reorganization and startup financing. In 2005 he became a lecturer at various universities, which finally lead to his professorship at the FOM.

Werner Gleißner is the CEO of FutureValue Group AG and chief risk researcher of Marsh GmbH. His R&D activities and projects focus on risk management, ratings, and strategy development, as well as the development of methods for aggregating risks and value-based management. His academic research focuses on risk management, valuation, and decision making under uncertainty and imperfect capital markets. He lectures at different universities in Germany in the fields of rating, risk management, value-based management, and entrepreneurship.

Dominique Guégan is a full professor in mathematics at the University of Paris, Panthéon-Sorbonne. She obtained her Ph.D. (mathematics) in 1975 from the University of Grenoble (France). She has published seven books on mathematics, nonlinear time series, and chaos theory, and written 10 book chapters and produced over 55 articles in finance, probability, statistics, nonlinear time series, econometrics, and chaos theory, including *Statistics and Probability Letters*, *Economic Letters*, *Econometrics Review*, the *Journal of Statistical and Planning Inference*, and the *Journal of Applied Probability.*

Marc Gürtler has been since 2002 a full professor of finance at the Technical University of Braunschweig, Germany. Before coming to Braunschweig he was an assistant professor of finance at RWTH Aachen University. He earned his Ph.D. in 1997 at the University of Bonn and his habilitation degree in 2002 at RWTH Aachen University. From 1993 to 1994 he worked as a risk manager in the department of asset management at the AXA Colonia Insurance Company in Cologne, Germany. His research interests include, in particular, portfolio management, credit risk management, and international financial management. He has written several books and peer-reviewed journal articles, and contributed to other books.

Jeffrey Heidemann holds a professorship in business administration and applied mathematics at the FOM University of Applied Sciences in Essen, Germany. He received his Ph.D. in corporate finance from the Technical

University of Ilmenau, Germany. Professor Heidemann also owns a consulting firm for risk modeling. His recent projects include SME loan securitization and risk transfer pricing for banks. Before becoming self-employed in 2001, he worked for credit risk management departments at BHF-Bank and Deutsche Bank.

Florian Heitger works as a financial engineer in the Structured Solutions team in the business unit Capital Markets Structuring & Trading at HSH Nordbank AG. His main objective is to solve quantitative-based problems for the team and to develop models for tailor-made transactions structured by HSH. Particularly, he focuses on credit portfolio modeling for securitization transactions. Prior to joining HSH in May 2005, he worked as a research assistant in the Department of Monetary Economics and International Finance at the University of Kiel, Germany. He was a member of the International Doctoral Program "Quantitative Economics" of the University of Kiel, preparing a doctoral thesis in behavioral finance, which he will submit in early 2009. He holds a degree in physics from the University of Muenster, Germany.

Andre Horovitz has in over 20 years of financial services experience gained insights into industry best practices of some of the leading financial institutions in the world. Mr. Horovitz began his banking career at Lehman Brothers in 1988, and has subsequently held senior executive and top management positions at Oliver, Wyman & Co., Commerzbank, HVB Group (currently part of Unicredit), Erste Bank, Credit Suisse and Nagler & Company. Currently, Mr. Horovitz is president of Financial Risk Fitness, a Munich-based risk-management consultancy. He trains banking executives on financial risk management and financial engineering through case study-based seminars, and is a frequent speaker at various risk-management conferences and a contributor to various industry journals. He holds teaching assignments in financial risk management at the Technical Universities of Munich and Vienna. His areas of expertise cover all classes of financial risk management, including the important link to overall institutional strategy. As a member of the International Finance Risk Institute's executive board, Mr. Horovitz has provided key inputs to the Basel Committee on Banking Supervision in developing the Basel II framework for capital adequacy. Mr. Horovitz holds a diploma in hydraulics engineering from the Technical University of Bucharest and an MBA in finance from New York University's Stern School of Business. He is a licensed engineer in New Jersey and Michigan, a registered securities representative in New York, and a certified Global Association of Risk Professional Financial Risk Manager.

Jason C. Hsu oversees the research and investment management areas at Research Affiliates. He manages $35 billion in the firm's subadvisory and hedge fund businesses as well as direct researches on asset allocation models that drive the firm's global macro and GTAA products and equity strategies that underpin RA's Fundamental Index® concept. Mr. Hsu is an adjunct professor in finance at the UCLA Anderson Business School and served as visiting professor at the UC Irvine Paul Merage School of Management and the School of Commerce at Taiwan National Chengchi University. Mr. Hsu graduated summa cum laude from the California Institute of Technology and earned his Ph.D. in finance from the University of California, Los Angeles.

Georges Hübner holds a Ph.D. from INSEAD (European Institute of Business Administration) and is the Deloitte Professor of Financial Management, and the cochair of the finance department at HEC Management School of the University of Liège. He is an associate professor of finance at Maastricht University, and affiliate professor of finance at EDHEC and Solvay Business School. He is also the founder and CEO of Gambit Financial Solutions, a financial software spin-off company of the University of Liège. Professor Hübner has taught at the executive and postgraduate levels in several countries in Europe, North America, Africa, and Asia. He regularly provides executive training seminars for the preparation of Financial Risk Manager and Chartered Alternative Investment Analyst certifications. His research articles have been published in leading journals including the *Journal of Banking and Finance*, the *Journal of Empirical Finance*, *Review of Finance*, *Financial Management*, and the *Journal of Portfolio Management*. Professor Hübner was the recipient of the prestigious 2002 Iddo Sarnat Award for the best paper published in the *Journal of Banking and Finance* in 2001. He is also the inventor of the generalized Treynor ratio as the multifactor extension of the popular Treynor performance measure.

Vitali Kalesnik, vice president at Research Affiliates, is responsible for quantitative research using advanced econometric tools in the areas of asset pricing and active asset allocation to enhance Research Affiliates' products, including the Fundamental Index® concept and global tactical asset allocation products.

Prior to joining Research Affiliates, he conducted research in economics at the University of California, Los Angeles, where he studied international trade and macroeconomics. He also worked as a researcher at the Ministry of Economics in Belarus as well as Priorbank.

He earned his Ph.D. in economics from the University of California, Los Angeles, was a winner of the UCLA Graduate Division Fellowship from 2001–2005, and speaks fluent English, Russian, and French.

Fiodar Kilin is a Ph.D. student at Frankfurt School of Finance & Management. At the same time, he works as a senior consultant for Quanteam AG, where he has developed exotics pricing models for several trading desks of banks. His research areas are model risk, pricing and hedging of forward-skew-sensitive equity derivatives, and calibration of stochastic volatility models.

Anne Kleppe works as senior consultant for d-fine GmbH. Her major topics are credit portfolio modeling, risk-type aggregation, capital allocation, and further aspects of profitability oriented risk and capital management concepts. Currently she is in charge of d-fine's proprietary developments of d-fine's Credit EC engine and underlying calibration tools. Furthermore, Ms. Kleppe is a specialist in the framework of ICAAP and enterprise-wide risk management for banks, corporate,s and insurance companies. She studied physics and chemistry at the University of Heidelberg, Germany, and earned a Ph.D. in theoretical physics from University of Cambridge, United Kingdom.

Martin Knocinski is employed at Bayerische Hypo-und Vereinsbank AG (HVB) where he provides solutions and performs research in the field of regulatory and accounting issues within the Markets & Investment Banking division of the UniCredit Group. Prior to joining HVB in 2007, he was with KPMG's Regulatory Services Group in Frankfurt, where he worked as a specialist for regulatory questions within the context of different audit and advisory projects for large German banks. He studied at the J. Mack Robinson College of Business at Georgia State University and the University of Duisburg-Essen, where he holds a degree in business administration. He is author of publications on different aspects of accounting and auditing.

Andreas Krause is Lecturer in Finance at the University of Bath, Great Britain. His research focuses on risk management, market microstructure, and agent-based computational finance. He has published a wide range of academic papers in journals such as the *Journal of Fixed Income*, the *Journal of Risk Finance*, the *International Journal of Theoretical and Applied Finance*, *Physica A*, and has contributed chapters to various multiauthor books.

Alexander Kudrov is a Ph.D. student at the Higher School of Economics (Russia). He obtained a master's degree in mathematics at the department of Mechanics and Mathematics–M. V. Lomonosov Moscow State University in 2005. His research interests are multivariate statistical analysis, numerical methods, and extreme value theory and its applications.

Raphael W. Lam obtained his Ph.D. in economics from the University of California, Los Angeles. He is currently an economist at the International Monetary Fund. Previous work experience includes research positions at the Hong Kong Monetary Authority, Institute of Monetary Research, and UCLA. His research interests focus on macroeconomics and finance, in particular, the macro-finance linkage and volatility in the financial markets.

Giacomo de Laurentis is a full professor of banking and finance at Bocconi University, a senior faculty member, and the director of Executive Education Open Programs, Division of SDA Bocconi School of Management. Selected recent publications include "Recovery Rates Determinants. Evidence from the Italian Leasing Market" (with M. Riani), WP SDA Bocconi, n. 82; an editor of *Performance Measurement Frontiers in Banking and Finance*, EGEA (Milano), 2004; an editor of *Strategy and Organization of Corporate Banking*, Springer Verlag, 2005; with M. Riani, "Estimating LGD in the Leasing Industry: Empirical Evidence from a Multivariate Model," in Altman, Resti, Sironi (eds.), *Recovery Risk: The Next Challenge in Credit Risk Management*, RiskBooks, 2005; with G. Gandolfi (eds.), Il gestore imprese. Creare valore per la banca e il cliente con i sistemi informativi di ruolo, Bancaria Editrice, 2008; with J. Mattei, "Lessors' Recovery Risk Management Capability," in *Managerial Finance* (forthcoming).

Mario A. Maggi is assistant professor of mathematical finance at the University of Pavia. He holds an MS in economics from the University of Pavia and a Ph.D. in mathematical finance from the University of Brescia. He held positions at the Universities of Insubria (Varese), Piemonte Orientale (Alessandria), Bologna (Rimini) and Bocconi (Milano). His research interests are mathematical finance, decision theory, and numerical methods.

Marcus R. W. Martin is a professor of financial mathematics and stochastics at Hochschule Darmstadt. Prior to his current academic position, Prof. Martin headed the risk model auditing group of Deutsche Bundesbank, Hauptverwaltung Frankfurt am Main, being responsible for regulatory audits of IRBA, IAA, IMM, and market risk and liquidity risk models.

Scott Mixon is a director at Société Générale Corporate and Investment Banking in New York. He serves as a senior Alternative Investments strategist and as a portfolio manager for hedge funds. Prior to joining SG, he was a director in equity derivatives research at UBS Investment Bank. He began his career as a quantitative asset allocation analyst at SBC Warburg Dillon Read in London. Professor Mixon also worked at Bates White, LLC, an economic consulting firm in Washington, DC. He provided litigation consulting services related to allegations of commodity price manipulation and antitrust risk assessments for private equity acquisitions and a merger of derivatives exchanges. He received a Ph.D. in economics from Duke University. His academic research focuses on market volatility, financial history, and derivatives markets; he has been published in journals such as the *Journal of Financial Economics* and the *Journal of Economic History*.

Christian Oehler works as manager in the department Applied Credit Risk Methods & Strategy within d-fine GmbH. His main responsibilities are strategic developments and projects in the framework of portfolio modeling, risk-type aggregation, capital allocation, and further aspects of profitability oriented risk and capital management concepts. Prior to d-fine, he worked in the Application and Implementation team within Risk & Capital Management department of Deutsche Bank, where he was responsible for groupwide credit portfolio modeling, capital allocation, and IT implementation projects for credit risk management instruments, primarily the proprietary credit default engine. Prior to Deutsche Bank he worked as a senior consultant for Arthur Andersen as a specialist for Basel II–relevant topics with a focus on rating development and validation methods. He earned a Ph.D. in experimental particle physics from the University of Karlsruhe, Germany, and holds a MSc in financial mathematics from the University of Oxford, United Kingdom.

Don Ah Pak is reading for a D.B.A. in corporate strategy. His research interests include strategy, management, and international business with a focus on China. He has worked in South Africa, the United Kingdom, and India. Other achievements include being selected for a graduate development program in 1998–1999; and in 2008, Mr. Pak attended a Faculty Development Consortium sponsored by the London School of Business.

Katja Pluto is head of Risk Methodology at HSBC Holdings plc. She is responsible for all methodological issues arising from and within the economic capital framework of HSBC Holdings. Moreover, her current role includes the responsibility for the groupwide validation of economic capital models at the global, regional, and local levels. Prior to joining HSBC, Ms. Pluto worked for the Banking Supervision Department of Deutsche Bundesbank, where she assessed and approved market risk models and assisted in the Basel II negotiations, with particular focus on the IRBA rules. She began her career with rating system and credit risk model development at Dresdner Bank. She holds a degree in mathematics from Jena University and an M.Phil in finance from Cambridge University.

Marco Rossi received his Ph.D. in quantitative economics jointly from the Catholic University of Louvain and the London School of Economics. He is currently a senior economist at the International Monetary Fund. Prior to joining the IMF, he worked at the Bank of England in its research department. He has published several papers on financial and money markets, international finance, and emerging markets. He is the author of *Payment Systems in the Financial Markets* (Macmillan Press and St. Martin's Press).

Zeynep Iltuzer Samur is a teaching and research assistant at Istanbul Technical University (ITU). She earned a bachelor's degree in mathematical engineering at Y'ld'z Technical University and a master's degree in management engineering at ITU. She is preparing her Ph.D. at the same faculty and works on derivatives and risk management.

Torsten Seil began his career as a financial engineer in the securitization team at HSH Nordbank AG in 2003. During his time in the Global Market Department, he worked on portfolio modeling of various asset classes, including shipping, aircraft, and corporate loans. In 2006 he joined HSH Nordbank's shipping department, where he is responsible for modeling shipping portfolio risk using Monte Carlo techniques and stress testing. He holds a degree in business administration, with a focus on statistics and econometrics.

Shane Shepherd conducts quantitative research used to strengthen and expand the Research Affiliates Fundamental Index® concept, and to support the global tactical asset allocation model. He currently focuses on extending the Fundamental Index strategy into fixed income asset classes and bringing investment products based on this research to the marketplace.

Prior to joining Research Affiliates, Professor Shepherd served as a research assistant at the University of California, Davis Graduate School of Management, where he investigated issues in behavioral finance, and as a legal assistant at Morrison & Foerster, LLP. Professor Shepherd earned his Ph.D. in finance from the University of California, Los Angeles. He holds a bachelor of arts degree in political science and philosophy from Duke University, North Carolina.

John Lockyer Simpson is an associate professor in the School of Economics and Finance at Curtin University in Western Australia. His Ph.D. in economics and finance was earned from the University of Western Australia in the areas of international banking, finance, and country risk. These areas continue to be his focus of research. He has published over 30 refereed papers in well-respected international journals.

Oktay Tas is an associate professor and chair of accounting and finance at Istanbul Technical University. After a bachelor's and master's degrees in accounting and finance at Marmara University, he prepared a Ph.D. at Technical University of Berlin. Professor Tas teaches financial management, portfolio management, and investment. His areas of interest are corporate finance, financial analysis and auditing, and financial derivatives.

Sandrine Tobelem is completing a Ph.D. degree at the London School of Economics funded by CREST, the research department of the French National Institute of Statistics (INSEE) and Credit Suisse. She holds a statistics MSc from LSE and from Paris VI (DEA El Karoui) and graduated from the Statistics School ENSAE, the business school ESSEC, and the political science school IEP in Paris. After working two years at Goldman Sachs for the statistical arbitrage desk from 2002 to 2004, she joined the proprietary desk at Credit Suisse where she is currently working as a quantitative trader. Her research focuses on optimal asset allocation under model risk and more generally robust decision under ambiguity.

Peter Whitehead joined Lehman Brothers in 2001 and was the European and Asian head of Model Validation. The remit of the Model Validation Group at Lehman Brothers was the independent validation of all front office pricing models across all asset classes as well as the independent validation of market and credit risk exposure models. He has a Ph.D. in mathematical finance from Imperial College and is a regular speaker at industry conferences on model risk.

George Woodward is an assistant professor of finance in the College of Business at the University of Colorado at Colorado Springs. He has published a number of papers on applying smooth transition models to nonlinearities in financial markets, including papers in *Quarterly Review of Economics and Finance* and the *International Review of Economics and Finance*.

FOREWORD

The deep financial crisis of 2008 and 2009 has challenged numerous paradigms that were long accepted as standing piers of modern financial risk management. The unprecedented financial losses in the private and institutional sectors around the world and the near collapse of the global banking industry have occurred despite highly praised advances in risk management methodologies, such as regulatory developments meant to safeguard the financial system like the Basel II Accord and the *Markets in Financial Instruments Directive*, and not least, the multibillion dollars for IT installations in large financial institutions geared to limit and control financial risks.

Much of the financial innovation of the last two decades rests on the industry's drive to optimize portfolios of financial assets via diversification strategies along with hedging structures. As such, the emergence of financial derivatives and the wide use of structured financial products incorporating embedded derivative contracts has enabled practitioners to separate components of risk and allowed for warehousing separately such components via modern financial engineering. In consequence, large amounts of economic capital have been freed up in banks around the world. Coupled with the relatively easy access of debt capital throughout the 1990s and the first half of the new decade, banks have engaged in a lending spree that eased funding for numerous projects around the world, thus spearheading an unprecedented global economic growth.

Unfortunately, the originally U.S. originated subprime and "Alt-A" mortgage bubble exploded with an unprecedented vengeance in the face of those very financial institutions that prided themselves with the ability to best model complex products like CDOs and ABCDSs, triggering along the near collapse of the financial system as we know it. The consequences of the aforementioned crisis are still subject to analysis; however, it is certain that a new financial order is to emerge, employing new playing rules among participants: banks, investment funds/vehicles, rating agencies, and regulators.

While the root causes of the global financial crisis are numerous and convoluted, it is certain that the failure of some financial pricing models have catalyzed the spiraling events as they triggered investors' loss of confidence in the very financial models they were hailing for "best in class" and "best practices" only a few years ago.

Perhaps the case study of credit exposure mitigation via CDO securitization is best illustrative of the modeling failures that have been incurred in many banks around the world.

A financial institution typically gathers a large and diversified portfolio of debt receivables (henceforth the "collateralized debt obligation" denomination), places them in a separate financial vehicle (SPV) and issues tranches of securities backed by the pool of assets (in most cases circumventing the complexities of the cash flow transfers via writing credit default swaps on the baskets of the pool assets—in which case the vehicle would be called a "synthetic CDO"). The pecking order of seniority of the issued securities is regulated by the cumulative cascading order of defaults in the pool.

Naturally and intuitively sound—the higher the level of borrower diversification in the pool of assets, the lower the likelihood of simultaneous defaults and the more effective the construct from the SPV owner's point of view.

Financial modelers have been employing methodologies ranging from pair-wise (Pearson) default correlations to copula functions (copulas are functions that link marginal probability distributions with joint distributions and are widely applied in credit risk models) to assess and often stress such structures. Results have been calibrated to exhibited defaults over vast quantities of data and going back decades in a process termed by industry regulators and risk managers as "backtesting."

Such security tranches were calibrated to defaults modeled by past experiences. Correlations and copula functions were calibrated and used to predict future patterns over the lifetime of the securities issued, often in excess of 10 years. Investors and rating agencies alike assigned credit ratings to these CDO securities congruent with the expected loss patterns implied by asset defaults. For example, if the implicit expected loss of a mezzanine tranche would not exceed the equivalent inherent expected loss of a BBB-rated corporate security, it would be rated alike and priced accordingly, tempered by liquidity-driven bid/offer spreads.

By mid–2007 when the first signs of the systemic collapse in U.S. subprime and Alt-A residential mortgages occurred, it became apparent that the historically calibrated correlation patterns were far from indicative of future joint default behaviors, and "tail-end probability" or "conditional/regime-dependent" models were built "in a hurry" in an effort to replace the original pricing models. For many investors this came unfortunately too late.

As prices collapsed (also driven by the new pricing models' results) and investors lost confidence in the pricing models, they scrambled to liquidate as many securities as they could, which triggered an imbalance in supply and demand in secondary markets, which further depressed both the liquidity and the values of the inventories in such tranches. Rating agencies consequently downgraded numerous CDO tranches on the basis that their fair values implied higher default probabilities, which compounded the sell-off in the markets (some investors were prohibited by their investment policies and funds prospectuses to hold securities below a certain agency credit rating). In short, the credit market panic of 2008 was born: lenders of short-term commercial paper to the SPVs failed to renew the credit commitments and the banks were forced either to liquidate the vehicles or consolidate them on their own balance sheets—triggering expensive capital allocations.

From a systemic point of view, a rather dangerous phenomenon has occurred, one linked to the risk mitigation effects of securitization: financial institutions were laying off piles of assets by synthetically buying credit default protection (most common vehicles are credit-linked notes and credit default swaps, but also via spread forward and options contracts), thus trading direct credit exposure for counterparty exposure.

At the margin, a typical transaction would result in a significant risk-weighted asset relief (and consequently regulatory capital congruent with the Basel II stipulations for risk mitigation) against a marginal pickup of counterparty exposure with the protection seller, often a highly rated financial counterparty such as AIG. The financial institution is in no position to know how many such credit default swaps (in terms of size, issuers, and counterparties) the "highly rated protection seller" has on its own books, therefore, in most cases, no reason to fear the simultaneous collapse of the insured asset and the counterparty. Unfortunately, this imbalance and uncertainty proved lethal in the evidence of some CDS players' overexuberance in providing credit protection (for handsome fees), many times insufficiently hedged, which triggered the near collapse of AIG and a few other "systemic relevant" institutions. By 2008, many of these institutions became dependent on "life vests to stay afloat" offered by OECD governments around the world in highly publicized "weekend rescue operations."

As a consequence, it is perhaps more imperative than ever for any participant in modern financial markets to grasp the importance of modeling financial products and assess the strengths but also the weaknesses of the models employed in valuing financial assets. In short, model risk assessment is gaining in importance if not centrality to any financial activity.

There could be no better timing for a comprehensive and deep compendium on assessing model risk in finance. The appearance of the *The Risk Modeling Evaluation Handbook* by Greg N. Gregoriou, Christian Hoppe, and Carsten S. Wehn is a happy and timely event in this direction. The authors have done a superb job in compounding the relevant contributory articles in an all-encompassing effort to gather the most recent industry and academic progress in the field. The book stands out for its ability to match the theoretical thoroughness with the high level of practicality in a "post subprime crisis" economic environment.

Part 1 (Introduction to Model Risk) reviews some significant articles on systemic risk from a central banker's perspective and draws conclusions from previous financial crises. As such, it builds a bridge in time between the financial events that have shaped the way professionals assess model risk in the industry.

Parts 2 and 3 (Model Risk Related to Equity and Fixed Income Investments and Model Risk Related to Credit and Credit Derivatives Investments) assess model risk from an investor and asset manager's viewpoint across all significant risk categories inherent in modern financial products. The aforementioned risks as relating to modeling corporate defaults and more generally migration are being scrutinized and lessons from the ongoing financial crisis are being critically analyzed with a practitioner's implementation oriented eye.

Part 4 (Model Risk Related to Valuation Models) specifically analyzes the risks associated with valuation models and addresses techniques for mitigating wrong valuations of complex financial securities.

Part 5 (Limitations to Measure Risk) primarily addresses risk management professionals as it introduces alternative measures of risk beyond the "value at risk" concept. It addresses the limitation of VaR as a risk measure (specifically identifying the subadditive property of a coherent risk measure as specified by Artzner in a seminal article that appeared in 1999 and proving through practical portfolio composition that VaR can fail this important property when the risk factors are not log-normal distributed). Other modern risk measures such as expected shortfall and copula VaR are analyzed and their suitability is assessed for practical implementation into overall bank-wide risk management architectures.

Part 6 (Modeling Model Risk for Risk Models) specifically addresses the "model risk" within widely used industry risk management models: counterparty risk, credit (portfolio) models, and internal market risk models.

Part 7 (Economic Capital and Asset Allocation) wraps the analysis within the enterprise risk oversight framework as it addresses economic capital models while including the interaction among various risk categories, not least market and funding liquidity risk as well as asset repricing risk (both typically present in asset liability management activities)

The contributors to the compendium are some of the most prominent academics and practitioners in the field of modern risk management and have distinguished themselves over decades with seminal articles that have taken the financial risk management profession into the twenty-first century. Dr. Gregoriou, Mr. Hoppe, and Dr. Wehn have done an outstanding job in carefully selecting the most appropriate articles pertaining to each subject in a flawlessly structured manner.

The *Handbook* will undoubtedly become a key reference book in the shelves of any modern financial professional and will likely contribute to elucidating the still shaded areas of managing model risk, especially in light of the recent happenings in financial markets.

Andre Horovitz
Financial Risk Fitness, GmbH
Munich, Germany

PART I

INTRODUCTION TO MODEL RISK

This page intentionally left blank

CHAPTER 1

MODEL RISK
Lessons from Past Catastrophes

Scott Mixon

ABSTRACT

Financial engineers should study past crises and model breakdowns rather than simply extrapolate from recent successes. The first part of this chapter reviews the 2008 disaster in convertible bonds and convertible arbitrage. Next, activity in nineteenth-century option markets is examined to explore the importance of modern theory in pricing derivatives. The final section reviews the linkage between theory and practice in bridge building. The particularly interesting analogy is the constructive direction taken by engineers after the highly visible, catastrophic failure of the Tacoma Narrows Bridge in Washington in 1940. Engineers were reminded that highly realistic models may increase the likelihood of failure if they reduce the buffer against factors ignored by the model. Afterward, they focused efforts on making bridges robust enough to withstand eventualities they did not fully understand and could not forecast with accuracy.

MODEL RISK: LESSONS FROM PAST CATASTROPHES

The holddown cables stabilizing the bridge began to vibrate in the wind around 3:30 in the morning, as the storm increased. The bridge was oscillating by 8 a.m., allowing a driver to watch the car up ahead disappear into a trough. Yet this was nothing new for the bridge nicknamed “Galloping Gertie.” An hour later, the tiedown cables cracked like whips as they alternately tightened and slackened.

The twisting began around 10 a.m. Winds were only 45 miles per hour, but the bridge was pivoting around the center line of the two-lane roadway.

One lane would lift up 45 degrees, and the other would twist down by the same amount. The bridge was twisting itself to pieces. A major section of roadway, hundreds of feet long, fell into Gig Harbor at 11 a.m. Ten more minutes and the remainder of the bridge (plus two abandoned cars and a dog) was gone. Leon Moisseiff, prominent engineering theorist and designer of the bridge, was completely at a loss to explain the disaster.

The collapse of the Tacoma Narrows Bridge took with it the entire trajectory and hubris of the suspension bridge building industry. The bridge lasted just four months, and it failed in winds less than half the 100 miles per hour for which it was designed to withstand. The dramatic collapse was captured on film, so the world saw the images repeated over and over in newsreels.

Suspension bridges had become longer and thinner in the decades before the Tacoma Narrows collapse in 1940. The George Washington Bridge, spanning the Hudson River into New York City, epitomized this trend. Built according to the most advanced theories, it had doubled the length of the longest suspension span extant before its 1931 completion. Engineers were under pressure from the economic realities of the Great Depression to reduce costs, while theoretical advances showed that conventional bridge designs were considerably overengineered. Suspension bridges had moved from designs in which the suspending cables were superfluous to ones in which cables were the dominant element.

Moisseiff and Lienhard's (1933) extension of existing theory meant that wind loads carried by the suspended structure of the George Washington Bridge, for example, were 80 percent lower than previously thought.[1] The cables carried the weight. It was only natural that Moisseiff's Tacoma Narrows design would incorporate this economically appealing feature. The bridge had a width to span ratio of 1:72, even though a more traditional ratio was 1:30. It was a narrow, two-lane highway suspended in the sky, and it was theoretically justified to carry its loads adequately. For four months, it did.

Engineers had built little margin of safety into the Tacoma Narrows Bridge. It was elegant, economical, and theoretically justified. Moving to more realistic theories of load bearing ironically led to catastrophic failure, since the theories provided little buffer against risk factors not considered in the models. Henry Petroski (1994) has noted that the reliance on theory, with little first-hand experience of the failures from decades before, led to a design climate in which elements of recent successes were extended beyond their limits. Petroski therefore recommended that engineers carefully study past design failures in order to improve their current design process. This chapter adopts Petroski's logic to model risk evaluation for financial engineers.

INTRODUCTION

First, we discuss the disaster in the convertible bond market during 2008 and pay particular attention to the difficulties faced by the model-intensive convertible arbitrage strategy. Next, we review the activity in option markets in the nineteenth century to explore the importance of modern theory in pricing derivatives. This is followed by a section reviewing the linkage between theory and practice in bridge building. It highlights the direction taken by the engineering profession after the highly visible, catastrophic failure of the Tacoma Narrows Bridge in 1940. The final section provides concluding thoughts.

The goal of bringing together such disparate topics is to provide a broad perspective on model risk. Some of the key issues addressed are: (1) How can we best frame themes regarding model breakdowns during market crises? (2) What can the careful study of the evolution of derivatives markets and no-arbitrage pricing models suggest for robustifying market practice against traumatic shocks? (3) What can be learned from the experiences of other disciplines that have suffered catastrophic failures when moving from theory to reality?

There are some straightforward conclusions for this chapter. First, financial engineers should study past crises and model breakdowns rather than simply extrapolate from recent successes. Second, theoretical advances have had a profound impact on the pricing of derivative securities and the mindset for financial engineering, but the real world is tricky. Highly realistic models may increase the likelihood of failure if they reduce the buffer against factors glossed over by the model. For example, highly complex no-arbitrage models may fail miserably when trading is not continuous and arbitrage opportunities exist for a period of time (i.e., when liquidity disappears).

Financial engineers can look to the experience of the civil engineering profession after the traumatic failure of the Tacoma Narrows suspension bridge in 1940. Bridge builders could have concluded that their theories were too naïve to build economical bridges as long as the ones proposed. They did not throw out the models or abandon the practice of bridge construction. Rather, they focused efforts at making bridges robust enough to withstand eventualities they did not fully understand and could not forecast with accuracy.

Faced with calamities such as the recent one in the financial markets, one could take the position that highly quantitative models are dangerous and should be thrown out. This idea is completely unjustified. A key idea running through this chapter is that some of the assumptions behind derivatives modeling, such as continuous, frictionless trading at a single market price (i.e., the absence of arbitrage) can fail at critical moments. Engineers (and drafters of legal documentation) should seek ways to robustify the structures so that these disruptions do not lead to catastrophic failures.

For example, some of the hedge fund industry might move to longer lock-ups and increased acceptance of a secondary market in ownership interests (a cross between ETFs and closed-end funds). During extreme liquidity crunches, as in the fall of 2008, investors requiring liquidity could have raised cash by selling their hedge fund interests, yet the underlying funds would not have been forced to sell assets to fulfill redemption requests. Is this not, in fact, a substantial part of the logic behind having a liquid secondary market for corporate ownership of companies?

CONVERTIBLE CATASTROPHE, 2008

The bottom line is that everything went wrong for convertibles in the last part of 2008, and the convertible arbitrage strategy gave back about six years of gains in a few months. There was tremendous selling pressure on convertibles, about 70 percent of which were held by hedge funds.[2] Convertible arbitrageurs typically were long convertibles and hedged various risks to profit from perceived mispricing of the convertible. In extremely volatile and thin markets, the ability to hedge was disrupted: the ability to short many stocks was eliminated (partially by government fiat), synthetic credit hedges were not following the same trajectory as cash bonds, and the implicit call options in convertible bonds were priced completely out of sync with listed options. Margin requirements increased massively. Redemptions loomed for hedge funds of all stripes, and raising liquidity became the overriding goal. Convertible arbitrage was hit hardest among the hedge fund strategies during 2008, with some indexes showing the strategy down more than 50 percent.

This section puts a finer point on these issues. The conceptual framework of the model risk in mind for a crisis of this type is $p_{it} = E^m\ [p_{it}] + e_{it}$, $e_{it} = c_t + d_{it}$, where p_{it} is the market price of asset i at time t, $E^m\ [p_{it}]$ is the expected value of the asset at time t under the model m, and there is an error term e_{it} that causes the observed price to deviate from the model price ("fair value"). We can divide the error term into two components: the idiosyncratic disturbance d_{it} and the market-wide term c_t. Roughly speaking, a crisis is when c_t, on rare occasions, dominates price movements. Evaluating model risk and figuring out ways to work around model failure during a crisis means understanding when c_t is the driving force behind market prices.

What Happened?

Table 1.1 displays convertible arbitrage index performance in 2008, along with performance for the S&P 500 and Vanguard's Intermediate Corporate Bond Fund. Returns are standardized by subtracting the average and dividing by the unconditional standard deviation. Averages and standard deviations are computed using data through December 2007. The Center for International Securities and Derivatives Markets (CISDM) index data begin in January 1992, the Credit Suisse/Tremont data begin in January 1994, and the Hedge Fund Research Performance Index (HFRI) data begin in January 1990. The S&P 500 and Corporate Bond Fund data begin in January 1990. Table values in bold are t-statistics of at least 2 in absolute value.

March 2008 seemed to be a particularly unlucky month for convertible arbitrage funds, according to the table, as returns were extremely unlikely. Yet things worsened: September and October were greater than 10 standard deviations to the downside. The corporate bond fund and the S&P 500 returns were also much larger in absolute magnitude than history would suggest, so perhaps some of these shocking moves were simply volatility increasing above the unconditional levels. Still, convertible arbitrage appears to have experienced some very rare events.

Table 1.1 Standardized Monthly Convertible Arbitrage Returns, 2008 (Standard Deviations)

	CISDM	Credit Suisse/ Tremont	HFRI	Average CB Arbitrage Index	Vanguard Corporate Bond Fund	S&P 500
Jan	−0.4	−0.9	**−2.2**	−1.2	0.0	−1.8
Feb	−0.9	−1.5	−1.6	−1.3	−0.3	−1.1
Mar	**−4.1**	**−5.1**	**−5.0**	**−4.7**	0.0	−0.3
Apr	−0.5	0.3	0.5	0.1	0.1	1.0
May	−0.1	0.6	0.0	0.2	−1.0	0.1
Jun	−1.6	−0.8	**−3.1**	−1.8	0.3	**−2.4**
July	**−3.0**	**−2.2**	**−2.4**	**−2.5**	−0.8	−0.4
Aug	−1.9	−1.0	−1.9	−1.6	0.6	0.1
Sep	**−12.0**	**−10.4**	**−13.3**	**−11.9**	**−2.8**	**−2.6**
Oct	**−16.4**	**−10.7**	**−18.2**	**−15.1**	**−4.5**	**−4.8**
Nov	−1.2	**−2.0**	**−3.3**	**−2.2**	**4.0**	**−2.1**
Dec	3.0	−1.3	0.2	0.6	**6.3**	0.0

CISDM, Center for International Securities and Derivatives Markets; HFRI, Hedge Fund Research Inc.; CB, Convertible Bond.

Figure 1.1 illustrates a measure of convertibles mispricing from January 2005 to April 2009. The chart shows the ratio of the median price (average of bid and ask) to the median theoretical price for the entire U.S. convertible universe tracked by Deutsche Bank. Both the bid/ask midpoints and the theoretical prices are Deutsche Bank calculations. From 2005 until the end of 2007, this mispricing hovered around zero; one can see occasional divergences, such as the spring 2005 cheapening documented by Mitchell and colleagues (2007). Convertibles then cheapened dramatically in spring 2008 but bounced back toward fair value by July. Then the bottom fell out, with market prices collapsing to more than a 15 percent discount by November. The mispricing steadily diminished after that and fell to less than half by April 2009.

"Fair value" models of convertibles suggested a much higher value than observed market quotes of late 2008 would suggest. Outright buyers of convertibles should have moved in, according to the textbook no-arbitrage condition, and bought up the cheap instruments. Anecdotes suggest this is what happened eventually, but it was not instantaneous. September through November felt like really long months to market participants staring at the wrong side of arbitrage opportunities.

Figure 1.2 displays another way of exploring convertible mispricing in 2008, focusing on the option component of the bonds. The chart displays a market-capitalization weighted index of one-year, at-the-money implied volatilities for S&P 100 stocks (constructed from Bloomberg data) and an issue-weighted index of convertible bond implied volatilities. The convertible index is from Deutsche Bank; it is computed from near-the-money

Figure 1.1 Convertible Bond Prices Versus Theoretical Values

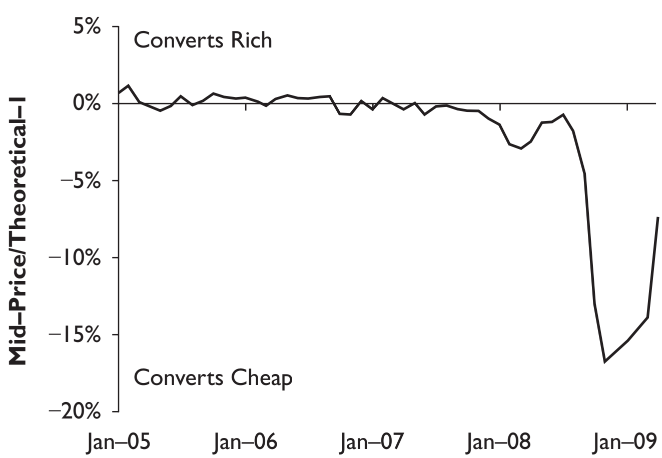
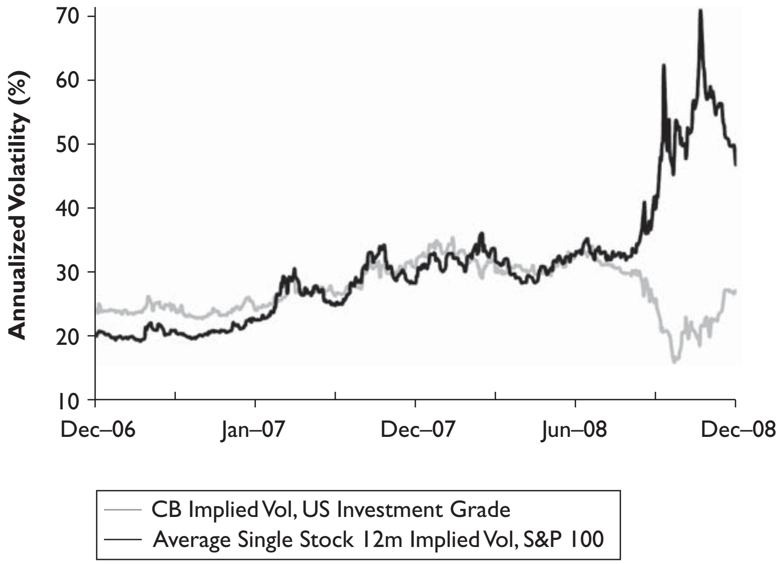


Figure 1.2 Convertible and Option Implied Volatility



investment grade convertibles and is reweighted quarterly. Throughout 2007 and the first half of 2008, the two series track each other quite closely. The exception during this period is for March 2008, when listed option volatilities spiked even as convertible volatilities moved in the opposite direction. (This spike is distinctly noticeable in the convertible arbitrage standardized returns for March.)

The dramatic divergence in the two markets is in the second half of 2008, when listed volatilities doubled to more than 70 percent and convertible implieds sank to less than 20 percent. Over the course of autumn, listed volatilities began to decline as markets began to stabilize. By the end of the year, the listed option volatility index was hovering around 50 percent, yet convertible implieds were still less than 30 percent.

Convertibles apparently had a negative vega, that is, buying a convertible meant going short volatility. This is clearly counterintuitive and stands in contrast to the expected exposure. Arbitrage managers had hedged themselves using models that suggested they were close to delta neutral, but the models would never have suggested such perverse moves in pricing as had occurred, and the theoretical option models would surely have suggested a positive vega for convertibles.

Industry analysts suggested that by mid-November, one-third of convertible arbitrage funds had gated redemptions, suspended net assest value calculations, or were in some other way impaired.[3] Some of the stabilization in pricing occurred as selling pressure abated due to reduced redemption pressure. The redemption pressure was artificially capped as some managers gated and simply refused to return investments to investors. The investor response was not a happy one. Anecdotally, some investors declared funds that gated as "uninvestible" going forward.

This discussion of convertible mispricing can be made less abstract, particularly with respect to hedging instruments. Figure 1.3 illustrates a particularly clear example using traded prices for two bonds issued by Smithfield Foods. The gray bars represent the straight bond due May 15, 2013, and the squares represent the convertible bond due June 30, 2013. The chart begins with the June 2008 issuance of the convertible and ends in late May 2009. Some interesting observations can be made. First, there are many gaps in the trade history of the bonds, particularly in November 2008. Only 12 of the 18 trading days in that month saw any action, and the median daily volume over the month was 60 bonds. The convertibles traded only six days during the month. Second, the price of the convertible is often below that of the straight bond, suggesting a negative value for the option to convert the bond into stock (these observations are the white squares).

Figure 1.4 focuses on the valuation of the option implicit in the convertible. The boxes represent the implied volatility (provided by Deutsche

Figure 1.3 Convertible and Straight Bond Prices, Smithfield Foods

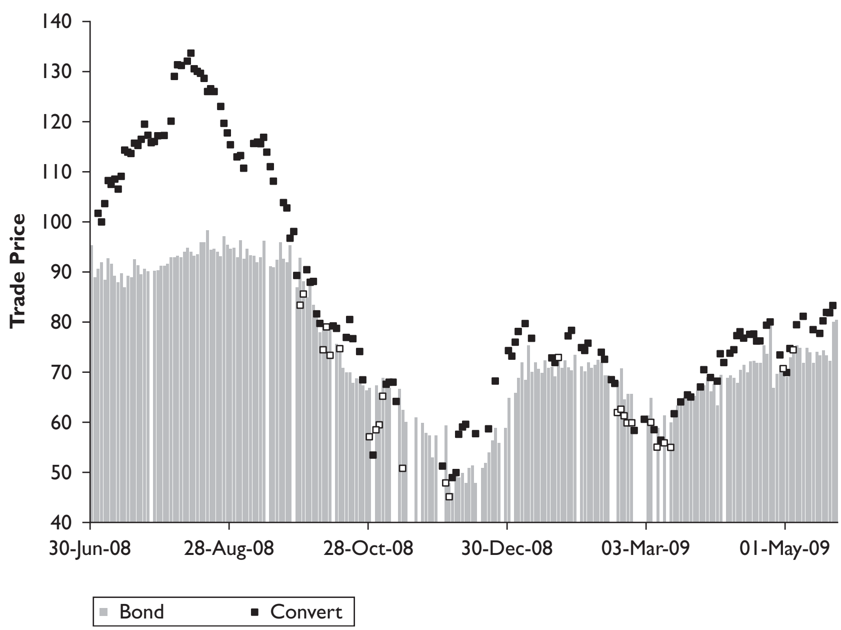
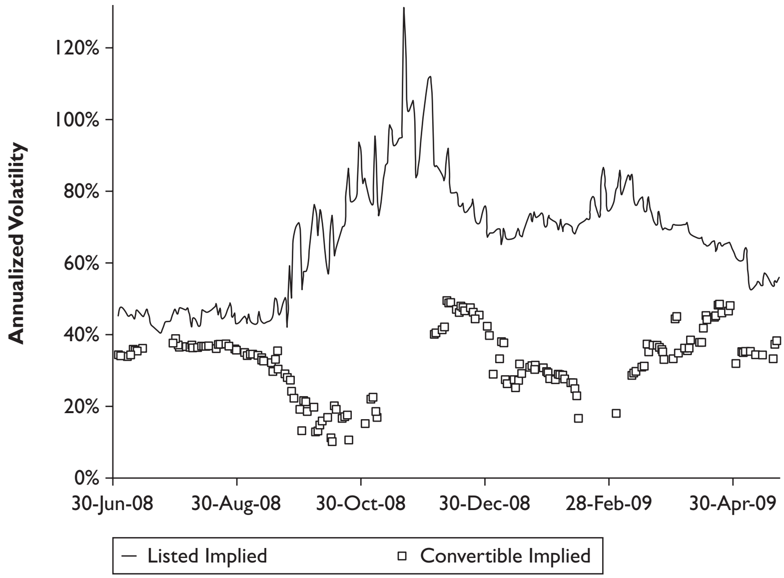


Figure 1.4 Convertible and Option Implied Volatility, Smithfield Foods



Bank) for the convertible, and the solid black line represents the implied volatility from the closest comparable listed option (provided by iVolatility.com). The convertible bond provides for a conversion price (i.e., the strike price of the option) of $22.685 per share. The listed implied volatility is for the January 2010 25-strike call option until December 12, 2008, and for the January 2011 22.5-strike call thereafter.

Corresponding to the many days shown in Figure 1.3 with a convertible price below the straight bond price, there are many dates for which no meaningful convertible implied volatility can be calculated. Even on the days when it can be calculated, it is far below the listed volatility and often moves in precisely the opposite direction as the listed volatility. The implicit option in the convertibles was not a good substitute for the comparable listed option during this period. A model that might have worked for listed options would have failed miserably for convertibles.

Why Did It Happen?

Market participants were convinced about the causes of the shock to convertibles. "It is the cost and availability of leverage that is driving the market," said one portfolio manager.[4] Because prime brokerage desks cut back on financing activity in the midst of increased perceived risk and bank-wide deleveraging, convertible margins increased. Margin for convertibles was typically around 15 percent before September 2008 but had risen to 30 percent by mid-November.[5] "At the same time, some funds focused on convertibles borrowed as much as $5 for each $1 of equity in recent years, leverage that has turned small losses into huge ones. . . . Now prime brokers are either cutting off hedge fund clients, raising borrowing rates or forcing them to produce more collateral to back the borrowed money."[6] Credit lines had been reduced, and the market for more speculative grade convertibles had vanished, leaving cash-strapped managers to sell liquid assets wherever they could be found.

Forced deleveraging by convertible managers was not the only issue related to increased haircuts: multistrategy hedge funds were said to be deleveraging to meet margin calls in a variety of areas after losses and margin increases. Table 1.2 shows estimated changes in margin for a variety of instruments. Any liquid securities were fair game for raising cash. The remaining portfolios after the liquidations were highly illiquid. Many funds imposed gates or put illiquid assets into sidepockets, increasing the gridlock for investors in need of liquidity.

Another major roadblock that no model expected was the imposition of arbitrary short-sale constraints by regulators. One regulation by the U.S. Securities and Exchange Commission (SEC) eliminated the exemption for options market makers to deliver shares of companies placed on threshold lists (i.e., stocks for which short sales had failed after the shares were not delivered). Next, the SEC ordered that market participants could not initiate new short positions against nearly 800 financial stocks, as of September 19. "Not only does the SEC rule hamper hedge funds," concluded the financial press, "Wall Street trading desks also have been handcuffed. . . . While an exemption provided by the SEC allows dealers to sell stocks short as part of

Table 1.2 Typical Haircuts or Initial Margins

	April 2007 (%)	August 2008 (%)
U.S. Treasuries	0.25	3
Investment-grade bonds	0–3	8–12
High-yield bonds	10–15	25–40
Equities	15	20
Investment grade corporate credit default swap	1	5
Senior leveraged loans	10–12	15–20
Prime mortgage–backed securities	2–4	10–20
Asset-backed securities	3–5	50–60

Source: International Monetary Fund (2008), p. 42.

their function as market makers, that leeway doesn't apply to market-making in convertibles."[7] Nor did it initially apply to option market makers. Once again, the opinion of market participants was unequivocal: "[O]ptions market makers were refusing to quote without the ability to hedge. If they don't fix it, there just won't be an options market on Monday."[8]

The week before the elimination of the exemption was overturned, option trading desks faced difficulties facilitating customer orders to hedge exposure in financials. As an indication of the dislocation in the option market due to the SEC short-selling regulations, consider the percentage bid/ask spread of XLF options. The XLF is a market capitalization weighted basket of financial stocks, and options on it should reflect the difficulties of shorting these stocks. The following numbers reflect a weighted average closing bid/ask spread using the put option slightly above and below the closing spot price for the XLF (using Bloomberg data). For the two weeks prior to the SEC rulings (September 2–12), the spread averaged 3.7 percent. During the week of the short sale rulings, the spread increased to an average of 15.3 percent. The following week, market maker exemptions were re-established and spreads declined to 6.2 percent. Based on the sharp change in spreads over these event dates, I infer that option market liquidity was severely diminished by the rulings.

Frictions remained after the exemption was reinstated. Market makers could not knowingly effect a short sale for a customer who was establishing or increasing a short position in restricted stocks. Desks required a memo from customers stating that the purchase of put options, for example, would not increase the customer's economic short position in that name (lying on the memo was fraud, said the SEC). Selling calls on a financial stock was similarly difficult.

Post Mortem

Convertible bond managers saw their assets sold off due to exogenous shocks, and they faced thin markets if they tried to sell. The buyers' strike meant that price discovery was difficult, if not impossible, for convertibles. Continuing supply shocks (prop desks or hedge funds unwinding entire books by sending out bid lists of portfolios) and tighter budget constraints for remaining funds meant that "good deals" went ignored. Gates, side-pockets, or suspensions were likely to be imposed on some funds. End-investors had little visibility into the portfolios and redeemed more than they really wanted, just to get the redemptions they needed. The cycle was vicious. Textbook arbitrageurs with unlimited capital and the ability to swoop in and buy underpriced assets were a myth.

What models would have coped with this situation better and managed losses better? Hedge funds needed to raise cash or hedge their convertible positions, but convertible markets were minimally open and hedging instruments were unavailable. Even if they had previously traded a hedge, such as selling a call option or shorting a similar bond, the hedge moved the wrong way and they lost money. What are the lessons?

I suggest that tweaking valuation and hedging models that work 99 percent of the time, by adding transactions costs or occasional massive jumps, may not be the most useful way to protect against the hundred year floods that happen every few years in financial markets. Green and Figlewski (1999) found that writing options with an implied volatility slightly higher than "expected"—a dealer mark-up, if you will—protects against many small losses but does not protect much against the most violent shocks. Going back to the conceptual model $p_{it} = E^m [p_{it}] + e_{it}$, $e_{it} = c_t + d_{it}$ described earlier, the "fair value" of the asset may well be a valid concept for the fund manager, even if (especially if?) it occasionally diverges dramatically from the market price. It may be also be perfectly adequate for hedging local risks. Tweaking the model may produce slightly different values or hedge ratios, but any variation is likely to be swamped by a huge value for c_t during crisis periods. Factoring in huge risk premiums to account for every eventuality, no matter how remote, effectively means walking away from the strategy.

The most important model to fix, in this instance, might be the investment and distribution model for hedge funds rather than the pricing model. Ensuring effective communication to investors, along with appropriate transparency, should help diminish noise trading by investors. Given the reluctance of hedge funds to "talk their books," perhaps this is best accomplished by passing the information through an intermediary (e.g., a risk aggregator service, or a managed account platform with established risk controls) who can vouch for the fund's prudence without revealing sensitive portfolio information. Second, the gridlock and uncertainty faced by investors at a market-wide level might be minimized to the extent possible by allowing better access to secondary markets in claims on hedge fund interests. This might facilitate better price discovery that is not buffeted by liquidity shocks that could amplify fundamental market disturbances. An open question would be how to minimize the principal-agent problems with a setup similar to this one.

BEFORE THERE WERE OPTION PRICING MODELS...

This section provides a brief review of how option pricing actually worked before modern no-arbitrage models dominated thinking. It turns out that option pricing in the nineteenth century, for example, did not work so differently than it does today. After the widespread adoption of no-arbitrage models (based on continuous trading and perfectly liquid markets), the major change was the sharp decline in the gap between implied and realized volatility. Option prices using the old rules of thumb were considerably overengineered, at least according to the textbooks.

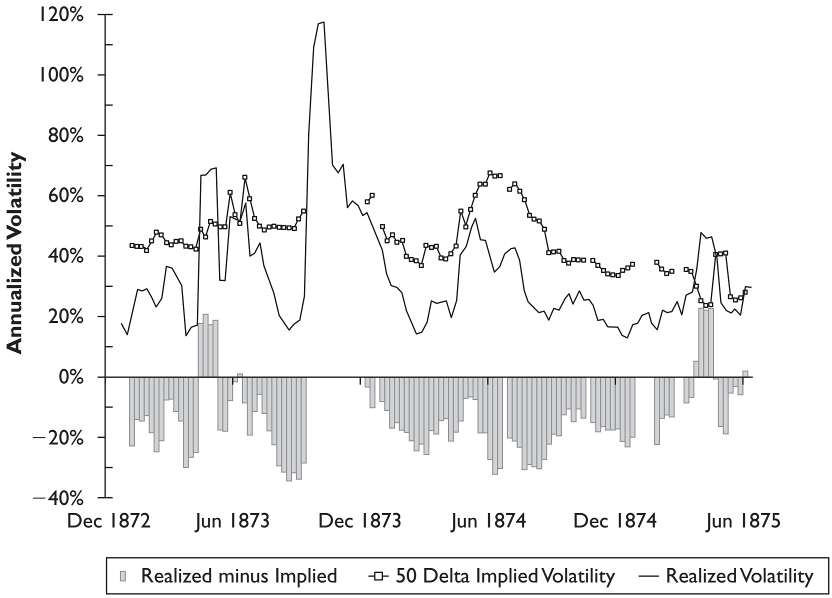
We start by observing that option markets existed and worked reasonably well in the centuries before no-arbitrage models came to the fore. This should not be such a surprise, as markets for other assets or commodities still function well without no-arbitrage models. Option prices conformed to a number of empirical regularities seen in today's markets: implied volatility typically exceeded realized volatility and followed its general movements, the volatility skew existed and moved similarly across stocks, and implied volatility matched realized volatility in the cross section of stocks (i.e., high volatility stocks had higher implied volatility) (Mixon, 2009). Supply and demand conditions plus trader intuition was enough to generate the same empirical regularities as modern markets. The striking difference is that implied volatility was vastly higher than realized volatility during the nineteenth century, and the gap between the two fell sharply as soon as the Chicago Board Options Exchange opened in the early 1970s. Theory did not create option markets, but it sharply impacted them along that one dimension.

Traders used rules of thumb that approximated the precise values prescribed by modern theories. For example, options in London featured a strike price generally set at the market price of the stock at initiation (plus interest until expiration, hence the options were effectively struck at the forward). If the stock was at $75 and the option was quoted at 1.25 percent, for example, the price of a 75-strike call was $125 per hundred shares. If a buyer wanted to cheapen the cost of the option, the rule of thumb was to adjust the option price by half the adjustment in the strike price. Modifying the example above, a call strike at $76 would cost 0.75 percent or $75.[9]

Figure 1.5 readily demonstrates the sophistication of option markets in the nineteenth century. The heavy black line is trailing one-month realized volatility for Union Pacific, an active railroad stock. The measure is the annualized Parkinson volatility computed using high and low daily price data. The black squares represent a 50 delta implied volatility for one-month Union Pacific options, computed from indicative quotes in the *Commercial and Financial Chronicle*. Mixon (2008) provides a full discussion of the market and the construction of the data used here.

Visual inspection shows that volatility moved around quite a bit during this period as Union Pacific's and the wider market's prospects varied. Baseline volatility appears to be around 20 percent, with spikes up to the 40 to 50 percent range. Volatility spiked up over 100 percent during the Crisis of

Figure 1.5 Union Pacific Implied and Historical Implied Volatility, 1873–1875

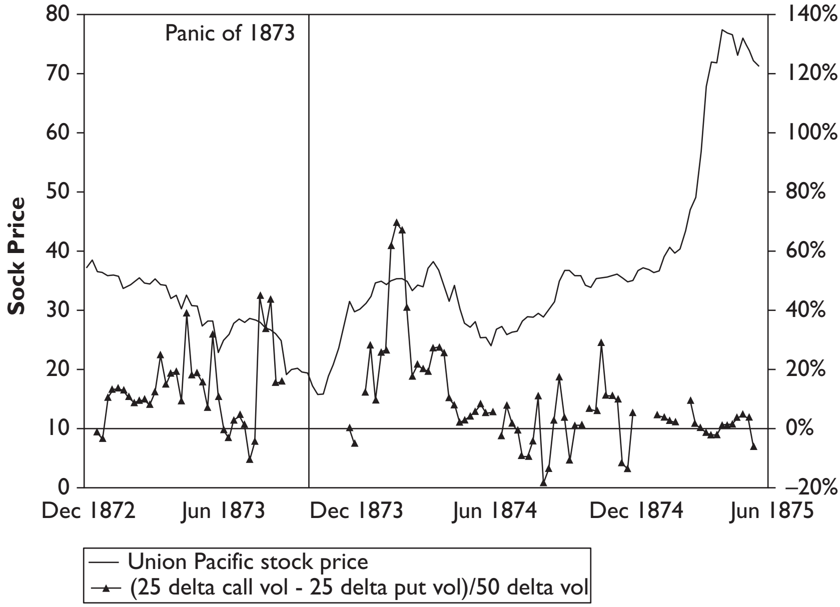


September 1873. Some observations are apparent. First, implied volatility generally tracks the movements in realized volatility, gradually declining as markets quieted down after the crash in September 1873, rising during the April 1874 market decline, and so forth. Second, the gap between implied and realized volatility averaged around 14 volatility points for dates when both values are available. This contrasts with the decidedly smaller values observed in modern-day markets.

Figure 1.6 illustrates the skew for equity options during the ninteenth century. When Union Pacific sold off during the Crédit Mobilier of America scandal of 1873, 25 delta call volatilities were bid up relative to 25 delta puts. This pressure also applied during the rally in early 1874. As the market traded in a range and then rallied sharply into 1875, the skew tended to fluctuate around zero.

These illustrations suggest that option markets were, and remain, highly influenced by supply and demand factors. Generally speaking, modern markets generate option prices the same way they did long before sophisticated mathematical models became available. After the introduction of "fair value" prices based on no-arbitrage models and perfect markets, the major change was the sharp decline in the gap between implied and realized volatility. Option prices using the old rules of thumb were considerably overengineered. Whereas the textbook no-arbitrage Black-Scholes solution

Figure 1.6 Union Pacific Stock Price and Volatility Skew, 1873–1875



predicts that option supply (long or short) is unlimited at the no-arbitrage price, this appears not to hold exactly in the real world. Careful research on modern option markets (e.g., Bollen and Whaley [2004]) underscores this proposition. Surely the impact is magnified during crisis periods, as the section above documents for convertibles during 2008.

LESSONS FROM BRIDGE BUILDING

Like option traders in the nineteenth century, bridge designers in that era had to rely on intuition and experience. They watched light suspension bridges blown away in storms, and a primary focus in bridge building was building a bridge that would not fail. Suspension bridge builders made their designs heavy and stiff and virtually ignored the presence of the cables holding up the roadway. They added stays and trusses, and they added weight so that the bridge would not swing like a pendulum in the wind. Yet, as the field progressed, the theoretical basis for their designs became far more realistic regarding how loads were distributed among the various components of the bridge. All that stiffness became expensive, unnecessary overkill.

Suspension bridges so long as to be virtually unbuildable using mid-nineteenth-century theories became feasible and turned into reality. The George Washington Bridge in New York spans 3,500 feet and was completed in 1931; it provides a useful example. Buonopane and Billington (1993) computed the deflection at the midpoint of the span, for a particular load, under several theories. The deflection theory, which we can treat as approximately the "true" value, indicates that the bridge would flex by 34 inches under the load. The less sophisticated elastic theory, dating from the 1880s, predicts a deflection over 10 times that amount at 363 inches. The Rankine theory from the 1850s would predict a deflection of 5,800 inches. The simple unsupported beam theory, which does not account for the cables at all, predicted a bending of nearly 2.4 million inches (nearly 200,000 feet). The old theories were far too conservative.

The upshot was that engineers in the 1930s could calculate quite precisely how bridges would perform under various loads. They economized on steel and cable, as the theories prescribed exactly how much they could safely carry. The newer, more realistic theories allowed designers to build longer bridges, build them more cheaply, and to make them quite elegant and graceful, instead of the expensive, bulky bridges from the past. For example, it has been estimated that building the George Washington Bridge with the deflection theory saved US$10 million (in 2008 dollars, that's more than US$100 million) due to reduced requirements for steel, cable, and anchorage costs.[10] In general, the application of the improved theory for the design of suspension bridges led to savings in truss steel estimated from 20 to 65 percent.[11]

And yet, the precise theories that allowed the designers to eliminate the expensive deck stiffening reintroduced wind-induced motion as a major design issue into bridge construction. The theories were fine as far as they went, but they were not sophisticated enough to deal with the aerodynamic issues emerging even at a relatively low velocity wind. The designers had become so sophisticated they forgot the lessons their predecessors had learned the hard way about making a bridge stable in the wind.

After the humbling experience of the Tacoma Narrows Bridge failure, bridge designers began carefully studying bridge aerodynamics. They stiffened the bridges. They utilized numerical methods to analyze the problems and built scale models. They still debate the exact mechanism by which the Tacoma Narrows twisted itself apart.[12] Perhaps the most important lesson for financial engineers is that they did not abandon sophisticated models, nor did they abandon building ever-more sophisticated and longer bridges.

Shoehorning economics questions into a physics or engineering framework may seem misplaced, but perhaps creative thinking can be stimulated by analogy, that is, by examining how those disciplines have responded to practical problems. For example, Strogatz and colleagues (2005) concluded that pedestrians on a swaying footbridge find that falling into step with the bridge's swaying is more comfortable than fighting it. Yet this individually rational "phase locking" behavior amplifies the swaying and could lead to disaster. The solution is to modify the bridge to dampen the feedback. Little imagination is needed to find the analogy of hedge funds or other institutions selling assets to reduce leverage in a crisis, amplifying the effects. What kind of dampeners could reduce such feedback during a marketwide financial deleveraging?

CONCLUSION

Some straightforward conclusions on the art of financial model building emerge from this review. First, financial engineers should carefully study past failures. Extrapolating from seemingly successful structures without understanding potential causes of failure can lead to catastrophes. Ironically, reliance on more precise theories can lead to disaster, as the structure can become more exposed to forces not incorporated into the more sophisticated theories. Other professionals, such as MBAs, analyze case studies from the real world in order to stimulate critical thinking; we should expect no less from financial engineers. The book by Bruner (2005) is an example of such work in the corporate finance area; it draws lessons from corporate mergers that, in retrospect, were disasters.

Second, continuous trading and the absence of arbitrage are crucial to most modern financial models. However, the real world contains frictions that make this broad assumption too strong to generate realistic predictions at all times. Financial engineers should be mindful of these extreme cases and design structures that can withstand such bursts of chaos.

Finally, despite the disasters and public disrepute associated with financial engineering, the last thing we should do is give up on building sophisticated models to understand how the world works. We have to learn how to work around our limited knowledge and make incremental progress.

REFERENCES

Bollen, N.P. and Whaley, R.E. (2004) Does Net Buying Pressure Affect the Shape of Implied Volatility Functions? *Journal of Finance*, 59: 711–753.

Bruner, R.F. (2005) *Deals from Hell: M&A Lessons That Rise Above the Ashes*. Hoboken, NJ: John Wiley & Sons.

Buonopane, S.G. and Billington, D.P. (1993) Theory and History of Suspension Bridge Design from 1823 to 1940. *Journal of Structural Engineering*, 119: 954–977.

Green, T.C. and Figlewski, S. (1999) Market Risk and Model Risk for a Financial Institution Writing Options. *Journal of Finance*, 54: 1465–1499.

International Monetary Fund. (October 2008) Global Financial Stability Report, Washington, DC.

Mitchell, M., Pedersen, L.H., and Pulvino, T. (2007) Slow Moving Capital. *American Economic Review*, 97(2): 215–220.

Mixon, S. (2008) The Crisis of 1873: Perspectives from Multiple Asset Classes. *Journal of Economic History*, 68(3): 722–757.

Mixon, S. (2009) Option Markets and Implied Volatility: Past versus Present. *Journal of Financial Economics*, 94(2): 171–191.

Moisseiff, L.S. and Lienhard, F. (1933) Suspension Bridges under the Action of Lateral Forces, with Discussion. *Transactions of the American Society of Civil Engineers*, 98: 1080–1141.

Peterson, I. (1990) Rock and Roll Bridge. *Science News*, 137(22), 344–346.

Petroski, H. (1994) *Design Paradigms: Case Histories of Error and Judgment in Engineering*. New York: Cambridge University Press.

Scott, R. (2001) *In the Wake of Tacoma: Suspension Bridges and the Quest for Aerodynamic Stability*. Reston, VA: ASCE Press.

Strogatz, S. Abrams, D., McRobie, A., Eckhardt, B., and Ott, E. (3 November 2005) Theoretical Mechanics: Crowd Synchrony on the Millennium Bridge. *Nature*, 438: 43–44.

NOTES

1. Scott, 2001, p. 32.
2. "All Pain, No Gain for Convertible Arbitrage Funds in 2008," Credit Suisse/Tremont LLC, November 20, 2008.
3. "All Pain, No Gain for Convertible Arbitrage Funds in 2008," Credit Suisse/Tremont LLC, November 20, 2008.
4. James Mackintosh, "The Search for a Floor in Convertibles," *Financial Times*, September 10, 2008, p. 23.
5. "All Pain, No Gain for Convertible Arbitrage Funds in 2008," Credit Suisse/Tremont LLC, November 20, 2008.
6. Gregory Zuckerman, "Convertible Bonds Cause Hedge Funds Serious Pain," *Wall Street Journal*, November 4, 2008, p. C1.
7. Tom Lauricella, "The Crisis on Wall Street: Short-Sale Ban Wallops Convertible-Bond Market," *Wall Street Journal*, September 26, 2008, p. C2.
8. "Options Market Avoids 'Halt' as SEC Studies Revision on Shorts," Bloomberg News, September 20, 2008.
9. "The London Privilege Market," *Wall Street Journal*, August 13, 1897, p. 2.
10. Scott, 2001, p. 24.
11. Scott, 2001, p. 16.
12. For example, see the discussion in Peterson, 1990.

CHAPTER 2

TOWARD A MARKET SECTOR–BASED COMPOSITE POLITICAL RISK INDICATOR MODEL

John Simpson

ABSTRACT

Model risk involves the risk of model misspecification. In this chapter it is argued that unsystematic risk is an indicator of country-specific and human factors. With a strong theoretical base for a systemic capital asset pricing model specification, these factors may be classified as political risk, which is influenced by social, legal, and cultural effects. When adjusted for degrees of systemic information efficiency and country–industry interaction with global stock markets, this uncomplicated analytical tool could reduce model risk and be used to calculate composite political risk, which can be used as an adjunct to other risk indicators. Government and industry risk analysts may be able to preempt market and political risk problems and to price risk premia in international bank lending with greater frequency of information than is currently available. The example used in study involves a hypothetical country banking industry and its interaction with the global banking industry. Future research will test the model's efficacy.

INTRODUCTION

This chapter is motivated by the need to return to a basic political risk indicator and risk valuation model in light of the recent global turbulence in economies and financial markets. A lesson to be learned from the economic and financial crises of 2008 and the Middle Eastern political turmoil from 2001 is that an additional analytical tool utilizing country and global stock market indexed data would be useful. Much of the blame for the current global financial crisis is leveled at model risk relating to the econometric models produced by financial economists. The model expounded in this chapter perseveres with a theoretical view on the basis that financial economists cannot be held responsible for the human element that produces dysfunction in stock market sectors. Nevertheless, the model discussed in this chapter strives to measure human or political influences in country stock market sectors and takes the example of the banking sector. Such an indicator would provide a broad picture of composite political risk in each country and its impact on the banking sector in that country on a daily basis.

Country risk is the risk that a country will be unable or unwilling to service to its external commitments. The inability to perform relates to economic and financial factors. The unwillingness to perform is related to human factors or, in other words political, social, legal, and cultural factors. In this chapter these human factors are classified under the broad category of composite political risk. Economic and financial risk is objectively assessed by examining factual and historical economic and financial information. Political risk is essentially the slowing down in the meeting of external commitments for overtly political, social, legal, and cultural reasons. Up to this point, however, political risk analysis has been infrequent and subjectively assessed but, based logically, on expert opinion.

Risk ratings agencies,[1] canvassing the opinions of credit risk experts, have attempted to quantify political risk by scoring various countries according to degrees of such risks as corruption, quality of bureaucracy, and history of law and order. These subjective assessments also provide an indication of the propensity of that country to experience and transmit political unrest and this unrest would include extreme political acts. One problem with the ratings is that they do not respond as frequently as they should to randomly arriving good and bad news. Changes in political risk ratings are only reported monthly at best.

A more frequent composite political risk indicator should be available, at least on a daily basis. A simplified model is needed where human behavioral inferences can be made from randomly arriving good and bad news relating to stock markets which are major indicators of a system's financial and economic health. Such an indicator would also reflect the country's degree of interaction with the global economy as another element of composite political risk. It is not suggested that the indicator replace existing risk ratings. It is suggested that it be used as another political risk management tool for use in conjunction with political risk ratings information.

Stock markets are major indicators of a country's economic and financial health and growth. In regression analysis of a basic systemic[2] market model, the regression coefficients represent systematic or market risk. The returns of a country's stock market index (adjusted by the risk-free rate of interest in that country) are regressed on a global stock market index (adjusted for an appropriate global proxy of a risk-free interest rate[3]). Market risk in financial and economic systems is dependent on economic and financial factors which are the same for all and cannot be diversified. These factors are objectively assessed as they are based on factual and historical information.

If it is assumed market risk captures all financial and economic factors in a stock market system in an economy as well as the effect of the interaction of one system with the global stock market system when stock markets are efficient (that is, where stock market returns are a random walk[4]), it may also be assumed that the error term in such a regression represents unsystematic or idiosyncratic risk. The error of such a regression is thus country specific and reflects the composite human element related to political, social, legal, and cultural factors. It contains all subjective factors that are by themselves difficult to measure and predict. It would, apart from the effect of any natural disaster, reflect composite political risk. Herein lies the basis for a new composite political risk indicator. The higher the error, the higher the degree of political risk. This information is as frequently available as stock market share price indices information.

To illustrate the functioning of this model, samples of developed and developing country financial systems are examined. The findings should be of interest to investors, regulators, and government trade and security policy makers. The model could help to anticipate not only financial crises (with the examination of market risk) but also political crises and climates for extreme political acts (with the examination and comparison of unsystematic risk components adjusted for the degree of information efficiency in each compared country).

EXISTING RISK RATINGS SYSTEMS

Sovereign credit rating history is published by world credit risk rating agencies such as Standard and Poor's, Moody's, and Fitch-IBCA. The ratings scales and assessments are comparable and are largely reflective of economic and financial factors. The scales extend from extremely strong ability to repay through to default. The agencies also report credit watches (short-term potential direction) and ratings outlooks (long-term potential directions). This is very useful information particularly when such information can be disaggregated into specific economic and financial risk components.

Simpson (2002) undertook a cross-sectional study of country and international banking risk ratings and economic and financial data for 1995, and from this study several comments may be made about the leading country/sovereign risk ratings agencies. First, the risk ratings from these agencies are highly positively correlated. Second, country risk ratings may be largely replicated using primarily trade performance and debt serviceability data. Third, country risk ratings are also highly positively correlated with international banking risk ratings, thus reflecting the importance of banks as key economic agents. Fourth, pure political risk factors have a very small role in the ratings replication process.

Finally, from a cross-sectional analysis of risk ratings alone, it is not possible to tell whether or not the ratings lead or lag either financial or economic crises. It is also evident that sovereign risk, country risk, and political risk definitions are often confused. In this chapter country risk is the broad concept (total risk) composed of economic and financial risk (systematic component) and political risk (the unsystematic component).

COUNTRY RISK, SOVEREIGN RISK, AND POLITICAL RISK

Most researchers have failed to differentiate country risk components conceptually, thus ignoring pure political risk. Nevertheless, several studies involving the relationship between market data and country risk ratings have been useful. Holthausen and Leftwich (1986), Hand, Holthausen, and Leftwich (1992), and Maltosky and Lianto (1995) argued that sovereign risk rating downgrades were informative to equity markets, but upgrades did not supply markets with new information. Cantor and Packer (1996) examined a sample of developed and emerging markets over the period 1987 to 1994 and found that sovereign risk ratings had a significant impact on bond yield spreads.

Erb, Harvey, and Viskanta (1996) discussed the importance of an understanding of country risk for investors. They found that country risk measures are correlated with future equity returns, but financial risk measures reflect greater information. They also found that country risk measures are also highly correlated with country equity valuation measures and that country equity value-oriented strategies generated higher returns.

Diamonte, Liew, and Stevens (1996) used analyst's estimates of country risk to show that country risk represents a more important determinant of stock returns in emerging rather than in developed markets. They also found that over the past 10 years country risk had decreased in emerging markets and increased in developed markets. They speculated that if that trend continued the differential impacts of country risks in each of those markets would narrow. In this chapter, the size of the error will also reflect the degree of globalization achieved by each country. Any reduction in the average size of the error for individual countries and for all countries will reflect increased globalization and a reduction in political risk.

Larrain, Reisen, and von Maltzan (1997) incorporated country risk data up to the Mexican crisis of 1994 to 1995 and found that the overall impact of ratings changes on bond prices was insignificant. Hill (1998) found that in times of crisis many investors may be determined to minimize exposure to securities affected by country risk until they have more information, but after a period of calm the spreads being offered appear to be too high relative to the risks. After more investors return to the market, the spreads lessen and when there is another crisis the cycle recommences.

Specifically in regard to the Asian currency crisis, Radelet and Sachs (1998) suggested that country/sovereign risk ratings agencies were too slow to react and when they did react it was suggested that their ratings intensified and prolonged the crisis. Ferri, Liu, and Stiglitz (1999) argued that the ratings agencies behaved in a procyclical manner by upgrading country/sovereign risk ratings during boom times and downgrading them during crises.

Reisen and von Maltzan (1999) argued that ratings agencies exacerbated boom-bust cycles in financial markets and put emerging markets at greater risk. Hooper and Heaney (2001) studied regionalism, political risk, and capital market segmentation in international asset pricing. They concluded that multi-index models should be tested that incorporate a regional index, an economic development attribute, commodity factors, and a political risk variable in order to more effectively price securities.

Brooks and colleagues (2004) argued that equity market responses to country/sovereign risk ratings changes revealed significant responses following downgrades. Hooper, Hume, and Kim (2004) found that ratings agencies provided stock markets and foreign exchange markets in the United States with new tradeable information. Ratings upgrades increased stock markets returns and decreased volatility significantly. They also discovered significant asymmetric effects of ratings announcements where the market responses were greater in the case of ratings downgrades. Few authors have examined pure political risk factors.

However, Busse and Hefeker (2005) explored the connection between pure political risk, institutions, and foreign direct investment flows (some of which is channeled into stock markets). They found that government stability, the absence of internal conflicts and ethnic tensions, basic democratic rights, and the ensuring of law and order are highly significant determinants of foreign investment flows. Evidence of the direct adverse effects of extreme political acts on industries and economies is provided and cited in Simpson (2007a, 2007b).

The evidence is mixed but most evidence points to country/sovereign risk having a significant relationship with stock market returns. Some arguments imply that financial crises reflected in reduced stock market returns are the drivers of sovereign risk ratings. If this is the case, risk ratings agencies cannot contribute new information to financial markets for investors and nor could they be useful to regulators and government policy makers. Under the surface, the unwillingness to service external debt may be influenced by economic and financial factors, such as acute shortages of foreign exchange (Bourke and Shanmugam, 1990).

MARKET RISK, MARKET EFFICIENCY, AND CONTAGION

The proposed model is strongly based in portfolio, market efficiency, and financial contagion theories and is therefore more likely (subject to thorough testing) to avoid problems of model risk and model misspecification.

Market Risk Models

Markowitz (1959) developed a basic portfolio model for securities based on a series of broad assumptions relating to investor behavior.[5] He demonstrated that the variance of the returns was a meaningful measure of portfolio risk. Under his assumptions, a single asset or a group of assets in a portfolio is efficient if no other asset or group of assets provides a higher expected rate of return for the same or lower risk or lower risk with the same or higher rate of return. Capital market theory has built on the Markowitz portfolio model and requires similar investor behavioral assumptions with additional assumptions that include consideration of the risk-free rate of return.[6] The proposed model in this chapter contains similar behavioral assumptions and controls for the risk-free rate and thus excess returns.

The capital asset pricing model (CAPM) developed by Sharpe (1964) and arbitrage pricing theory (APT) developed by Ross (1976) differ in that the latter includes several risk factors. This permits a more comprehensive definition of systematic investment risk than that in the CAPM's single market portfolio. Fama and French (1992) found a weak association between the returns of an asset and its beta. They found statistically significant relationships between returns, firm size, and the ratio of book to market values. Roll (1977) suggested that the market proxy for CAPM may not be mean-variance efficient. A criticism of the APT is that the risk factors in the model are not defined in terms of their quantity, but significantly, the APT asserts that a security's return has an expected and an unexpected component. By implication it has a measurable or quantifiable or systematic component based on fact and a difficult to measure or unsystematic component that is based largely on opinion.

This is consistent with the model adopted in this chapter, although the model does not control for the factors discussed by Fama and French (1992) or Roll (1977). The model in this chapter is in accordance with APT assertions where the systematic or quantifiable components are economic and financial in nature and the unsystematic component is reflective of human behavior in a country's political system, which in turn is affected by social, legal, and cultural factors in that country.

More recently, multifactor models have attempted to turn theory into practice and use a variety of macro- and microeconomic factors to explain risk and return. Many of these multifactor models may not be firmly founded in capital market or economic theory and there are many different specifications (Reilly and Brown, 2003). Ultimately, if political, social, legal, and cultural factors are to be taken into account in a model of country stock market returns, it is necessary to assume that they are incorporated in such a basic market model. This avoids the myriad of problems encountered in more advanced versions of the CAPM or the APT or the multifactor models. Reilly and Brown (2003) imply that it is feasible to apply a basic market model to a financial system using systemic stock price index data provided the constituents of the indices used are representative of the industry in the country concerned.

Contagion and Spillovers

The global financial crisis has highlighted the concept of contagion, spillovers, and the importance of the interconnection of global financial markets and economies. Researchers that have studied stock market spillovers are many and include Baig and Goldfajn (1998), Forbes and Rigobon (1999), Dungey and Zhumabekova (2001), Caporale, Cipollini, and Spagnolo (2003), Rigobon (2004), with currency market literature in

Ellis and Lewis (2000). This literature has focused on the manifestation of financial contagion.

Market Efficiency

According to Fama (1970), security markets can be tested for information efficiency at three levels. They are weak-form efficient if stock prices and/or returns are a random walk; semi-strong-form efficient if stock prices and/or returns immediately reflect all available public information; and they are strong-form efficient if stock prices and/or returns reflect all public and private information. However, many developing markets have not achieved even weak-form efficiency. Strong-form efficiency is yet to be attained in even the most developed market, as evidence by the frequency of cases of investigation of insider trading and market manipulation.

It is quite clear that the raw composite political risk scores derived from the market model would need to be adjusted for varying degrees of information efficiency in each system. The scores will only be consistent and provide meaningful indication if all markets have a similar degree of information efficiency or if they are weighted according to the degree of efficiency achieved in each system. The level of market efficiency for a country's stock market can be proven through autocorrelation tests (that is, testing the independence or randomness of prices and/or returns) and event studies (that is, testing the timing of the changes in prices and/or returns around public news events).

THE MODEL

The whole point about political risk is that it is largely composed of legal differences between countries and that these differences are exacerbated by other human factors relating to social and cultural environments. The model that follows cannot control for the various components of political risk. However, the model recognizes that there is a composite political risk value that comprises all these human and legal components. Political news good or bad arrives randomly. If we are examining daily data, models must attempt to provide daily composite political risk indication. Such models do not do this at present. Let it be assumed that the study involves a country banking system. A simple capital assets pricing model is specified and expanded to not only control for the country stock market, but also the interaction with the global banking market and the global stock market.

The major assumptions of the portfolio, efficient markets, and financial contagion theories are carried over to the specified model. That is, that all economic and financial influences on a country banking sector are captured in the regression intercept and its coefficients. All country specific and therefore all social, cultural, legal, and political influences (which collectively make up composite political risk) on the country banking industry are captured in the unsystematic risk component. That is, in the error term of the regression. The basic capital asset pricing model is expanded to be assumed to be applicable to an industry sector rather than a firm within that sector. The model is also expanded to include control for global interaction and relative market efficiency variables. These are unique features of the model which, in its expanded form, is more likely to avoid model risk problems and model misspecification.

Step One

The first step is the specification of a basic market model of unlagged returns variables. According to this model, systematic risk (financial and economic) components are assumed captured in the regression intercept and beta coefficient and idiosyncratic (unsystematic or country specific political, social, legal, and cultural factors) risk components are assumed captured in the error term. Note: Returns of the country banking industry price index, the country stock market index, and the global banking industry price index below are as follows:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{2.1}$$

$$R_{i_t} - R_{ifi_t} = \alpha_{i_t} + \beta 1_{i_t}(R_{iM_t} - R_{fw_t}) + e_{i_t} \tag{2.2}$$

where R_{it} is the return on a country's banking share price index i at time t. R_{ift_t} is the risk-free rate in the country system i at time t. α_{it} and β_{it} are the regression coefficients representing the proportion of systematic or market risk in system i at time t arising from the country stock market and the global stock and banking markets. R_{iM_t} is the return on a country stock market price index at time t. e_{i_t} is the error term of the regression indicating the unsystematic risk in banking system i at time t.

Note 1: The error term is the raw composite political risk component and it also reflects the degree of a country's global interaction in banking markets. Thus the development of the model in Equation (2.1) needs to control for global interaction.

Note 2: The model deals with returns and thus reduces the likelihood of dealing with nonstationary data as well as dealing with serial correlation in the error term. It is, however, likely that heteroscedasticity[7] is persistent in the error terms. This could be controlled for by the specification of a generalized least-squared regression or an autoregressive conditional heteroscedasticity model.

Step Two

Models that try to control for country and political risk have not taken into account the interaction between their stock market returns and the global stock market and, in the context of this chapter, the global banking market. A better way to control for these factors is to examine the interrelationships between the country banking market returns and those of the global stock market (I_{iGM_t}) and the global banking market (I_{iGB_t}). These interrelationships may be proxied by the correlations (ρ_t) between each market in returns at time t where

$$I_{iGM_t} + I_{iGB_t} = \rho_{iGMi} + \rho_{iGB_t} \tag{2.3}$$

The interrelationship terms for the country banking market with the global stock market and the global banking market may then be added to Equation (2.2) as follows:

$$R_{i_t} - R_{ifi_t} = \alpha_{i_t} + \beta 1_{i_t}(R_{iM_t} - R_{ifw_t}) + \beta 2_t(I_{iGM_t}) + \beta 3_t(I_{iGB_t}) + e_{i_t} \tag{2.4}$$

Step Three

The remaining problem is to expand the model to control for the differing levels of information efficiency in different country banking markets. There are several ways that this could be handled. The logical proposal is to test for a random walk using autocorrelation tests on each set of systemic data. This test will indicate whether the country banking sector lacks any degree of information efficiency or is at least weak-form efficient. The analysis could then move to test semi-strong and strong-form efficiency by running event studies around structural breaks in the weak-form efficient systems[8] where the structural breaks are proven to exist around major political or economic events.[9] Each system will be shown to be

1 = Not information efficient
2 = Weak-form efficient
3 = Semi-strong-form efficient
4 = Strong-form efficient

It might be adequate to divide the error term (as it represents a raw composite political risk score out of 100) by the level of information efficiency. That is, divide the inefficient markets error by 1, the weak-form efficient markets by 2, the semi-strong-form efficient markets by 3 and the strong form efficient markets by 4. In reality, in a full analysis of all country banking sectors there will be few, if any, countries in category 4 due to the frequency, even in very developed markets, of insider trading and market manipulation.

Perhaps a better way of capturing the level of efficiency of each banking market would be to specify a dummy variable (that is, a dummy variable denoted D_{ieff_t} with the level of efficiency of each market ascribed a number, for example inefficient market = 4 through strong-form efficient = 1), as follows:

$$R_{i_t} - R_{ifi_t} = \alpha_{i_t} + \beta 1_{i_t}(R_{iM_t} - R_{if_t}) + \beta 2_t(I_{iGM_t}) + \beta 3_t(I_{iGB_t}) + \beta 4(D_{ieff_t}) + e_{i_t} \quad (2.5)$$

Note: The speed with which each country banking market change reacts to changes in the country stock market, the global stock market, and the global banking market could also be obtained by running vector autoregressive model based causality tests on lagged data and/or impulse response functions.[10] The degree of information efficiency of the particular market could be proxied by the relative speed with which each banking market reacts to changes in the domestic stock market, the global stock market, and the global banking market and to any changes in the efficiency of that market.

For Example

Table 2.1 shows the results of hypothetical regression analysis of Equation (2.4) for a sample of developed and developing countries.[11]

Table 2.1 Regression Results of the Basic Banking Market Model for Each Country

Country	Adjusted R-Squared Value Systematic or Market risk	Beta Coefficients Summary	t Statistic of Beta Summary	Standard Errors Summary	Raw Composite Political Risk Score (Regression Errors)
United States	0.6332	1.2979	46.0952	0.0282	**36.68**
United Kingdom	0.4717	1.1467	33.1446	0.0346	**52.83**
Australia	0.0956	1.1048	8.3246	0.0457	**90.44**
Malaysia	0.0266	0.3635	5.8011	0.0553	**97.34**
Philippines	0.0019	0.0884	2.5048	0.0662	**99.53**
Thailand	0.0047	0.0660	1.8247	0.0727	**99.81**

Note 2: The value and significance of regression parameters, t statistics, and standard errors may also be useful for ranking either market risk or unsystematic risk in the country banking industry.

In an overall comparison of the selected countries, the developed country system regressions (particularly those for the United States and the United Kingdom, also, to a lesser extent, Australia) have, on the day of testing, higher adjusted R-squared values, higher regression coefficients, higher t statistics, and lower standard errors than the developing country banking industry sectors. In this way a running regression can be employed, using past daily data up to the date of analysis.

It may be concluded that the developed country banking sectors have higher levels of systematic risk and lower levels of unsystematic risk than the developing country banking sectors, when interacting with their own stock market as well as the global stock and banking markets in returns. When the unsystematic risk component is converted to a score out of 100 for political risk, the ranking of least to most risky country in the sample is the United States, the United Kingdom, Australia, Malaysia, the Philippines, and Thailand. Logically this is the reverse of the market risk (systematic risk) ranking, as more developed countries are likely to have a greater component of banking market risk as their banking sector markets are more information efficient and they possess a greater degree of global integration.

Australia is a developed country with a sophisticated financial system and a stable economic and political environment. A political risk score of more than 90 does not provide a true indication of its riskiness in terms of social, legal, and cultural factors that impact its political willingness to service its external debt and commitments. Australia's banking market might be tested and demonstrated to be semi-strong-form efficient. Hypothetically, the United States and the United Kingdom may have tested as strong-form and semi-strong-form efficient, respectively. Malaysia and the Philippines may have tested as weak-form efficient. Thailand may have tested as lacking any level of information efficiency. The scores could be adjusted daily, until the markets are again tested for levels of efficiency some time in the future when deemed appropriate.

However, if the control for the level of information efficiency of a particular banking market is implemented using a dummy variable [as described in Equation (2.5)], the results of the testing of the model may reveal the following information in Table 2.2. In this table it would be concluded that the U.S. market possesses the lowest composite political risk and the Thailand market the highest. While Australia has a high raw composite political risk rating due to a lack of global interaction, a suitable adjustment has been

Table 2.2 Adjusted Raw Composite Political Risk Scores

Country	Level of Information Market Efficiency	Raw Composite Political Risk Score	Market Efficiency-Adjusted Composite Political Risk Score (Rank)
United States	4	36.68	**9.17 (1)**
United Kingdom	3	52.83	**17.61 (2)**
Australia	3	90.44	**30.15 (3)**
Malaysia	2	97.34	**48.67 (4)**
Philippines	2	99.53	**49.77(5)**
Thailand	1	99.81	**99.81 (6)**

made to this rating by controlling in the final model for the semi-strong level of market efficiency that exists in the Australian market.

Step Four

When the daily errors terms are held for each country banking system they are compared with the composite political risk ratings from risk rating agencies. If there is a high positive correlation between the errors and the risk ratings, this verifies the correct specification of the model and the nonexistence of model risk.

It is likely that the returns of the country banking sector and the political risk scores are stationary series. It is thus likely that the level series of banking industry prices (P_{it}) and level series errors (composite political risk score denoted PR_{it}) are integrated nonstationery processes and that a univariate vector auto regressive (VAR) model can be specified by lagging the variables in the single-period model. The single-period model is as follows:

$$P_{i_t} = \alpha_{i_t} + \beta(PR_{i_t}) + e_{i_t} \tag{2.6}$$

When these variables are optimally lagged in a VAR specification, the VAR-based tests of cointegration and exogeneity can demonstrate the strength of the relationship between country banking industry stock market prices, and the composite political risk variable and the exogeneity of the composite political risk variable can also be tested. Pairwise Granger causality tests and impulse-response functions can indicate whether country banking industry prices are influenced by the composite political risk variable or vice versa. If the prices are driven by political risk, then political risk is an important variable that adds new information to the country banking market and can assist in the pricing of credit risk premia for international lending purposes.

CONCLUSION

In terms of theoretical consistency, developed country banking sectors have higher levels of systematic risk and lower levels of unsystematic risk than those in developing countries. Unsystematic risk, which includes country-specific political factors impacted by social, legal, and cultural effects, particularly legal effects, is greater in developing country banking sectors. These country banking sectors exhibit less information efficiency due to higher political risks in areas such as government stability, corruption, and quality of bureaucracy. Their degree of globalization and interaction with the world market is expected to be less. The specified model controls for global interaction between a country banking industry and the global stock market and the global banking industry with the use of correlation variables.

It is posited that developed economies with banking markets that are semi-strong-form efficient but which possess low global interaction (for example, Australia and New Zealand) still require a substantial reduction in their raw composite political risk score because they are assumed to and can be proven to possess greater information efficiency than developing country systems. The specified model controls for varying degrees of information efficiency in differing country banking industries by introducing a dummy variable that ascribes proven levels of information efficiency to each system.

The specified model has a strong theoretical base in portfolio, market efficiency, and financial contagion theories. If the assumptions of these theories are adopted for the specified model it may be tested for each country banking system. If the errors of the regressions are highly positively correlated with actual composite political risk ratings by reputable risk rating agencies, then the model has zero model risk and is correctly specified. The errors of the model become the composite political risk ratings and VAR-based cointegration and causality tests may be run on the level series. If the new composite political risk rating drives the banking industry prices, it may be assumed that political risk adds new information to the banking industry; and this information may be used by the banking industry in each country to price international lending and credit risk premia.

Previous studies have demonstrated that country/sovereign risk ratings from leading ratings agencies may be replicated using nonpolitical data and largely reflect economic and financial information. The scoring of pure political risk (such as changes of government, corruption, the role of the military, the quality of bureaucracy, and other factors that are either the cause or the effect of social, legal, and cultural factors) by reputable political risk rating agencies remains valuable. However, a composite political risk score appropriate for the banking industry or any industry derived from more frequent market data and adjusted for country market efficiency levels and global interaction is also considered to be useful for investors and policy makers. It is posited that the possibility exists that a significant reduction in market risk in one country might mean that substantial increases in political risk (from overseas or domestically) have been priced into the market. It remains for the model to be properly tested for all country stock market sectors across all countries.

REFERENCES

Baig, T. and Goldfajn, I. (1998) Financial Market Contagion and the Asian Crisis. IMF working paper 155.

Bourke, P. and Shanmugam, B. (1990) Risks in International Lending. In Bourke (ed.), *An Introduction to Bank Lending*. Sydney, Australia: Addison-Wesley Publishing.

Brooks, R., Faff, R., Hillier, D., and Hillier, J. (2004) The National Market Impact of Sovereign Rating Changes. *Journal of Banking and Finance*, 28(1): 233–250.

Busse, M. and Hefeker, C. (2005) Political Risk, Institutions and Foreign Direct Investment. HWWA discussion paper 315.

Cantor, R. and Packer, F. (1996) Determinants and Impact of Sovereign Credit Ratings. *FRBNY Economic Policy Review*, 2(2): 37–54.

Caporale, G.M., Cipollini, A., and Spagnolo, N. (2003) Testing for Contagion: A Conditional Correlational Analysis. *Australasian Meeting of the Econometric Society*, October.

Diamonte, R., Liew, J.M. and Stevens, R.L. (1996) Political Risk in Emerging and Developed Markets. *Finance Analysts Journal*, 52(3): 71–76.

Dungey, M. and Zhumabekova, D. (2001) Testing for Contagion Using Correlations: Some Words of Caution, WP No. PB01-09, Centre for Pacific Basin Monetary and Economic Studies, Economic Research Department, Federal Reserve Bank of San Francisco.

Ellis, L. and Lewis, E. (2000) The Response of Financial Markets in Australia and New Zealand to News about the Asian Crisis. BIS Conference on International Financial Markets and the Implications for Monetary and Financial Stability, Basel, Switzerland.

Erb, C.B., Harvey, C.R., and Viskanta, T.E. (1996) Political Risk, Economic Risk and Financial Risk. Fuqua School of Business working paper 9606.

Fama, E.F. (1970) Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance*, 25(2): 383–417.

Fama, E. and French, K. (1992) The Cross Section of Expected Stock Returns. *Journal of Finance*, 47(2): 427–465.

Ferri, G., Liu, L., and Stiglitz, J. (1999) The Procyclical Role of Rating Agencies: Evidence from the East Asian Crisis. *Economic Notes*, 28(1): 335–355.

Forbes, K.J. and Rigobon, R. (1999) No Contagion, Only Interdependence: Measuring Stock Market Co-Movements. NBER working paper 7267.

Hand, J., Holthausen, R., and Leftwich, R. (1992) The Effect of Bond Rating Agency Announcements on Bond and Stock Prices. *Journal of Finance*, 47(2): 733–752.

Hill, C.A. (1998) How Investors React to Political Risk. *Duke Journal of Comparative and International Law*, 8(2): 283–314.

Holthausen, R. and Leftwich, R. (1986) The Effect of Bond Rating Changes on Common Stock Prices. *Journal of Financial Economics*, 17(1): 57–89.

Hooper, V. and Heaney, R. (2001) Regionalism, Political Risk and Capital Market Segmentation in International Asset Pricing. AFA Meetings, New Orleans.

Hooper, V., Hume, T., and Kim, S. J. (2004) Sovereign Rating Changes—Do They Provide New Information for Stock Markets? University of New South Wales working paper.

ICRG. (2005) International Country Risk Guide. Political Risk Services Group.

Larrain, G., Reisen, H., and von Maltzan, J. (1997) Emerging Market Risk and Sovereign Credit Ratings. OECD Development Centre technical paper 124.

Maltosky, Z. and Lianto, T. (1995) The Incremental Information Content of Bond Rating Revisions: The Australian Evidence. *Journal of Banking and Finance*, 19(5): 891–902.

Markowitz, H. (1959) *Portfolio Selection-Efficient Diversification of Investments*. New Haven, CT: Yale University Press.

Radelet, S. and Sachs, J. (1998) The Onset of the East Asian Financial Crisis. NBER working paper 6680.

Reilly, F.K. and Brown, K.C. (2003) *Investment Analysis and Portfolio Management* (pp. 176–285). Mason, Ohio: Thomson South-Western.

Reisen, H. and von Maltzan, J. (1999) Boom and Bust Sovereign Ratings. OECD Development Centre technical paper 148.

Rigobon, R. (2004). Identification through Heteroskedasticity. *Review of Economics and Statistics*, 85(4): 777–792.

Roll, R. (1977) A Critique of the Asset Pricing Theory's Tests. *Journal of Financial Economics*, 4(2): 129–176.

Ross, S. (1976) The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory*, 13(2): 341–360.

Sharpe, W.F. (1964) Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance*, 19(3): 425–442.

Simpson, J.L. (2002) An Empirical Economic Development Based Model of International Banking Risk and Risk Scoring. *Review of Development Economics*, 6(1): 91–102.

Simpson, J.L. (2007a) Expert Political Risk Opinions and Banking System Returns: A Revised Banking Market Model. *American Review of Political Economy*, 5(1): 14–28.

Simpson, J.L. (2007b) The Effects of Extreme Political Acts and Political Risk on International Banking Systems. *The International Journal of Business and Management*, 2(5): 3–19.

NOTES

1. For example, ICRG (2005) published by the Political Risk Services Group.
2. Note the different meanings in this chapter of the words *systemic* and *systematic*. Systemic relates to a financial or economic system. Systematic relates to systematic risk, which is one of the components of total risk in a financial system.
3. For example, a eurocurrency interbank offered rate.
4. Where returns immediately reflect all available randomly arriving news and returns are independent of each other.
5. For example, investors maximize one-period expected utility and their utility curves demonstrate diminishing marginal utility of wealth, and, for a given risk level, investors prefer higher to low returns and for a given level of return lower for higher risk.
6. Other principal assumptions are that capital markets are in equilibrium with all investments priced accurately in line with their risk levels and that there is no inflation or change in interest rates or inflation is fully anticipated. Also assumed is that there are no taxes or transaction costs in buying or selling assets.
7. Unequal variance of the error term.
8. For example, Chow and other structural break tests indicate significant changes in regression parameters comparing the full sample period, the period up to the break and the period after the break. If prices and/or returns have demonstrated the production of abnormal positive returns prior to the event, this may mean either insider trading or market manipulation based on withheld good news.

9. Practically, one could select reputable empirical evidence of country stock market efficiency and grade countries accordingly.
10. These tests provide a one standard deviation shock to the endogenous (dependent) variable. The response time of the exogenous (independent) variables can then be observed.
11. The hypothetical sample of country banking sectors was selected to represent strong, globally integrated developed economies in the United States, United Kingdom, and Australia as well as a group of developing South East Asian economies in Thailand, Malaysia, and the Philippines.

PART II

MODEL RISK RELATED TO EQUITY AND FIXED INCOME INVESTMENTS

CHAPTER 3

ANALYSTS' EARNINGS FORECASTS, PORTFOLIO SELECTION, AND MARKET RISK PREMIA

An Empirical Comparison of Four Different Valuation Approaches

Franziska Becker, Wolfgang Breuer, Marc Gürtler

ABSTRACT

The most relevant, practical impediment to an application of the Markowitz portfolio selection approach is the problem of estimating return moments, in particular, return expectations. We present four valuation models based on analysts' forecasts that are utilized for the derivation of implied expected stock returns: the dividend discount model, the residual income model, the Ohlson/Jüttner-Nauroth model, and the discounted cash flow model. In an empirical capital market study, we implement these four models and several benchmark strategies to obtain and compare the out-of-sample performance of the respective strategies from June 1, 1999, to December 1, 2008. Furthermore, we estimate corresponding market risk premia for six out of all nine strategies we examine. Though theoretically equivalent, practical

results across the four approaches under consideration vary to a great extent. Moreover, it is hard to systematically beat a simple naïve portfolio selection strategy even on the basis of analysts' forecasts.

INTRODUCTION

The estimation of expected returns of investments is one of the central problems of portfolio management and asset pricing. Within the framework of portfolio management, this information is used to carry out quantitative portfolio optimizations. In the context of asset pricing, the knowledge of the cost of capital is necessary, since it is used for discounting expected cash flows. The determination of cost of capital is closely related to the calculation of the market risk premium which corresponds to the difference of the weighted average cost of equity of all risky securities and the riskless interest rate.

While the application of historical return realizations for the estimation of expected returns has been discussed for quite a long time, the idea of utilizing analysts' forecasts concerning corporate cash flows and performance indicators is quite new. Depending on which analyst's estimation is taken as a basis, four different models can be distinguished: the dividend discount model (DDM), the residual income model (RIM), the Ohlson/Jüttner-Nauroth model (OJM), and the discounted cash flow model (DCM). In the second section of this chapter, all four approaches are briefly introduced. Thereafter we follow Breuer, Feilke, and Gürtler (2008) who utilized a modification of the DDM for the case of a nonflat term structure of interest rates, and we extend the application of this approach to the RIM and the DCM. However, a similar straightforward implementation is not possible for the OJM.

Despite the theoretical equivalence of these four approaches, resulting estimators for expected one-period stock returns need not be identical as a consequence of inconsistent parameter fixation and varying underlying assumptions. Although in Breuer, Feilke, and Gürtler (2008), the DDM has already been examined as a basis for portfolio optimization and for the estimation of market risk premia, it seems interesting to study the adequacy of the other three approaches as well. We do this based on monthly data for HDAX or DAX100 firms from January 1, 1994, to January 1, 2009. Our results are presented in this chapter's fourth section. In general, portfolio selection results and corresponding estimators for market risk premia vary considerably across all four approaches. Thereby, we find that the OJM performs best for low-risk aversion and the RIM outperforms all other models for moderate and higher risk aversion.

UTILIZING ANALYSTS' FORECASTS FOR EXPECTED RETURN ESTIMATION

The Dividend Discount Model (DDM)

We start our analysis with a closer look at DDM. This approach can be traced back to Williams (1938) and Gordon (1959, 1966). According to this model, at a given point in time t, the expected stock return $\hat{\mu}_i^{(D)}$ with respect to firm i will be determined as an internal rate of return of the expected dividend payments on the basis of the market value of equity $EQ_{i,t}$ of firm i at time t. For reasons of practicability, typically a two-phase model is employed. In the first phase, starting at time t, estimations of dividends $\hat{D}_{i,\tau}$ of the points in time $\tau = t+1, \ldots, t + T$ are available. In the second phase, starting from time $t + T + 1$ on, only a constant relative growth $g_i^{(D)}$ of dividends is assumed. If necessary, an intermediate phase will be modeled in order to procure a gradual adaption to the dividend growth rate from the end of the first phase to the final phase (see, for instance, Stotz, 2004, 2005). The two-phase model leads to

$$EQ_{i,t} = \sum_{\tau=1}^{T} \frac{\hat{D}_{i,t+\tau}}{(1+\hat{\mu}_i^{(D)})^{\tau}} + \frac{\hat{D}_{i,t+T} \cdot \left(1 + g_i^{(D)}\right)}{(\hat{\mu}_i^{(D)} - g_i^{(D)}) \cdot (1 + \hat{\mu}_i^{(D)})^{T}} \tag{3.1}$$

Under the assumption that analysts' forecasts are (on average) true and representative for all investors on the capital market, $\hat{\mu}_i^{(D)}$ in Equation (3.1) by definition corresponds to the cost of equity of the firm over its entire lifetime. However, it is not necessarily identical to the expected one-period return from holding this firm's stocks from t to $t + 1$. Despite this problem, this additional (implicit) assumption is exactly what is needed to make use of $\hat{\mu}_i^{(D)}$ for portfolio management purposes. Then it is possible to determine the expected return of the market portfolio as the weighted average of the expected individual stock returns. Subtraction of the riskless interest rate leads to the one-period market risk premium. This premium is an essential component of many valuation formulas, especially those which are based on the capital asset pricing model (CAPM).

In the case of heterogeneous expectations, 'the' market risk premium no longer exists, thus such a question leads nowhere. However, the approach, according to Equation (3.1) can still be used to determine $\hat{\mu}_i^{(D)}$, although this value can no longer be interpreted as the cost of equity over a firm's lifetime. However, it may still be utilized for portfolio selection purposes.

Seemingly, several shortcomings of the basic approach have been introduced so far. First of all, $\hat{\mu}_i^{(D)}$ will be generally valid only as an average discount rate over the whole time horizon of the firm under consideration. In particular, in situations with a nonflat term structure of interest rates, corresponding one-period discount rates will not be constant. In order to take varying risk free interest rates into account, one may replace $\hat{\mu}_i^{(D)}$ by $\hat{\phi}_i^{(D)} + r_{t+\kappa}^{(f)}$, with $r_{t+\kappa}^{(f)}$ being the risk free (forward) interest rate from time $t+\kappa-1$ to $t+\kappa$ and $\hat{\phi}_i^{(D)}$ being the average equity risk premium over the whole lifetime of the firm under consideration. Under the assumption of constant risk-free interest rates from time T on, Equation (3.1) becomes

$$EQ_{i,t} = \sum_{\tau=1}^{T} \frac{\hat{D}_{i,t+\tau}}{\prod_{\kappa=1}^{\tau}(1+\hat{\phi}_i^{(D)} + r_{t+\kappa}^{(f)})} + \frac{\hat{D}_{i,t+T} \cdot \left(1+g_i^{(D)}\right)}{(\hat{\phi}_i^{(D)} + r_{t+T}^{(f)} - g_i^{(D)}) \cdot \prod_{\kappa=1}^{T}(1+\hat{\phi}_i^{(D)} + r_{t+\kappa}^{(f)})} \tag{3.2}$$

and could be used to determine $\hat{\phi}_i^{(D)}$. Adding the risk-free interest rate $r_{t+\kappa}^{(f)}$ would then result in another, hopefully better, estimator for the expected one-period stock return.

In Equation (3.2) it is assumed that information for forward risk-free interest rates are available for the same time horizon as analysts' forecasts for future dividends. However, with T_1 being the former relevant time horizon and T_2 being the latter, we typically have $T_1 > T_2$. This simply means that in Equation (3.2), T has to be substituted by T_1 and for future dividends between $\tau = t + T_2 + 1$ and $\tau = t + T_1$, we have to write $\hat{D}_{i,t+\tau} = \hat{D}_{i,t+T_2} \cdot (1 + g_i^{(D)})^{t+\tau-T_2}$.

Moreover, up to now, we have been silent on tax considerations. However, corporate taxes are already allowed for by analysts when estimating future dividend payments. It remains to take a closer look at personal taxes. Nevertheless, as situations may be quite different for different investors, we will refrain from this possible model extension (see, however, Breuer, Feilke, and Gürtler, 2008, for an analysis of this topic).

The Residual Income Model (RIM)

The DDM is not the only approach by which the market value of a firm's equity can be computed as a sum of discounted earnings figures. Another one is RIM. It makes use of the fact that, under certain circumstances, the market value of equity cannot only be computed cash flow oriented, according to Equation (3.1), but can also be determined income oriented as a result of the Preinreich-Lücke theorem (Preinreich, 1937; Lücke, 1955). As long as an increase in corporate income either leads to higher dividends or to a higher book value of equity and thus a balance sheet extension (so-called clean-surplus condition), the market value of equity can be determined by discounting the corporate residual gains with the appropriate cost of equity (estimator) $\hat{\mu}_i^{(RG)}$

$$EQ_{i,t} = BEQ_{i,t-1} + \sum_{\tau=1}^{T} \frac{\widehat{RG}_{i,t+\tau}}{(1+\hat{\mu}_i^{(RG)})^{\tau}} + \left(\frac{\widehat{RG}_{i,t+T} \cdot (1+g_i^{(RG)})}{(\hat{\mu}_i^{(RG)} - g_i^{(RG)}) \cdot (1+\hat{\mu}_i^{(RG)})^{T}} \right) \tag{3.3}$$

In addition, in this context, it is essential that expected gains $G_{i,t}$ are reduced by the absolute costs of equity $\hat{\mu}_i^{(RG)} \cdot BEQ_{i,t-1}$ (with $BEQ_{i,t-1}$ being the book value of equity at time $t-1$) to arrive at the so-called residual gain or income $RG_{i,t}$ (see Breuer, 2007, p. 460)

$$RG_{i,t} = G_{i,t} - \hat{\mu}_i^{(RG)} \cdot BEQ_{i,t-1} \tag{3.4}$$

In the same way as with respect to DDM, we may utilize the solution $\hat{\mu}_i^{(RG)}$ of Equation (3.3) as an estimator for the one-period expected stock return and, after subtraction of the relevant risk-free interest rate, as an estimator for the corresponding one-period equity risk premium of firm i. Moreover, with the replacement of $\hat{\mu}_i^{(RG)}$ by $\hat{\phi}_i^{(RG)} + r_{t+\kappa}^{(f)}$, we could once again allow for nonflat term structures of interest rates. Corporate taxes have already been taken into account by analysts when estimating future expected gains and, as in the case of DDM, we refrain from personal taxes.

The Ohlson/Jüttner-Nauroth Model (OJM)

The OJM is another approach that aims at determining a firm's equity value by discounting earnings figures by the equity cost of capital. In contrast to the RIM, it is based on the assumption of a constant growth rate $g_i^{(G)}$ of expected firm gains. In order to circumvent the necessity of the clean-surplus condition for valuation approaches which rely on gains instead of payments, the concept of excess income is introduced. The excess income $z_{i,t+1}$ at time $t+1$ corresponds to the discounted income growth of the time period from $t+1$ to $t+2$ minus the reinvested retained income of time $t+1$ (with the cost of equity estimator $\hat{\mu}_i^{(OJM)}$ as the relevant discount rate)

$$z_{i,t+1} = \frac{1}{\hat{\mu}_i^{(OJM)}} \cdot \left(\hat{G}_{i,t+2} - \hat{G}_{i,t+1} - \hat{\mu}_i^{(OJM)} \cdot \left(\hat{G}_{i,t+1} - \hat{D}_{i,t+1} \right) \right) \tag{3.5}$$

For this excess income, a constant growth rate $g_i^{(z)}$ is assumed as well

$$z_{i,\tau+1} = (1 + g_i^{(z)}) \cdot z_{i,\tau} \tag{3.6}$$

Under this assumption, one gets (Ohlson and Jüttner-Nauroth, 2005, p. 354; Reese, 2005, p. 7)

$$EQ_{i,t} = \frac{\hat{G}_{i,t+1}}{\hat{\mu}_i^{(OJM)}} + \frac{\hat{G}_{i,t+2} - \hat{G}_{i,t+1} - \hat{\mu}_i^{(OJM)} \cdot \left(\hat{G}_{i,t+1} - \hat{D}_{i,t+1}\right)}{\hat{\mu}_i^{(OJM)} \cdot \left(\hat{\mu}_i^{(OJM)} - g_i^{(z)}\right)} \tag{3.7}$$

Thus, the value of equity is determined as a capitalized income $\hat{G}_{i,t+1}/\hat{\mu}_i^{(OJM)}$ of the following period (i.e., the value of a perpetuity amounting to $\hat{G}_{i,t+1}$) and an additional premium for future excess income. Again, corporate taxes have been deducted from the income.

An application of Equation (3.7) is possible even without the validity of the clean-surplus condition. However, a flat-term structure of interest rates is assumed in order to arrive at Equation (3.7). As a nonflat modification of OJM is not necessarily solvable, we refrain from extending our examination in this direction. Furthermore, only future expected gains of the next two periods and the future dividend of time $t + 1$ have to be estimated. These are clearly practical advantages of OJM. However, in order to reach these advantages, constant growth rates $g_i^{(G)}$ and $g_i^{(z)}$ as well as a flat term structure of interest rates must be assumed.

The Discounted Cash Flow Model (DCM)

In contrast to the three approaches discussed above, DCM is used to derive an estimator for a firm's total cost of capital; this is then applied to compute the corresponding value for a firm's equity cost of capital. To be precise, DCM is based on the fact that the value of a firm is calculated as the net present value of the future expected free cash flows. With market values $EQ_{i,t}$ and $DB_{i,t}$ for equity and debt at time t, $CF_{i,\,t+\tau}$ as the (as seen from time t) estimated free cash flow of company i at time $t + \tau$, and $g_i^{(CF)}$ as the growth rate of the free cash flow, a firm's weighted costs of capital $WACC_i$ over its whole lifetime can be calculated as follows:

$$EQ_{i,t} + DB_{i,t} = \sum_{\tau=1}^{T} \frac{\widehat{CF}_{i,t+\tau}}{\left(1+\widehat{WACC}_i\right)^{\tau}} + \frac{\widehat{CF}_{i,t+T} \cdot (1+g_i^{(CF)})}{\left(\widehat{WACC}_i - g_i^{(CF)}\right) \cdot \left(1+\widehat{WACC}_i\right)^{T}} \tag{3.8}$$

Under the assumptions that $WACC$ is identical to a firm's expected one-period return on total firm value and that debt financing is risk free, we may derive from this a corresponding estimator of expected one-period stock return by the help of the following relationship:

$$\widehat{WACC}_i = \frac{EQ_{i,t}}{EQ_{i,t} + DB_{i,t}} \cdot \hat{\mu}_i^{(WACC)} + \frac{DB_{i,t}}{EQ_{i,t} + DB_{i,t}} \cdot r_{t+1}^{(f)} \cdot \left(1 - s^{(F)}\right) \qquad (3.9)$$

From Equation (3.9), it is once again possible to derive an estimator $\hat{\mu}_i^{(WACC)}$ for the expected one-period stock return and the corresponding stock risk premium $\hat{\mu}_i^{(WACC)} - r_{t+1}^{(f)}$. Different from the other three approaches, the corporate tax rate $s^{(F)}$ explicitly appears in the relevant formula. Since debt capital reduces the tax burden of companies, the term $(1 - s^{(F)})$ is added. For example, for Germany, a typical value of $s^{(F)} = 40\%$ is utilized. At any rate, as is the case for analysts' dividend and gain forecasts, in analysts' future cash flow estimations, corporate taxes are taken into account as well.

Again, it would be possible to allow for nonflat term structures of forward risk-free interest rates in Equation (3.8) by replacing $WACC_i$ by the corresponding sum $\hat{\phi}_i^{(WACC)} + r_{t+\kappa}^{(f)}$ of a risk premium on total firm value and the relevant risk-free forward interest rate.

PRACTICAL APPLICATIONS

In general, valuation Equations (3.1), (3.3), (3.7), and (3.8) are equivalent. Nevertheless, in practical applications, resulting estimates for expected one-period stock returns may be different, as there are different underlying assumptions which may be violated and, in addition, parameter estimates may not be consistent with each other even if there are no violations of assumptions. This gives rise to the question of which approach performs best in practical applications of portfolio selection and market risk premium estimation.

The accuracy of the four estimation models described hinges crucially on the validity of analysts' dividend estimates. In fact, there is much evidence for analysts' forecasts to be 'biased.' In particular, two basic causes can be adduced for this. On the one hand, there are reasons for analysts' forecasts tending to be too optimistic, even under the assumption of unlimited rationality. For instance, an investment bank can ameliorate the business connection to a company by publishing a positive assessment and thus realize additional earnings (Dugar and Nathan, 1995; Lin and McNichols, 1998; Michaely and Womack, 1999; Wallmeier, 2005). However, the advantage that has been achieved by this means stands in opposition to the disadvantage of the analyst's forecast impreciseness which reduces his reputation and consequently generates lower earnings. This mitigates analysts' incentives for too optimistic forecasts, but does not completely eliminate it (Jackson, 2005). Besides such a rational argument, there is also a cognitive bias which results from analysts' misperception and improper processing of information as, empirical analyses show. Whether analysts' forecasts generally react too strongly (e.g., DeBondt and Thaler, 1990) or too weakly (Abarbanell, 1991; Abarbanell and Bernard, 1992) to new information, or that over- and underreactions depend on the type of information (Easterwood and Nutt, 1999), is still under discussion. The same is true for the question of whether analysts rather tend to give overly pessimistic (Brown, 1996; Chan, Karecski, and Lakonishok, 2007) or overly optimistic forecasts (Crichfield, Dyckman, and Lakonishok, 1978; O'Brien, 1988; Lys and Sohn, 1990; Hong and Kubik, 2003) or both depending on the situation (Abarbanell and Lehavy, 2003). Breuer, Feilke, and Gürtler (2008) show that biases in analysts' forecasts are only of secondary importance concerning questions of portfolio management, if all expected returns determined on this basis are similarly biased up- or downward. Moreover, as we want to compare different approaches which are all based on analysts' estimates, this issue is of only minor relevance for our investigation, as it will not primarily influence the relative performance of all four approaches. We therefore refrain from a separate analysis of this problem in this chapter.

EMPIRICAL EXAMINATION

In this section, the different models are applied with real capital market data.[1] For this purpose, monthly data from January 1, 1994 to January 1, 2009 of all HDAX and DAX100 stocks are available. Since DAX100 was replaced by HDAX not before March 24, 2003, it will form the basis of our empirical examination in the beginning. DAX100 was composed of 30 DAX shares and 70 MDAX shares. However, HDAX is composed of 30 DAX shares, 50 restructured MDAX shares, and 30 TecDAX-shares. We are considering 100 shares until March 24, 2003 and 110 shares thereafter. Only stocks that are included in the index (either DAX100 before April 2003 or HDAX from April 2003) at a specified point in time are considered in the optimization at this time. Furthermore, we examine whether all data are available for the implementation of the models, for example, the consensus forecasts regarding dividends per share or earnings per share. If not all data that are required for the empirical examination are available, the respective share is not used in the optimization for this point in time. This procedure allows for an optimization that, at a certain point in time, contains a stock, which is no longer in the index after this point in time; thus, in this empirical examination, a survivorship bias is nonexistent. The number of shares which are optimized over time fluctuates between 8 and 49. Risk-free interest rates are calculated on the basis of the interest yield curve as provided by the Deutsche Bundesbank for (remaining) maturities of 1 to 15 years. Data are extracted from the Thomson Reuters Datastream database.

For any point in time t from June 1, 1999 to December 1, 2008, we apply all portfolio selection strategies under consideration and several benchmark strategies. What all our portfolio selection strategies have in common is that (if necessary) we estimate excess return variances and covariances on the basis of historical return realizations by way of a single index model, as the number of stocks sometimes exceeds the number of months in the estimation period, which is 24. The resulting variance-covariance matrix would not be invertible and the optimization could not be applied. The approaches under consideration only differ with respect to the estimation or consideration of expectation values of excess stock returns.

As outlined in the second section of this chapter, we take into account four different valuation models to determine estimates of expected excess returns on the basis of analysts' forecasts. In case 1, we apply Equation (3.2) of DDM for the special situation with a nonflat term structure of interest rates with $T_1 = 15$ and $T_2 = 3$. The stock price and the dividend forecasts for the next three years are taken from Datastream. Case 2 implements Equation (3.4) according to RIM for a nonflat term structure and also $T_1 = 15$ and $T_2 = 3$. In order to determine expected returns the following values are required: dividends per share for the next two years, earnings per share for the next three years and the last reported book value per share. With the help of dividends per share and earnings per share and applying the clean-surplus-relation the book value per share for the next three years can be calculated. The current book value per share is determined as the compounded last reported book value per share $BEQ_{i,lr}$ using the predicted return on equity ($\hat{G}_{i,t+1}$ / $BEQ_{i,lr}$) as the appropriate interest rate

$$BEQ_{i,t} = BEQ_{i,lr} \cdot \left(1 + \hat{G}_{i,t+1} / BEQ_{i,lr}\right)^{\frac{months(\,fiscal\ year\ end\ firm\ i,t\,)}{12}} \tag{3.10}$$

where months (fiscal year-end firm i, t) denotes the number of months between the end of the fiscal year of firm i and the estimation date t.[2] Case 3 is based on Equation (3.7) according to OJM and case 4 on the DCM of Equations (3.8) and (3.9). The OJM requires the knowledge of current stock price, dividends per share for the next year, and earnings per share for the next two years. The cash flow per share for the next three years, the current stock price, and the debt to equity ratio are input parameters of DCM. With respect to the determination of the expected rate of return via

Equation (3.9), firm debt is assumed to be risk free. The true cost of debt capital will typically be greater than $r_{t+1}^{(f)}$ due to a positive credit spread. However, such data are not available from the database. The OJM is applied with a flat term structure of interest rates because the nonflat model is not necessarily solvable. For all four valuation models, it is necessary to define the annual growth rate g_i of the specified forecasts of firm i beyond the horizon of current analysts' forecasts. While Stotz (2004) used $g_i = g = 6$ % on the basis of the average annual growth rate of the (nominal) gross national income in Germany from 1980 to 1999, we apply the average of the last 20 years of the annual growth rate of the (nominal) gross national income just before the point in time when the respective portfolio selection takes place. We thus take into account time-varying estimators for future national growth rates. As a robustness check we also tested the average of the last 5 or 10 years of the gross national income growth rate and our results were stable.

Cases 5 to 7 describe portfolio selection strategies that do not rely on explicit estimations of expected excess returns. In case 5 (holding the market portfolio), at each point in time from June 1, 1999, until December 1, 2008, the investor realizes a portfolio structure of risky assets that is identical to that of the whole supply of all equity shares under consideration (MKT). In case 6, we assume that the investor adheres to a risky subportfolio with a share of $1/n_t$ for each of the different stocks with n_t being the number of assets in the optimization at time t (EW). Case 7 is defined by the holding of the variance minimal stock portfolio at each point in time from June 1, 1999 until December 1, 2008 (MVP). Case 8 refers to the estimation of expected excess returns at a point in time t as the average of 24 (monthly) excess return realizations from $t-23$ to t (HIST). Case 9 is an application of a Bayesian approach of Kempf, Kreuzberg, and Memmel (2002) (BAYES). Within the approach, a prior estimator of expected stock returns is combined with information on historical return realizations summarized in the vector M_{hist} of average historical stock returns. The prior estimator ϕ is called the grand mean and is identical to the average historical return realization over all stocks under consideration. The mean of the predictive density function M_{KKM} of expected stock returns is then computed as a weighted average of $\phi \cdot \mathbf{1}$ ($\mathbf{1}$: a vector of n_t ones) and M_{hist}. To be more specific, define Σ as the estimator of the variance-covariance matrix based on the single index model, E as the unit matrix, T as the number of historical return realizations under consideration (for our analysis: $T = 24$) and τ^2 as the estimated variance of the historical return estimators of each stock in relation to the grand mean as the corresponding expectation value

$$\tau^2 = \frac{1}{n_t - 1} \cdot \sum_{i=1}^{n_t} \left(M_{hist,i} - \phi \right)^2 \tag{3.11}$$

Then

$$M_{KKM} = \left(\Sigma + T \cdot \tau^2 \cdot E \right)^{-1} \cdot \Sigma \cdot \phi \cdot 1 + \left(\Sigma + T \cdot \tau^2 \cdot E \right)^{-1} \cdot T \cdot \tau^2 \cdot M_{hist} \tag{3.12}$$

As a peculiarity of Bayesian approaches, we also have to adjust the variance-covariance estimator

$$\Sigma_{KKM} = \Sigma \cdot \left(E + \tau^2 \cdot \left(\Sigma + T \cdot \tau^2 \cdot E \right)^{-1} \right) \tag{3.13}$$

For all nine cases under consideration, we compute 115 successive optimal (myopic) portfolios from June 1, 1999 to December 1, 2008 with a time horizon of one month each, subject to short sales constraints $0 \le x_j \le 1$ for all stocks $j = 1, \ldots, n_t$ under consideration. We maximize the following preference function

$$\theta = \mu_P - \frac{\lambda}{2} \sigma_P^2 = X'M - \frac{\lambda}{2} X'\Sigma X \quad \rightarrow \max_X ! \tag{3.14}$$

Assuming constant absolute risk aversion and normally distributed returns involves this certainty equivalent, which can be maximized instead of the expected utility (Anderson and Bancroft, 1952; Freund, 1956). Here, X stands for the n_t vector of asset weights; Σ describes the $N \times N$ variance-covariance matrix of asset returns; M is the expected return vector; λ a parameter of relative[3] risk aversion; and μ_P and σ_P^2 are the expected portfolio return and the portfolio variance, respectively. We apply different values for λ from 0.5 to 3.5 to account for more or less risk averse investors.[4] The optimization is constrained as portfolio weights are restricted to be between 0 and 1 and the sum of asset weights has to be 1. This seems to be the practically most relevant consideration, as many investors cannot realize short sales or get into debt.

For all nine cases and 115 periods of the rolling optimization procedure, we determine corresponding realized portfolio excess rates $r_{P,t+1}^{(exc)}$ of return at time $t+1$. By this procedure, we get 115 excess return realizations for nine strategies, that means we obtain $9 \times 115 = 1{,}035$ optimized portfolios. With $\hat{\mu}^{(exc)}$ as the mean excess return over all 115 excess return realizations and $\hat{\sigma}^{(exc)}$ as the corresponding estimator for the excess return standard deviation, we are able to compute (estimators for) resulting Sharpe ratios $\phi_s = \hat{\mu}^{(exc)} / \hat{\sigma}^{(exc)}$ for any portfolio selection strategy under consideration. We extend our analysis to several other performance measures, e.g., Jensen's alpha, the Treynor ratio, and the certainty equivalent according to Equation (3.14).

Table 3.1 presents our empirical results for the nine strategies described above with a risk aversion parameter of 0.5. The OJM performs best for all considered performance measures (Sharpe ratio, 5.74%). Regarding the Sharpe ratio, the equally weighted portfolio (3.41%) and the MVP (1.10%) follow directly. There are only four strategies that are able to attain a positive Sharpe and Treynor ratio over the optimization period. The other rankings (negative ratios) are not listed, as negative Sharpe and Treynor ratios imply the superiority of a simple riskless investment. The DCM also performs quite well except for the certainty equivalent. Notice that the market portfolio is not able to reach a positive excess return on average from June 1, 1999, to December 1, 2008.

Table 3.2 displays analogous results for a risk aversion parameter of 2.0. Now RIM outperforms all other strategies for three out of four performance measures. The OJM and the equally weighted portfolio share rank 2 and DDM reaches a positive Sharpe and Treynor ratio at least (rank 5).

The outcomes for the most risk averse investor with $\lambda = 3.5$ are presented in Table 3.3. Again RIM performs best (Sharpe ratio, 4.93%) and the equally weighted portfolio follows directly (Sharpe ratio, 3.41%). The OJM is ranked on position 4 on average. Therefore, a less risk averse investor should implement OJM and a more risk averse investor should hold a portfolio according to RIM. Particularly we can advise these two models from the models based on analysts' forecasts for an implementation. This is remarkable, as, e.g., OJM is based on the assumption of a flat term structure of

Table 3.1 Sharpe Ratios, Certainty Equivalents, Jensen's Alphas, and Treynor Ratios for Nine Different Portfolio Strategies With Risk Aversion $\lambda = 0.5$

No.	Strategy	Sharpe Ratio	Rank	Certainty Equivalent	Rank	Jensen's Alpha	Rank	Treynor Ratio	Rank
1	DDM	−0.0563		−0.0085	8	−0.0065	8	−0.0068	
2	RIM	−0.0123		−0.0022	7	−0.0008	7	−0.0013	
3	OJM	0.0574	1	0.0060	1	0.0075	1	0.0067	1
4	DCM	0.0092	4	−0.0014	6	0.0021	3	0.0011	4
5	MKT	−0.0086		0.0008	5	0.0000	5	−0.0006	
6	EW	0.0341	2	0.0039	2	0.0028	2	0.0027	2
7	MVP	0.0110	3	0.0027	3	0.0008	4	0.0011	3
8	HIST	−0.0717		−0.0114	9	−0.0090	9	−0.0094	
9	BAYES	−0.0075		0.0008	4	−0.0001	6	−0.0008	

Table 3.2 Sharpe Ratios, Certainty Equivalents, Jensen's Alphas, and Treynor Ratios for Nine Different Portfolio Strategies With Risk Aversion $\lambda = 2$

No.	Strategy	Sharpe Ratio	Rank	Certainty Equivalent	Rank	Jensen's Alpha	Rank	Treynor Ratio	Rank
1	DDM	0.0032	5	−0.0038	7	0.0007	5	0.0004	5
2	RIM	0.0412	1	−0.0002	3	0.0037	1	0.0044	1
3	OJM	0.0277	3	−0.0025	5	0.0029	2	0.0029	2
4	DCM	−0.0096		−0.0150	9	−0.0006	7	−0.0012	
5	MKT	−0.0086		−0.0032	6	0.0000	6	−0.0006	
6	EW	0.0341	2	0.0007	2	0.0028	3	0.0027	3
7	MVP	0.0110	4	0.0012	1	0.0008	4	0.0011	4
8	HIST	−0.0181		−0.0122	8	−0.0015	9	−0.0022	
9	BAYES	−0.0311		−0.0020	4	−0.0014	8	−0.0032	

Table 3.3 Sharpe Ratios, Certainty Equivalents, Jensen's Alphas, and Treynor Ratios for Nine Different Portfolio Strategies With Risk Aversion $\lambda = 3.5$

No.	Strategy	Sharpe Ratio	Rank	Certainty Equivalent	Rank	Jensen's Alpha	Rank	Treynor Ratio	Rank
1	DDM	0.0180	5	−0.0029	4	0.0015	5	0.0019	5
2	RIM	0.0493	1	−0.0008	2	0.0034	1	0.0050	1
3	OJM	0.0224	3	−0.0045	6	0.0020	4	0.0023	4
4	DCM	−0.0087		−0.0246	9	−0.0004	8	−0.0011	
5	MKT	−0.0086		−0.0071	7	0.0000	7	−0.0006	
6	EW	0.0341	2	−0.0026	3	0.0028	2	0.0027	2
7	MVP	0.0110	6	−0.0003	1	0.0008	6	0.0011	6
8	HIST	0.0223	4	−0.0108	8	0.0026	3	0.0026	3
9	BAYES	−0.0466		−0.0040	5	−0.0020	9	−0.0048	

interest rates, while the DDM which performs considerably poorer seems to be a much more straightforward approach.

Nevertheless, across all three scenarios, the equally weighted portfolio performs quite well. As argued in DeMiguel, Garlappi, and Uppal (2009), it is indeed difficult to find a portfolio selection strategy that is systematically better than a simple naïve diversification.

Table 3.4 presents the annual market risk premia for six out of all nine strategies. After estimating individual expected stock returns, the implied expected excess return of the market portfolio is computed for given current market capitalizations. As strategies 5 to 7 are not based on explicit expected return estimations, we can only refer to the remaining strategies in order to compute the market risk premia. According to Table 3.4, the

Table 3.4 Annual Market Risk Premia for Six Different Approaches

No.	Strategy	Market Risk Premia
1	DDM	0.0129
2	RIM	0.0444
3	OJM	0.0863
4	DCM	0.1525
8	HIST	0.1353
9	BAYES	0.1069

market risk premium estimator is lowest for DDM. Low risk premia resulting from implied expected returns are in line with the study of Claus and Thomas (2001) who utilized RIM. With respect to our analysis, RIM and OJM involve premia of 4.44 percent and 8.63 percent, respectively. This is much lower than the risk premium of the models based on historical data with 13.53 percent for the historical strategy and 10.69 percent for the Bayesian approach. The DCM provides the highest market risk premium with 15.25 percent.

Apparently, market risk premia estimators vary considerably across these six approaches as well as across those four which are based on analysts' forecasts. Certainly, deeper theoretical investigations are necessary in order to identify the reasons for this finding. At any rate, one may not conclude that those approaches which perform best for portfolio selection purposes are also most suited for market risk premia estimation, as a superior performance points to the fact that the corresponding return expectations are *not* identical to market expectations. The adequate estimation of market risk premia will thus remain an issue.

CONCLUSION

The object of this chapter was to analyze the relevance of estimation models on the basis of analysts' forecasts for portfolio selection purposes and for the computation of market risk premia estimates. We presented four valuation models that are utilized for the derivation of implied expected stock returns: the dividend discount model, the residual income model, the Ohlson/Jüttner-Nauroth model, and the discounted cash flow model. In our empirical study, we implemented these four models and several benchmark strategies in order to obtain and to compare the out-of-sample performance of the corresponding portfolio selection strategies. We found that the Ohlson/Jüttner-Nauroth model performs best for a low risk aversion and the residual income model outperforms all other models for moderate and higher risk aversions. Nevertheless, simple naïve diversification performs in a satisfying way in all three scenarios. This chapter presents further evidence that it is difficult to systematically beat simple passive portfolio selection strategies.

Furthermore, we estimated the market risk premium with six out of all nine strategies under consideration. The lowest premium was estimated with the dividend discount model, and the discounted cash flow model produced the highest market risk premium. As market risk premia vary considerably across all approaches under consideration, further effort will be necessary in order to identify the computation method which is most adequate.

REFERENCES

Abarbanell, J. (1991) Do Analysts' Forecasts Incorporate Information in Prior Stock Price Changes? *Journal of Accounting and Economics*, 14(2): 147–165.

Abarbanell, J. and Bernard, V. (1992) Tests of Analysts' Overreaction/Underreaction to Earnings Information as an Explanation for Anomalous Stock Price Behavior. *Journal of Finance*, 47(3): 1181–1207.

Abarbanell, J. and Lehavy, R. (2003) Biased Forecasts or Biased Earnings? The Role of Reported Earnings in Explaining Apparent Bias and Over/Underreaction in Analysts' Earnings Forecasts. *Journal of Accounting and Economics*, 36(1): 105–146.

Anderson, R.L. and Bancroft, T.A. (1952) *Statistical Theory in Research*. New York: McGraw-Hill.

Breuer, W. (2007) *Investition I: Entscheidungen bei Sicherheit*, 3rd ed. Wiesbaden, Germany: Gabler.

Breuer, W. and Gürtler, M. (2008) Kimball's Prudence and Two-Fund Separation as Determinants of Mutual Fund Performance Evaluation. *Kredit und Kapital*, 41(4): 501–539.

Breuer, W., Feilke, F., and Gürtler, M. (2008). Analysts' Dividend Forecasts, Portfolio Selection, and Market Risk Premia. In: J. Kalcsics and S. Nickel (eds.), *Operations Research Proceedings 2007* (pp. 251–256).

Brown, L.D. (1996) Analyst Forecast Errors and Their Implications for Security Analysis: An Alternative Perspective. *Financial Analysts Journal*, 52(1): 40–44.

Chan, L.K.C., Karecski, J., and Lakonishok, J. (2007) Analysts' Conflicts of Interest and Biases in Earnings Forecasts. *Journal of Financial and Quantitative Analysis*, 42(4): 893–914.

Claus, J. and Thomas, J. (2001) Equity Premia as Low as Three Percent? Evidence from Analysts' Earnings Forecasts for Domestic and International Stock Markets. *Journal of Finance*, 56(5): 1629–1666.

Crichfield, T., Dyckman, T., and Lakonishok, J. (1978) An Evaluation of Security Analysts' Forecasts. *Accounting Review*, 53(3): 651–668.

Daske, H., Gebhardt, G., and Klein, S. (2006) Estimating the Expected Cost of Equity Capital Using Analysts' Consensus Forecasts. *Schmalenbach Business Review*, 58(1): 2–36.

DeBondt, W.F.M. and Thaler, R.H. (1990) Do Security Analysts Overreact? *American Economic Review*, *Papers and Proceedings*, 80(2): 52–57.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009) Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy? *Review of Financial Studies*, 22(5): 1915–1953.

Dugar, A. and Nathan, S. (1995) The Effect of Investment Banking Relationships on Financial Analysts' Earnings Forecasts and Investment Recommendations. *Contemporary Accounting Research*, 12(1): 131–160.

Easterwood, J.C. and Nutt, S.R. (1999) Inefficiency in Analysts' Earnings Forecasts: Systematic Misreaction or Systematic Optimism? *Journal of Finance*, 54(5): 1777–1797.

Freund, R.J. (1956) The Introduction of Risk into a Programming Model. *Econometrica*, 24(3): 253–263.

Gordon, M.J. (1959) Dividends, Earnings, and Stock Prices. *The Review of Economics and Statistics*, 41(1): 99–105.

Gordon, M.J. (1966) *The Investment*, *Financing and Valuation of the Corporation*. Homewood, IL: Irwin.

Hong, H. and Kubik, J.D. (2003) Analyzing the analysts: Career concerns and biased earnings forecasts. *Journal of Finance*, 58(1): 313–352.

Jackson, A.R. (2005) Trade Generation, Reputation and Sell-side Analysts. *Journal of Finance*, 60(2): 673–717.

Kempf, A., Kreuzberg, K., and Memmel, C. (2002) How to Incorporate Estimation Risk into Markowitz Optimization. In: P. Chamoni (ed.), *Operations Research Proceedings 2001* (pp. 175–182). Berlin: Springer.

Lin, H. and McNichols, M. F. (1998) Underwriting Relationships, Analysts' Earnings Forecasts and investment Recommendations. *Journal of Accounting and Economics*, 25(1): 101–127.

Lücke, W. (1955) Investitionsrechnung auf der Basis von Ausgaben oder Kosten? *Zeitschrift für handelswissenschaftliche Forschung*, 7(7): 310–324.

Lys, T.Z. and Sohn, S. (1990) The Association Between Revisions of Financial Analysts Earnings Forecasts and Security-price Changes. *Journal of Accounting and Economics*, 13(4): 341–363.

Michaely, R. and Womack, K.L. (1999) Conflict of Interest and the Credibility of Underwriter Analyst Recommendations. *Review of Financial Studies*, 12(4): 653–686.

O'Brien, P. (1988) Analysts' Forecasts as Earnings Expectations. *Journal of Accounting and Economics*, 10(1): 53–83.

Ohlson, J.A. and Jüttner-Nauroth, B. (2005) Expected EPS and EPS Growth as Determinants of Value. *Review of Accounting Studies*, 10(3): 349–366.

Preinreich, G.A.D. (1937) Valuation and Amortization. *Accounting Review*, 12(3): 209–224.

Reese, R. (2005) Alternative Modelle zur Schätzung der erwarteten Eigenkapitalkosten—Eine empirische Untersuchung für den deutschen Aktienmarkt. Working paper.

Stotz, O. (2004) *Aktives Portfoliomanagement auf Basis von Fehlbewertungen in den Renditeerwartungen*. Berlin: Duncker & Humblot.

Stotz, O. (2005) Active Portfolio Management, Implied Expected Returns, and Analyst Optimism. *Financial Markets and Portfolio Management*, 19(3): 261–275.

Wallmeier, M. (2005) Analysts' Earnings Forecasts for DAX100 Firms During the Stock Market Boom of the 1990s. *Financial Markets and Portfolio Management*, 19(2): 131–151.

Williams, J.B. (1938) *The Theory of Investment Value*. Cambridge, MA: Harvard University Press.

NOTES

1. For the general empirical setting, see Breuer, Feilke, and Gürtler (2008).
2. See Daske, Gebhardt, and Klein (2006), p. 11.
3. To be more precise, λ is an investor's relative risk aversion for a rate of portfolio return of just zero.
4. In Breuer and Gürtler (2008) it is shown that plausible values of λ approximately are between 1 and 2. We slightly extend the interval upward and downward to allow for nonconventional investors.

CHAPTER 4

THE MARKET TIMING ABILITY OF AUSTRALIAN SUPERANNUATION FUNDS

Nonlinearities and Smooth Transition Models

George Woodward and Robert Brooks

ABSTRACT

The market timing ability of fund managers remains a major research issue in finance. In the Australian context, greater personal responsibility is now required for retirement incomes via superannuation fund investments of individuals. Accordingly, the performance of these funds is an issue of major public policy importance. To some extent any assessment of performance is model specific, and therefore a possible concern is that results obtained are due to model risk or misspecification. This chapter extends the previous literature by exploring the market timing ability of Australian superannuation funds using the class of nonlinear smooth transition models. This chapter develops a tiered definition of market timing ability in terms of strong form, mild form and weak form. The analysis shows that the conventional models used to measure market timing ability of Treynor and Mazuy (1966), Henriksson and Merton (1981), and Merton (1981) are appropriate for measuring market timing ability and therefore in this context the findings are not due to model risk. Consistent with the

previous literature, we find only limited evidence in support of strong market timing ability.

INTRODUCTION

The issue of an aging workforce has led Australia, like many other countries, to move from a publicly funded pension system to a retirement incomes system in which individuals now have greater personal responsibility for their retirement incomes through long-term investment vehicles such as superannuation investments. This has led to a strong increase in individual share ownership and managed funds investments by individuals. Thus, comparable to a range of other countries, there is now a greater interest in fund performance in the Australian context, around a range of issues such as performance persistence and market timing ability. The question of market timing ability is of great importance, in terms of the benefits of investments through managed funds.

The market timing ability of Australian funds have been studied by a number of authors. Hallahan and Faff (1999) find little evidence of market timing ability on the part of Australian equity trusts. International equity funds have been studied by Benson and Faff (2003) and Gallagher and Jarnecic (2004), who find an absence of market timing ability and no gains in terms of superior returns from active management. Do, Faff, and Wickramanayake (2005) find no market timing ability on the part of Australian hedge funds. Sinclair (1990) finds evidence of perverse market timing ability on the part of Australian managed funds. However, Sawicki and Ong (2000) find reduced evidence of perverse market timing ability for managed Australian funds after conditioning on publicly available information along the lines of Ferson and Schadt (1996), with a particularly strong role found for the dividend yield variable. Further, Prather, Middleton, and Cusack (2001) find no evidence of market timing ability for Australian managed funds in an analysis focusing on the role of management teams. Holmes and Faff (2004) study multisector trusts and find greater evidence of negative market timing once an allowance is made for volatility timing in the analysis. Gallagher (2001) finds that the lack of evidence supporting market timing ability is not overcome by including a wider set of assets than only equities in the benchmark portfolio. The range of models used in the analysis raise queries about the extent to which the results are driven by model risk or misspecification; the present study explores this matter via consideration of a more general model.

The majority of previous studies make use of the Treynor and Mazuy (1966) quadratic market models and/or the Henriksson and Merton (1981) and Merton (1981) dual beta models in their assessment of the measurement of market timing ability. In a recent paper Chou, Chung, and Sun (2005) propose extending the Henriksson and Merton (1981) model to a more general case of the threshold regression model in the assessment of the performance of U.S. mutual funds. This chapter extends the Chou, Chung, and Sun (2005) analysis in a variety of ways. First, using the Henriksson and Merton (1981) and Merton (1981) dual beta models as a base, we provide an extended definition of market timing ability that allows for three tiered states of market timing ability, specifically, strong, mild, and weak. Second, we then provide a more general nonlinear model in which to test market timing ability that nests the Treynor and Mazuy (1966) quadratic market model, the Henriksson and Merton (1981) and Merton (1981) dual beta model, and the logistic smooth transition regression model of Teräsvirta and Anderson (1992), and its subsequent generalization by Granger and Terasvirta (1993) as special cases. Third, we then conduct a test of the market timing ability of Australian wholesale and retail superannuation funds using this generalized modeling framework. Interestingly, our results support the use of the Treynor and Mazuy (1966) and Henriksson and Merton (1981) and Merton (1981) models as appropriate for measuring market timing ability, relative to the more complex nonlinear models. Thus, our findings show in the present context that model risk or misspecification issues are not driving our timing performance results. Further, consistent with the previous literature, we find only limited evidence in support of strong market timing ability.

The plan of this chapter is as follows. In the second section in this chapter, we outline our modeling framework that develops our tiered approach to measuring market timing ability, and our nested modeling structure. This chapter's third section then provides the empirical results of our test of the market timing ability of Australian wholesale and retail superannuation funds. The final section then contains some concluding remarks.

MODELING FRAMEWORK

A major model on which tests of the market timing ability of funds is assessed is the dual beta model of Henriksson and Merton (1981) and Merton (1981). In this section we propose a more general framework for modeling market timing that allows us to generalize to a tiered definition (strong, mild, weak) of market timing ability and a general model that includes the Treynor and Mazuy (1966) quadratic market model, the Henriksson and Merton (1981) and Merton (1981) dual beta model, the Terasvirta and Anderson (1992) and Granger and Terasvirta (1993) logistic smooth transition model, and the constant beta model as special cases. Our aim is to derive a basis for determining the market timing ability of fund managers in the general case of K market conditions and when fund managers react in a continuous and smooth manner to all changes in market conditions.

For the purpose of the analysis we assume that the capital asset pricing model is the base model, however, we allow beta to vary depending on market conditions. This, in part captures the Pettengill, Sundaram, and Mathur (1995) critique of the Fama and French (1992) results, although we potentially allow for the more general setting of K states. Thus, we allow the betas to vary depending on the contemporaneous excess market return, $(R_{mt} - R_{ft})$ into K sets of market conditions. As an extension of Merton (1981), we partition the excess market return into K ordered subsets defined by the boundary parameters $\xi_1 < \xi_2 < \cdots < \xi_{K-1}$. Assuming that K discretely different target betas $\eta_1, \eta_2, \cdots, \eta_K$ are chosen by the fund manager, where η_K is chosen if the forecast of the excess market return falls *kth* in the interval, $R^{(k)}$ then, $\beta(R^*) = \eta_1 I_1 + \cdots + \eta_k I_k$, where, $R_t^* = (R_{mt} - R_{ft})$, I_i is an indicator variable for the i^{th} set of market conditions. Provided the fund manager is rational, then $\eta_k < \eta_{k+1}$ for all $k = 1, \cdots, K-1$.

Let q^k; $k = 1, \ldots, K$; $0 \leqslant q^k \leqslant 1$ represent the fund manager's unconditional probability that the excess market return will fall in the k^{th} subinterval. Let p_{ik} be the known conditional probability that the one step ahead excess market return will fall in interval k at time $t+1$, given that it actually fell in interval i at time period t. Then

$$\beta_i = E[\beta(t) | (R_m - R_f)_t \in R^{(i)}] = \sum_{k=1}^{K} p_{ik}\eta_k \text{ and } E[\beta(t)] = b = \sum_{i=1}^{K}\sum_{k=1}^{K} q^i p_{ik}\eta_k = \sum_{i=1}^{K} q^i \beta_i$$

In this setting $\beta(t)$ is a random variable and we can define $\theta(t) = [\beta(t) - b]$ as the unanticipated component of β. Then conditional on $(R_{mt} - R_{ft}) \in R^{(i)}$ we have

$$\theta_i = \begin{cases} \eta_1 - b \ \textit{with probability}\ p_{i1} \\ \eta_2 - b \ \textit{with probability}\ p_{i2} \\ \vdots \\ \eta_K - b \ \textit{with probability}\ p_{iK} \end{cases} \text{ and } \bar{\theta} i = E[\theta \mid (R_{mt} - R_{ft}) \in R^{(i)}] = \sum_{k=1}^{K} p_{ik}\eta_k - b$$

Therefore, the excess fund return can be expressed as $(R_{pt} - R_{ft}) = \alpha + [\beta(t)](R_{mt} - R_{ft}) + \varepsilon_{pt}$ or $(R_{pt} - R_{ft}) = \alpha + [b + \theta(t)](R_{mt} - R_{ft}) + \varepsilon_{pt}$, which by substitution we can write as,

$$(R_{pt} - R_{ft}) = \alpha + [b + \sum_{i=1}^{K} I_i(\bar{\theta}_i + u_{it})](R_{mt} - R_{ft}) + \varepsilon_{pt}, \text{ where } u_{it} = \theta(t) - \bar{\theta}_i$$

Therefore, $(R_{pt} - R_{ft}) = \alpha + [\sum_{i=1}^{K} I_i (\sum_{k=1}^{K} p_{ik}\eta_k)](R_{mt} - R_{ft}) + w_{pt}$, where

$$w_{pt} = (R_{mt} - R_{ft}) \cdot \sum_{i=1}^{K} I_i \cdot u_{it} + \varepsilon_{pt}, \text{ and we have } (R_{pt} - R_{ft})$$

$$= \alpha + \sum_{i=1}^{K} \beta_i x_{it} + w_{pt}, \text{ where}$$

$$x_{it} = \begin{cases} (R_{mt} - R_{ft}) \; if \; (R_{mt} - R_{ft}) \in R^{(i)} \\ 0 \;\; otherwise \end{cases}, \; \beta_i = \sum_{k=1}^{K} p_{ik}\eta_k, \text{ and}$$

$$E[(R_{pt} - R_{ft}) \mid (R_{mt} - R_{ft}) \in R^{(i)}] = \alpha + (b + \bar{\theta}_i) \cdot E[(R_m - R_f) \mid (R_m - R_f) \in R^{(i)}].$$

Ordinary least squares estimation of the equation $y_t = \alpha + \beta_1 x_{1t} + \cdots + \beta_K x_{Kt} + w_t$, with x_{it} defined as above, $y_t \equiv (R_{pt} - R_{ft})$ and w_t the disturbance term, provides consistent estimates of the parameters $\beta_i = \Sigma_{k=1}^{K} p_{ik} \eta_k$ for $I = 1 \ldots$ K. Therefore, if we assume that a rational fund manager will be one who attempts to increase/decrease risk only when market conditions improve/worsen, a fund manager will be a good market timer if and only if $\sum_{s=1}^{K} p_{is} \eta_s > \sum_{s=1}^{K} P_{ks} \eta_s$ for $i > k$. Then to generalize these conditions to the case where the fund manager responds to all movements in the excess market return, we let $K \to \infty$ and assume that the fund manager's systematic risk is a continuous function of the excess market return $(R_m - R_f)$. We then define three tests for market timing ability. In order of the strength of the timing ability, they are as follows.

Strong-Form Market Timing Ability

This test requires that the value of beta satisfies $\beta(x + \zeta) > \beta(x)$ for all x and all $\zeta > 0$. In other words, beta is a monotonic increasing function of the excess market return. Strong-form market timing ability is concluded when the estimates of the coefficients governing the positive relationship between the excess market return and risk are significant.

Mild-Form Market Timing Ability

This test requires that the value of beta satisfies $\beta(x + \zeta) \geq \beta(x)$ for all x and all $\xi > 0$ and $\beta(x + \zeta) > \beta(x)$for some x and all $\zeta > 0$. Mild form market timing ability is concluded when the threshold model applies and $\hat{\delta}$, the estimated differential value of the up market slope, is significantly positive.

Note that this test does not give special attention to the behavior of beta for $(R_m - R_f)$ near zero. Despite the fact that up and down beta formulations have proven popular in tests of asset pricing (see Pettengill, Sundaram, and Mathur (1995)) and are based around the switching at a return interval around zero, we adopt a more general approach that allows for multiple thresholds that do not necessarily give special weighting to the zero value. If a fund beta is an increasing function of the excess market return in a small neighborhood of zero but a decreasing function of the excess market return everywhere else, then for most values of the excess market return, the manager acts in a perverse manner. In other words, even though the return on the fund is positively affected by his actions when the excess market return is in the small neighborhood around zero, outside this neighborhood the rebalancing decisions are perverse. Note that for the threshold Merton (1981) model, with a threshold value denoted by c, the mild-form market timing ability holds when the up-market beta is larger than its down-market counterpart because in that case $\beta(x + \xi) \geq \beta(x)\ \forall\ x$ and for $\forall\ \xi > 0$ and $\beta(c - \varsigma_1) < \beta(c + \varsigma_2)\ \forall\ \varsigma_1, \varsigma_2 > 0$. The point we are making with our mild-form market timing ability conditions is that for any gain to occur through timing decisions, the fund manager must rebalance appropriately for at least some values of the excess market return while not taking perverse timing decisions elsewhere in other return intervals.

Weak-Form Market Timing Ability

This is a test that requires that the average value of beta corresponding to $(R_{mt} - R_{ft}) < Median\ (R_{mt} - R_{ft})$ is less than the average value of beta corresponding to $(R_m - R_f)_t \geqslant Median\,(R_m - R_f)_t$. In other words since we have an estimated beta for each value of $(R_{mt} - R_{ft})$, we can take the average of this series on each side of the median value of $(R_{mt} - R_{ft})$. The procedure used for this 'test' is not inferential but rather based on descriptive statistics alone. We simply compare the average value of the estimated up-market beta with the average estimated value of the down-market beta. If the average value of the estimated up-market beta is larger than the average value of the estimated down-market beta, we conclude in favor of weak-form market timing ability.

The point of this condition is that for a fund manager to make a favorable difference, at the very least, we would expect that on average when the market is up the beta will be larger than when it is down. Consistent with the approach above of not assigning a special a priori significance to the zero value, we choose the median as the demarcating value.

The most general form of the empirical model that we consider is $y_t = \alpha_1 + \alpha_2 \mathrm{D}_t + \beta_1 x_{1t} + \delta \cdot D_t \cdot x_{1t} + \beta_2 x_{2t}\ \beta_3 x_{3t} + \beta_4 x_{4t} + \beta_5 x_{5t} + \beta_6 x_{6t} + \varepsilon_t$, where $y_t = (R_{pt} - R_{ft})$, $x_{1t} = R_{mt} - R_{ft}$ and $x_{lt} = (R_{mt} - R_{ft}) \cdot R_t^{*l-1}$ for $l = 2,3,4,5,6$, R_p, R_m, R_f and R_t^* represent the return on the fund portfolio, the market portfolio, the risk-free rate, and the market condition transition variable, respectively. In addition, D_t is where the excess market return is greater than a threshold and zero otherwise.

This general formulation nests within a number of the popular models used in the assessment of market timing ability and thus allows analysis of the importance of model risk. In the Merton (1981) testing framework $R_t^* = (R_{mt} - R_{ft})$ is the contemporaneous excess market return. The logistic smooth transition model (LSTM) can be approximated by a model in which α_2, β_5, β_6, and δ are set equal to zero. The quadratic market model of Treynor and Mazuy (1966) can be obtained by setting α_2, β_3, β_4, β_5, β_6, and δ equal to zero. The threshold dual beta model of Merton (1981) can be obtained by setting α_2, β_2, β_3, β_4, β_5, and β_6 equal to zero. Each of these models can be extended to allow the intercept to vary dichotomously with changes in market conditions by allowing for α_2 values not equal to zero. Finally, setting all of the parameters to zero except for α_1 and β_1 produces a model where the beta risk is constant. The Treynor and Mazuy (1966) model has been used in previous tests of Australian fund performance (see Sawicki and Ong (2000), Gallagher (2001), Prather, Middleton, and Cusack (2001), Benson and Faff (2003), Gallagher and Jarnecic (2004), Holmes and Faff (2004)). The Henriksson and Merton (1981) and Merton (1981) dual beta model has also been used in previous tests of Australian fund performance (see Sinclair (1990), Hallahan and Faff (1999), Gallagher (2001), Prather, Middleton, and Cusack (2001), Benson and Faff (2003)).

The first stage of our empirical analysis involves the nonlinearity tests of Luukkonen, Saikkonen, and Teräsvirta (1988) and Tsay (1989). The Luukkonen, Saikkonen, and Teräsvirta (1988) testing methodology has been shown by these authors (1988) and Petrucelli (1990) to have good power against the threshold (dual beta model with endogenous threshold) and LSTM forms in small samples. In addition it was designed to have good power against nonlinearity of a general form. For cases where we cannot reject the null hypothesis of linearity, we do not proceed with further modeling and conclude that there is an absence of market timing ability. However, if the null hypothesis of linear model is rejected, we then go on to estimate the general model and all 13 nested nonlinear alternatives that can be produced by setting zero restrictions on the alpha and beta parameter values. From this set of 13 models we then choose a best model on the basis of the modified Akaike information criterion (AIC) derived by Fox (2000). After a model is chosen, we then perform the strong, mild, and weak form tests of market timing ability.

EMPIRICAL RESULTS

Australian superannuation funds data obtained from the ASSIRT rating agency were used for our empirical analysis. Monthly returns of 30 wholesale and 40 retail funds for the period May 1991 to July 2002 were collected. We used the All Ordinaries Accumulation Index as a market proxy and the 13-week Treasury bill as a proxy for the risk-free asset. The excess market return was calculated as the difference between the return on the market portfolio and the risk-free rate.

We selected balanced funds with investment in multiple sectors to ensure that our sample captured active changes in portfolio weights in response to changing market conditions. Wholesale funds represent those superannuation funds for which the contribution from the employee/investor is indirect. The employer of the company makes the investment choice. Although some large employers allow some choice to employees, the choice is quite limited. For retail superannuation funds, employees invest directly in their fund of choice. Generally, retail funds are more expensive than wholesale funds because they engage financial advisers to market the funds. They also charge entry and exit fees to contributors and spend more on advertising to attract investors. Thus retail funds tend to be more competitive than wholesale funds. Wholesale funds enjoy more economies of scale and may not need investor-level competition. This data is unique in that with it we are able to address the role that choice plays on fund performance. This is a major public policy issue with the federal government in Australia legislating for greater choice in funds for employees.

Although the transition variable, R_t^*, can be any indicator of market conditions, our empirical results are based only on an excess market return–based model since this is the variable that must be used in tests of market timing ability. We begin our model selection process by testing for nonlinearity using the Luukkonen, Saikkonen, and Teräsvirta (1988) and Tsay (1989) test statistics. The rejection counts for each of the tests and the significance levels are reported in Table 4.1. In general, the results vary across the tests. For both the retail and wholesale fund, there is only slight evidence of nonlinearity.

Table 4.1 reports the results of nonlinearity testing using the Luukkonen, Saikkonen, and Teräsvirta (1988) and Tsay (1989) test statistics for the sample of wholesale and retail funds. Rejection counts are reported at three significance levels (1%, 5%, 10%) for the three versions of the Luukkonen, Saikkonen, and Teräsvirta (1988) test (S1, S1*, S3) and the two versions of the Tsay (1989) test (Tsay and Tsay*).

For cases where any of the five nonlinearity tests reject, we then estimated the most general form of the empirical model and all of the 13 nested nonlinear alternatives and choose the model with best performance on Fox's (2000) modified AIC. Using the excess market return as the transition variable, we estimate the nonlinear models for 14 of the 30 wholesale

Table 4.1 Nonlinearity Testing for Wholesale and Retail Funds

	S_1	S_1^*	S_3	*Tsay*	*Tsay**
WHOLESALE FUNDS (N = 30)					
Transition Variable					
$R_m - R_f$					
1%	0	1	0	0	2
5%	1	2	3	1	5
10%	7	4	7	5	9
RETAIL FUNDS (N = 40)					
$R_m - R_f$					
1%	0	0	0	1	0
5%	4	3	2	2	3
10%	9	5	7	13	9

funds, and 22 of the 40 retail funds. In our modeling we explore setting the threshold at zero versus endogenous estimation of the threshold value. In all cases the modified AIC supports the models with the endogenously estimated threshold values and as such these are the focus of the results reported in this chapter.

The results are reported for the wholesale funds in Table 4.2 and for the retail funds in Table 4.3. For the wholesale funds, the modified AIC tends to choose either the Treynor and Mazuy (1966) or the Merton (1981) dual beta model as the preferred model with a time-varying intercept. The endogenously estimated thresholds are evenly spread across positive and negative values. A total of five funds exhibit abrupt beta transition (the cases where δ is in the model chosen by the modified AIC), while the remaining nine funds exhibit smooth transition around the threshold. In terms of market timing ability, only one fund (w7) shows strong market timing ability in that beta is a monotonic increasing function of the transition variable and β_2 is significantly positive. Another two funds (w10 and w29) also exhibit weak-form market timing ability in that their average beta in the upper regime is larger than their average beta in the lower regime.

For the retail funds, the modified AIC tends to choose the endogenous threshold Merton (1981) dual beta model as the preferred form with and without a time-varying intercept. The endogenously estimated thresholds are evenly spread across positive and negative values. In total, 13 of the funds exhibit abrupt transition (the cases where δ is in the model chosen by the modified AIC), while the remaining nine funds exhibit smooth transition

Table 4.2 Estimates of the Nonlinear Models for Wholesale Funds: Excess Market Return as Transition Variable

Fund	α_1	α_2	β_1	β_2	β_3	β_4	β_5	β_6	δ	c|L	AIC	Model	$\bar{\beta}_L$	$\bar{\beta}_U$
w2	−0.189	0.596	0.463						−0.062	−0.605	0.8836	DBM	0.44	0.40
	(−1.046)	(2.760)	(13.655)						(−1.359)	49				
w4	−0.482	1.029	0.554	−0.007						−0.605	0.8570	TM	0.57	0.53
	(−2.227)	(3.298)	(16000)	(−2.003)						49				
w7	−0.288	−0.824	0.217	0.013						4.903	0.4754	TM	0.19	0.26
	(4.903)	(−2.181)	(10.615)	(3.642)						118				
w10	−0.749	0.915	0.183	0.004						−4.415	0.5359	TM	0.17	0.20
	(−1.857)	(2.255)	(7.451)	(0.981)						16				
w19	0.173	0.714	0.422	−0.018	−0.002					4.198	0.7681		0.44	0.33
	(2.055)	(2.377)	(13.561)	(−3.242)	(−4.051)					111				
w27	0.859	−0.931	0.731	−0.016	−0.003					−1.796	0.8022		0.73	0.63
	(2.776)	(−2.453)	(11.042)	(−3.064)	(−3.971)					38				
w29	0.138	−9.100	0.450						1.125	5.146	0.2461	DBM	0.45	0.72
	(1.571)	(−0.707)	(16.34)						(0.639)	119				
w31	0.851	−0.766	0.634	−0.017	−0.002					−1.796	0.8124		0.65	0.54
	(2.939)	(−2.196)	(9.963)	(−3.425)	(−2.161)					38				
w33	0.208		0.523						−0.118	−4.415	0.7848	DBM	0.43	0.41
	(2.529)		(9.474)						(−1.825)	16				
w36	0.066	0.562	0.401						−0.133	0.860	0.8117	DBM	0.40	0.27
	(0.572)	(2.698)	(13.553)						(−2.246)	69				
w37	0.276		0.560						−0.125	1.912	0.8169	DBM	0.56	0.47
	(2.496)		(21.351)						(−2.018)	86				
w40	−0.034	1.015	0.192	−0.075	0.011	0.002	−0.000	−0.000		0.860	0.774		1.40	−0.4
	(−0.204)	(2.847)	(1.813)	(−3.167)	(2.378)	(2.666)	(−2.117)	(−2.162)		69				
w42	0.267	1.108	0.394	−0.019						4.198	0.7948	TM	0.44	0.33
	(2.785)	(2.835)	(13.070)	(−3.450)						111				
w43	0.288	0.805	0.384	−0.016						4.198	0.7341	TM	0.42	0.33
	(2.773)	(2.206)	(12.096)	(−3.372)						111				

AIC, Akaike information criterion; DBM, dual beta model; TM, Treynor and Mazuy model.

Table 4.3 Estimates of the Nonlinear Models for Retail Funds: Excess Market Return as Transition Variable

Fund	α_1	α_2	β_1	β_2	β_3	β_4	β_5	β_6	δ	c|L	AIC	Model	$\bar{\beta}_L$	$\bar{\beta}_U$
rms36	0.208		0.29						−0.116	−2.202	0.5226	DBM	0.23	0.17
	(1.972)		(8.008)						(−2.103)	30				
rms37	1.161	−1.322	0.461	−0.050	−0.005	0.0006	0.00005			−1.649	0.4954		0.5	0.28
	(3.196)	(−3.463)	(5.447)	(−2.728)	(−1.928)	(2.228)	(1.796)			39				
rms38	1.084	−1.079	0.383						−0.212	−1.649	0.3071	DBM	0.29	0.17
	(3.009)	(−2.867)	(5.573)						(−2.762)	39				
rms320	−2.317	2.442	−0.150						0.328	−4.415	0.4967	DBM	0.1	0.18
	(−4.348)	(4.531)	(−2.441)						(4.966)	16				
rms321	−1.980	2.058	−0.053						0.229	−4.415	0.5181	DBM	0.12	0.18
	(−3.885)	(3.98)	(−0.797)						(3.219)	16				
rms324	−1.115	1.192	0.039						0.176	−2.476	0.6440	DBM	0.16	0.22
	(−4.949)	(5.028)	(1.365)						(4.714)	24				
rms56	−0.413	0.529	0.244	0.001	0.008	0.0003	−0.0001	−0.0000		−2.202	0.8454		0.31	0.32
	(−1.067)	−1.408	(4.63)	(0.047)	(2.642)	(0.522)	(−2.800)	(−1.642)		30				
rms57	−0.656	0.794	0.235						0.092	−2.202	0.7575	DBM	0.29	0.33
	(−2.521)	(2.874)	(6.067)						(1.895)	30				
rms74	0.039		0.329						−0.192	−2.476	0.2878	DBM	0.2	0.14
	−0.321		(4.239)						(−2.063)	24				
rms75	0.119		0.53						−0.154	−2.146	0.7375	DBM	0.45	0.38
	−0.999		(16.840)						(−2.595)	24				
rms710	−0.004	2.291	0.423						−0.431	4.333	0.8072	DBM	0.42	0.28
	(−0.056)	(2.836)	(24.135)						(−3.182)	112				

(Continued.)

Table 4.3 (Continued)

Fund	α_1	α_2	β_1	β_2	β_3	β_4	β_5	β_6	δ	c|L	AIC	Model	$\bar{\beta}_L$	$\bar{\beta}_U$
rms713	0.181	0.926	0.556	−0.026	−0.003					4.198	0.7715		0.58	0.42
	(1.559)	(1.862)	(14.900)	(−3.107)	(−3.212)					111				
rms714	0.225	0.959	0.368	−0.018						4.493	0.7306	TM	0.41	0.31
	(2.312)	(2.336)	(11.526)	(−3.747)						115				
rms715	0.092		0.301						−0.287	4.493	0.8211	DBM	0.3	0.22
	(0.976)		(14.753)						(−2.987)	115				
rms716	0.047	3.379	0.453						−0.587	4.493	0.7792	DBM	0.45	0.28
	(0.586)	(3.176)	(19.001)						(−3.334)	115				
rms721	0.214		0.508						−0.144	2.272	0.6925	DBM	0.51	0.41
	(1.897)		(15.361)						(−2.282)	88				
rms722	−0.022	−0.950	0.626	0.034	−0.003	−0.000				3.225	0.7622	LSTM	0.54	0.67
	(−0.196)	(−2.205)	(10.201)	(2.787)	(−2.971)	(−3.358)				99				
rms725	0.144	1.194	0.528	−0.026	−0.002					4.464	0.8271		0.57	0.41
	(1.365)	(3.363)	(15.555)	(−4.376)	(−3.565)					114				
rms726	0.115		0.516						−0.121	−2.146	0.8539	DBM	0.45	0.4
	(1.300)		−18.741						(−2.668)	32				
rms727	0.097	0.806	0.412	−0.010						4.198	0.8353	TM	0.44	0.38
	(1.132)	(3.402)	(17.037)	(−2.716)						111				
rms728	0.593	−0.820	0.482	−0.009						−1.640	0.65	TM	0.5	0.45
	−3.090	(−2.678)	(11.883)	(−2.029)						40				
rms730	0.272	1.363	0.432	−0.026						4.493	0.6926	TM	0.5	0.34
	(2.0510)	(2.256)	(9.509)	(−3.880)						115				

around the threshold. None of the funds exhibit strong market timing ability. Four funds (rms320, rms321, rms324, and rms57) exhibit mild-form market timing ability in that $\hat{\delta}$ is significantly positive. Another two funds (rms56 and rms722) also exhibit weak-form market timing ability in that their average beta in the upper regime is larger than their average beta in the lower regime. In general, the results support a lack of market timing ability for both wholesale and retail funds.

Table 4.2 reports the results of the model chosen by the modified AIC for the 14 wholesale funds for which nonlinearity was identified by any of the five tests when the excess market return is used as a transition variable. The table reports parameter estimates and t statistics (calculated using White (1980) standard errors) for the null hypothesis that the coefficient equals zero in parentheses. In the tables c and L are the threshold and number of observations in the lower regime, respectively, while $\overline{\beta}_L$ and $\overline{\beta}_U$ are the average of the fitted values for beta over for the excess market return less than the median and larger than the median, respectively.

Table 4.3 reports the results of the model chosen by the modified AIC for the 22 retail funds for which nonlinearity was identified by any of the five tests using the excess market return as a transition variable. The table reports parameter estimates and t statistics (calculated using White (1980) standard errors) for the null hypothesis that the coefficient equals zero in parentheses. In the tables c and L are the threshold and number of observations in the lower regime, respectively, while $\overline{\beta}_L$ and $\overline{\beta}_U$ are the average of the fitted values for beta over for the excess market return less than the median and larger than the median, respectively.

CONCLUSION

This chapter has explored the market timing ability of Australian superannuation funds. In the exploration, the chapter has generalized the Henriksson and Merton (1981) and Merton (1981) dual beta model to a definition that allows a tiered approach (strong, mild, weak) to market timing ability and developed it using a class of nonlinear models that captures the Treynor and Mazuy (1966) quadratic market model, the Henriksson and Merton (1981) and Merton (1981) dual beta model, and the logistic smooth transition model of Teräsvirta and Anderson (1992), and Granger and Teräsvirta (1993) as special cases. In this context our analysis of Australian wholesale and retail superannuation funds finds that the simpler models of Treynor and Mazuy (1966) and Henriksson and Merton (1981) and Merton (1981) are more appropriate than the more complex nonlinear models of market timing ability, suggesting our findings are not sensitive to the model risk of extending these simpler models to a set of general nonlinear alternatives. This is an important finding given that these two models constitute the workhorses for testing market timing ability in the literature. Finally, our empirical results show a general absence of market timing ability consistent with the previous literature.

REFERENCES

Benson, K. and Faff, R. (2003) A Performance Analysis of Australian International Equity Trusts. *Journal of International Financial Markets Institutions and Money*, 13(1): 69–84.

Chou, P., Chung, H., and Sun, E. (2005) Detecting Mutual Fund Timing Ability Using the Threshold Model. *Applied Economics Letters*, 12(13): 829–834.

Do, V., Faff, R., and Wickramanayake, J. (2005) An Empirical Analysis of Hedge Fund Performance: The Case of the Australian Hedge Funds Industry. *Journal of Multinational Financial Management*, 15(4–5): 377–393.

Fama, E. and French K. (1992) The Cross-section of Expected Stock Returns. *Journal of Finance*, 47(2): 427–465.

Ferson, W. and Schadt, R. (1996) Measuring Fund Strategy and Performance in Changing Economic Times. *Journal of Finance*, 51(2): 425–462.

Fox, K. (2000) Information Rich Expressions for Model Selection Criteria. *Applied Economics Letters*, 7(1): 59–62.

Gallagher, D. (2001) Attribution of Investment Performance: An Analysis of Australian Pooled Superannuation Funds. *Accounting and Finance*, 41(1–2): 41–62.

Gallagher, D. and Jarnecic, E. (2004) International Equity Funds, Performance and Investor Flows: Australian Evidence. *Journal of Multinational Financial Management*, 14(1): 81–95.

Granger, C. and Teräsvirta, T. (1993) *Modelling Nonlinear Economic Relationships*. New York: Oxford University Press.

Hallahan, T. and Faff, R. (1999) An Examination of Australian Equity Trusts for Selectivity and Market Timing. *Journal of Multinational Financial Management*, 9(3–4): 387–402.

Henriksson, R. and Merton, R. (1981) On Market Timing and Investment Performance: Statistical Procedures for Evaluating Forecasting Skills. *Journal of Business*, 54(4): 513–533.

Holmes, K. and Faff, R. (2004) Stability, Asymmetry and Seasonality of Fund Performance: An Analysis of Australian Multi-sector Managed Funds. *Journal of Business Finance and Accounting*, 31(3–4): 539–578.

Luukkonen, R., Saikkonen P., and Teräsvirta T. (1988) Testing Linearity Against Smooth Transition Autoregressive Models. *Biometrika*, 75(3): 491–499.

Merton, R. (1981) On Market Timing and Investment Performance: An Equilibrium Theory of Value for Market Forecasts. *Journal of Business*, 54(3): 363–406.

Petruccelli, J. (1990) A Comparison of Tests for SETAR-type Nonlinearity in Time-series. *Journal of Forecasting*, 9(1): 25–36.

Pettengill, G., Sundaram, S., and Mathur, I. (1995) The Conditional Relation between Beta and Return. *Journal of Financial and Quantitative Analysis*, 30(1): 101–116.

Prather, L., Middleton, K., and Cusack, A. (2001) Are N+1 Heads Better than One? The Timing and Selectivity of Australian Managed Investment Funds. *Pacific Basin Finance Journal*, 9(4): 379–400.

Sawicki, J. and Ong, F. (2000) Evaluating Managed Fund Performance Using Conditional Measures: Australian Evidence. *Pacific Basin Finance Journal*, 8(3–4): 505–528.

Sinclair, N. (1990) Market Timing Ability of Pooled Superannuation Funds January 1981 to December 1987. *Accounting and Finance*, 30(1): 51–65.

Teräsvirta, T. and Anderson, H. (1992) Characterizing Nonlinearities in Business Cycles Using Smooth Transition Autoregressive Models. *Journal of Applied Econometrics*, 7(S): S119–S136.

Treynor, J. and Mazuy, K. (1966) Can Mutual Funds Outguess the Market? *Harvard Business Review*, 44(July–August): 131–136.

Tsay, R.S. (1989) Testing and Modelling Threshold Autoregressive Processes. *Journal of the American Statistical Association*, 84(405): 231–240.

White, H. (1980) A Heteroscedasticity-consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity. *Econometrica*, 48(4): 817–838.

This page intentionally left blank

CHAPTER 5

MODEL RISK

Caring about Stylized Features of Asset Returns—How Does the Equity Market Influence the Credit Default Swap Market?

Hayette Gatfaoui[1]

ABSTRACT

Recent debacles have raised the alarm that risk modeling is the main challenge for the financial community in the coming years. Model risk arises either from employing inappropriate models to describe a given economic dynamic or violating modeling assumptions. In this view, model risk concerns pricing models as well as quantitative risk assessment tools. Indeed, the violation of the classic Gaussian assumption as well as the negligence or underestimation of the well-known correlation risk (i.e., leverage effect in credit derivatives market) strongly impairs valuation and risk assessment processes. Under such a setting, we investigate the distributional properties of asset returns, and then propose a sound and simple measure of the corresponding correlation risk. Nowadays, such concerns are of huge significance in the era of increased asset comovements and asymmetric reactions to financial and economic shocks (e.g., subprime crisis). We focus on the link prevailing between credit default swap spreads (as a credit risk proxy) and the U.S. financial market (as a market risk proxy) in a world free of distributional assumptions. The relationship between CDX spreads and Dow Jones Composite Average index return is investigated with the flexible least squares regression method. We care about bad scenarios where a decrease in the U.S. market index triggers an increase in CDX

spreads. Namely, we focus on the downside risk so that the evolution of the Dow Jones Composite Average index returns impairs CDX spreads (i.e., widening of credit spreads).

INTRODUCTION

Recent debacles (e.g., LTCM hedge fund default in 1998, Amaranth collapse in 2006, and the subprime crisis since summer 2007) have raised the alarm that proper risk modeling is the main challenge for the financial community in the coming years. Model risk arises from an estimation error embedded in the assessment process of a specific economic phenomenon. Basically, model risk can be considered as a component of operational risk because it represents the potential failure in assessing the phenomenon under consideration. Originally, such assessment risk arises from the potential violations of the assumptions underlying the model that is in use (Derman, 1996). For example, many valuation models are founded on basic assumptions such as Gaussian returns, market liquidity, and preference-free framework so that we can assume a risk-neutral world. However, current studies have shown the frequent violations of such assumptions. Another source of error comes from neglecting the correlation risk between either financial assets (Barberis, Shleifer, and Wurgler, 2005; Dungey et al., 2006; Fender and Kiff, 2004) or market segments, namely the link prevailing between various risk sources (i.e., violation of some independency assumptions). For example, some authors show the skewed and fat-tailed nature of asset returns (Black, 2006; Taleb, 2007), whereas behavioral finance focuses on the importance of investor preferences in asset price determination (Amromin and Sharpe, 2005). On the other hand, Abid and Naifar (2006), Collin-Dufresne, Goldstein, and Martin (2001), Ericsson, Jacobs, and Oviedo (2004), Gatfaoui (2005, 2008), Gordy (2000), and Merton (1974), among others, showed the interaction between credit markets and equity markets (i.e., systematic factors). Consequently, applying classic models, which rely on basic violated assumptions, generates valuation biases known as model risk. Hence, model risk results from a mistaken valuation process because either the model is inappropriate[2] (i.e., erroneous dynamic) or assumptions are violated (i.e., nonrealistic world representation). Such issues favored the emergence of back-testing techniques to check for the validity of models as well as related scenario analysis to check for the model output and behavior, depending on the future economic and/or financial state, which will materialize (Peters, Shevchenko, and Wüthrich, 2007). The model performance is usually assessed through the analysis of its corresponding error (e.g., pricing error and misspecification of parameters, incorrect risk measurement in terms of biased profit and loss profile, see Kato and Yoshiba, 2000; Hull and Suo, 2002; Frey and McNeil, 2003; Cont, 2006; Nalholm and Poulsen, 2006).

Most credit risk determinants in use consist of credit spreads, namely the difference between corporate yields and corresponding Treasury yields. Credit spreads represent a compensation for the credit risk borne by investors. Such credit risk indicators are highly correlated with credit default swap (CDS) spreads, which are mainly default risk fundamentals but also liquidity determinants (Longstaff, Mithal, and Neis, 2005; Zhu, 2006). Indeed, Blanco, Brennan, and Marsh (2005) studied the equivalence between CDS prices and credit spreads.[3] Moreover, the correlation between credit risk indicators and equity market determinants is widely documented (Merton, 1974; Ericsson, Jacobs, and Oviedo, 2004; Abid and Naifar, 2006; Gatfaoui, 2008).[4] For example, Merton (1974) supports the significance of equity volatility in explaining credit spread levels. Analogously, Ericsson, Jacobs, and Oviedo (2004) exhibited equity volatility as a key determinant of the CDS spreads describing senior debt. In the same line, Abid and Naifar (2006) studied the influence of equity volatility on CDS rates in terms of level and daily changes, whereas Gatfaoui (2008) quantified the impact of changes in both an equity market benchmark's returns and equity market's implied volatility on credit defaults swap spread changes. Basically, several credit-risky instruments bear a nonnegligible portion of market/systematic risk insofar as these assets are traded within the financial market. Since the equity market bears a huge part of systematic risk as well, credit and equity markets should be correlated to some large extent. Namely, a well-chosen equity market index should be a good representative of the systematic/market risk factor,[5] which should help explain observed credit and CDS spreads (Sharpe, 1963). Along with academic and empirical research, we investigate the link prevailing between CDS spreads and the Dow Jones Composite Average index (DJC) return on a daily basis. We study the response of CDS spreads to moves in stock market benchmark's returns without specific distributional assumptions (i.e., no a priori probability setting). Namely, we focus on a bad scenario where CDS spreads increase (i.e., an increase in credit risk level, a worsening of credit market conditions) when DJC return decreases (i.e., a degradation of financial market conditions, an increase in systematic risk level). Specifically, downside risk is focused on through a negative link between CDS spreads and DJC returns over time.

DATA

We introduce the data and related stylized facts over the studied time horizon. Rather than focusing on recent history, a general prevailing link is targeted. Such a link is all the more important during stable time horizons because it strengthens during disturbed times (Fisher, 1959).[6]

Description

Daily data run from September 20, 2005, to August 14, 2006, for a total of 225 observations per series. We first consider the return of DJC expressed in basis points (R_DJC) as a proxy of market/systematic risk factor. The fact an index represents the market is not necessarily linked to the number of assets it encompasses (provided that the number of assets lies above the minimum admissible threshold). It depends rather on the way it is built, and two schools deal with this topic. The first school relies on the Markowitz (1952, 1959) diversification principle telling us that 30 assets suffice to build a diversified portfolio. The second school relies on statistical principles advocating at least 100 assets or a few hundreds of assets in a market benchmark, such as the S&P 500 index. But Campbell and colleagues (2001) have shown that the S&P 500 index still encompasses a nonnegligible part of idiosyncratic risk. Therefore, the informational content of selected data and a convenient definition of the relevant phenomenon should be targeted.[7]

Second, we consider a set of eight Dow Jones CDX indexes (DJCDX), which are CDS-type indexes tracking the CDS market as well as related liquidity side.[8] Specifically, DJCDX indexes are Dow Jones aggregate credit derivative indexes, which represent credit risk fundamentals. The first six indexes under consideration are DJCDX North America credit derivative indexes. They refer to entities (i.e., issuers) domiciled in North America and distributed among five sectors. We label them NA_IG, NA_IG_HVOL, NA_HY, NA_HY_BB, NA_HY_B, and NA_XO representing investment grade, investment-grade high volatility, high yield, BB-rated high yield, B-rated high yield, and crossover DJCDX indexes, respectively. Investment-grade indexes consider good and higher credit quality reference obligations/credits (i.e., BBB- to AAA-rated credits with low default risk). High yield indexes consider speculative grade credits, distressed debt, as well as some weaker BBB-rated credits. Crossover index NA_XO expresses credit rating divergences between Standard & Poor's and Moody's rating agencies across BB/Ba–BBB/Baa rating classes. Finally, the last two indexes under consideration are DJCDX emerging markets credit derivative indexes. They refer to entities domiciled either in Latin America, Eastern Europe, Middle East, Africa, or Asia. We label EM and EM_DIV as the emerging markets and emerging markets diversified DJCDX indexes, respectively. The EM index is based on sovereign entities, whereas EM_DIV is founded on both sovereign and corporate entities.

Basically, DJCDX credit derivative indexes are equal-weighted indexes, except the EM index whose weights depend on the decisions of CDS IndexCo LLC. Moreover, CDX indexes are reviewed regularly (i.e., issuers' selection and corresponding reference obligations) and updated on a semi-annual basis. Finally, we consider the spreads of DJCDX indexes against appropriate LIBOR rates (see www.markit.com for more details about the aggregation and computation/update process of indexes).[9] Those CDX spreads are expressed in basis points.

Properties

As regards time series properties, DJC return and DJCDX spreads are asymmetric and fat tailed (Table 5.1). Apart from the EM_DIV index, DJCDX spreads have negative excess kurtosis. Moreover, the NA_HY index exhibits the highest average DJCDX spread, whereas the NA_IG index exhibits the lowest one (i.e., lowest credit risk level). An unreported Phillips-Perron test[10] showed a stationary DJC return and first order integrated DJCDX spreads (i.e., stationary daily changes).

Unreported Kendall and Spearman correlations between DJCDX spreads and DJC return yielded mitigated results with regard to their respective sign. Obtained estimates are insignificant at a 5 percent bilateral test level (i.e., two-sided Student test).[11] Specifically, only the EM, EM_DIV, and NA_HY DJCDX indexes exhibit negative correlation coefficients while the other DJCDX indexes exhibit positive correlation coefficients. Moreover, Kendall and Spearman statistics range from −0.0625 and −0.0941 for the EM index (i.e., minimum observed values) to 0.0296 and 0.0454 for the NA_HY_BB index (i.e., maximum values).

Consequently, data are generally far from exhibiting a Gaussian behavior as supported by the goodness-of-fit test in Table 5.2,[12] and classic statistical tools do not allow for detecting any joint link between our chosen market risk proxy and credit risk fundamentals.

QUANTITATIVE ANALYSIS

Investigating the link between DJCDX spreads and DJC returns, we address the following question: How does market risk impair credit risk? We focus specifically on the negative impact of the financial market on

Table 5.1 Descriptive Statistics for CDX Spreads and DJC Return

Index	Mean	Median	Standard Deviation	Skewness	Excess Kurtosis*	Minimum (Maximum)
EM	154.0639	154.7300	27.8717	−0.1584	−0.4742	90.8800 (221.1200)
EM_DIV	102.9907	102.1300	15.4991	0.7123	0.7566	71.9600 (152.2600)
NA_HY_BB	237.0725	240.4400	27.4675	−0.3275	−0.6396	176.5500 (292.9100)
NA_HY_B	313.9764	310.6200	26.9951	0.0117	−0.3855	250.9800 (382.7800)
NA_HY	348.1687	344.1000	32.8510	0.4284	−0.1431	284.0900 (449.4400)
NA_IG_HVOL	88.5607	89.1900	10.7279	−0.1817	−0.9118	64.3900 (108.9600)
NA_IG	43.7473	44.2200	3.7758	−0.2183	−0.5853	34.0000 (51.1700)
NA_XO	204.2981	211.5600	24.3584	−0.3938	−0.7453	147.6400 (257.5000)
R_DJC	3.5220	6.5810	80.0253	0.1164	0.4459	−214.4586 (245.6136)

* The excess kurtosis is simply the kurtosis coefficient minus 3 (i.e., deviation from the benchmark Gaussian kurtosis). It assists in identifying distribution tails relative to the Gaussian probability law. A time series with a positive excess kurtosis (i.e., peaked, leptokurtic distribution) exhibits fatter distribution tails than the Gaussian law. The probability that extreme values may occur is then higher than the probability of extreme values for the Gaussian distribution. In the reverse case (i.e., flat, platykurtic distribution), the time series exhibits thinner distribution tails than the Gaussian probability law.

Table 5.2 Goodness-of-Fit Test for a Gaussian Distribution

Index	Anderson-Darling*	Estimated Mean	Estimated Standard Deviation
EM	0.8052	154.0639	27.8717
EM_DIV	1.9641	102.9907	15.4991
NA_HY_BB	1.4296	237.0725	27.4675
NA_HY_B	1.1337	313.9764	26.9951
NA_HY	1.5667	348.1687	32.8510
NA_IG_HVOL	2.3511	88.5607	10.7279
NA_IG	1.5331	43.7473	3.7758
NA_XO	4.1120	204.2981	24.3584
R_DJC†	0.5021	3.5220‡	80.0253

* Adjusted Anderson-Darling statistic accounting for finite sample and parameter uncertainty.

† The Gaussian probability distribution assumption is validated.

‡ Nonsignificant at a 5% test level, all other parameters being significant.

corporate credit market, namely the downside risk (Gatfaoui, 2005). For this purpose, we run flexible least squares (FLS) regressions of observed DJCDX spreads on observed DJC returns over the time horizon under consideration (i.e., investigating a dynamic linear link). Such a method is powerful since no probability assumption is required for the data.

Econometric Study

The FLS regression method was formerly introduced by Kalaba and Tesfatsion (1988, 1989, 1990). Such econometric method allows for running regressions with time-varying parameters, capturing some instantaneous link between random variables with robustness in terms of sample size. The efficiency of the method is such that systematic moves of regression coefficients such as unanticipated regime shifts (e.g., structural breaks or jumps at dispersed time points) are handled. Moreover, this methodology is also robust to outliers, nonstationary data as well as correlated data, among others. Such a feature is quite convenient given that market returns are stationary, whereas credit spread returns are not stationary. Finally, FLS setting requires no distributional assumptions or properties about the data under consideration, except the following implicit assumptions. Indeed, applying FLS linear regression principle assumes that both a linear link exists locally between the variables under consideration, and regression coefficients evolve slowly over time. The local linear link is supported by considering data, which are expressed in the same units and exhibit the same order of magnitude.[13] Consequently, we apply FLS method to run regressions of a given DJCDX spread S on DJC return

$$S_t = a_t + b_t \times \mathrm{R\_DJC}_t + v_t \tag{5.1}$$

where time t ranges from 1 to 225, a_t and b_t are time-varying trend and slope regression coefficients, and v_t is a residual measurement error. Coefficient a_t represents the DJCDX spread component that is unexplained by DJC return (i.e., idiosyncratic/unsystematic trend over time), whereas b_t coefficient catches the dynamic link between DJCDX spread and DJC return (i.e., instantaneous correlation risk indicator).[14] Therefore, we expect the trend coefficient to be positive, whereas the slope coefficient can be either positive or negative. Such slope coefficient captures the asymmetric response of DJCDX spread moves to shocks on DJC returns. Indeed, it quantifies both how DJCDX spreads change in directional terms, and the extent to which these spreads change subsequent to a change in DJC return (i.e., magnitude of CDS spread moves). In particular, FLS method helps account for two market influences. First, the linear link between DJCDX spread and DJC return (i.e., equity volatility impacts credit spreads) catches market price risk. Second, market volatility risk is handled through the time variation of the slope coefficient (i.e., the magnitude of b_t over time reflects the impact and significance of market volatility risk).

Given optimal cost parameters c_1 and c_2, the following objective function F is minimized:

$$F(a_t, b_t, t = 1 \cdots 225) = \sum_{t=1}^{225} v_t^2 + c_1 \sum_{t=2}^{225} \left(a_t - a_{t-1}\right)^2 + c_2 \sum_{t=2}^{225} \left(b_t - b_{t-1}\right)^2 \qquad (5.2)$$

Minimizing function F yields a goodness-of-fit criterion since function F represents the estimation costs of relation (Equation (5.1)). The first term of F is simply the sum of squared regression residuals, whereas the remaining terms represent the weighted sums of squared dynamic specification errors.[15] Specifically, the weights represent the cost parameters, which account for FLS coefficient variations. The lower the cost parameters, the more volatile are the time-paths of related regression coefficients.[16] Conversely, the higher cost parameters are, the smoother (i.e., more regular and stable) corresponding coefficient time-paths. The set of possible FLS solutions is called residual efficiency frontier by Kalaba and Tesfatsion (1988). Such efficiency frontier represents the pairs of sums of squared regression residuals and dynamic specification errors, which satisfy the quadratic minimization in Equation (5.2), conditional on the observed data.

Discussing Optimal FLS Choice

The optimization program solving for the cost function's minimization yields FLS estimates, which belong to the efficiency frontier. Any FLS choice on the efficiency frontier is arbitrary (Kalaba and Tesfatsion, 1996).[17] However, Kalaba and Tesfatsion (1989) underline that "residual measurement errors and residual dynamic errors are anticipated to be symmetrically distributed around zero." Therefore, in line with the implicit assumption about generally smooth time-varying coefficients and previous statement, we condition the optimal criterion on the quality of regression residuals. Namely, the best FLS estimates yield regression residuals, which are approximately symmetrically distributed around zero so that they evolve around this threshold in a stable way with the lowest standard deviation. The approximate distribution criterion we apply is so that we allow for a deviation margin of 10 percent around the perfect 50 percent level. Basically, we allow for observing at least 40 percent (or at most 60 percent) of estimated regression residuals to be above zero, and the other 60 percent (or 40 percent) to be below zero.

Results

Optimal cost parameters are 0.001 apart from the EM, EM_DIV, and NA_HY indexes for which c_1 is 0.10, and NA_HY_B for which c_1 is 10. As a robustness check, Table 5.3 displays statistical information about regression residuals. Observed standard deviations and percentage statistics show the stable evolution as well as the closeness of residual levels to the zero threshold. Moreover, residuals exhibit an approximate balanced dispersion around zero.[18] Regression trends a_t are stable (Figure 5.1) over time, whereas b_t slope coefficients are highly volatile over time (Figure 5.2). As a result, DJCDX spreads exhibit a stable default component (i.e., a stable unsystematic/idiosyncratic component), whereas they exhibit an extremely volatile market-based component (i.e., volatile systematic/market component). Moreover, the descriptive statistics displayed in Table 5.4 advocate the nonconstancy assumption about regression parameters. Indeed, the standard deviation levels are far from being zero, which confirms Kalaba and Tesfatsion's (1989) statement.[19] As an extra investigation, unreported results confirmed the stationary and white noise patterns in regression residuals while estimating simple and partial autocorrelations as well as Phillips-Perron and Ljung-Box statistics.[20]

With regard to Figure 5.1, the NA_IG index exhibits the lowest unsystematic component over time, whereas NA_HY exhibits the highest one. Moreover, the idiosyncratic CDX spread components tend generally to decrease until the end of the first quarter 2006 and start increasing during the second quarter 2006 (i.e., higher default risk level). A structural break seems then to arise from a reversal in the previous decreasing trend of most CDX spreads. Generally speaking, the default risk level (i.e., idiosyncratic

Table 5.3 Statistics About Regression Residuals

Index	Median	Standard Deviation	Prop% < 0	Prop% > 0	|Prop%| < 0.0001*
EM	−4.2967E–07	4.16E–06	50.6667	49.333	88.8889
EM_DIV	5.4555E–08	1.0086E–06	49.7778	50.2222	93.7778
NA_HY_BB	4.4582E–07	9.6578E–08	44.8889	55.1111	95.5556
NA_HY_B	2.6733E–06	4.6557E–06	41.3333	58.6667	83.5556
NA_HY	1.7917E–06	5.0681E–06	43.5556	56.4444	85.3333
NA_IG_HVOL	5.6814E–07	1.1697E–08	40.4444	59.5556	96.8889
NA_IG	1.1589E–07	2.739E–09	40.8889	59.1111	98.2222
NA_XO	6.2802E–07	6.4137E–08	41.7778	58.2222	94.2222

* Proportion of residual observations lying below this threshold in absolute value.

Figure 5.1 Flexible Least Squares Regression Trend Coefficient

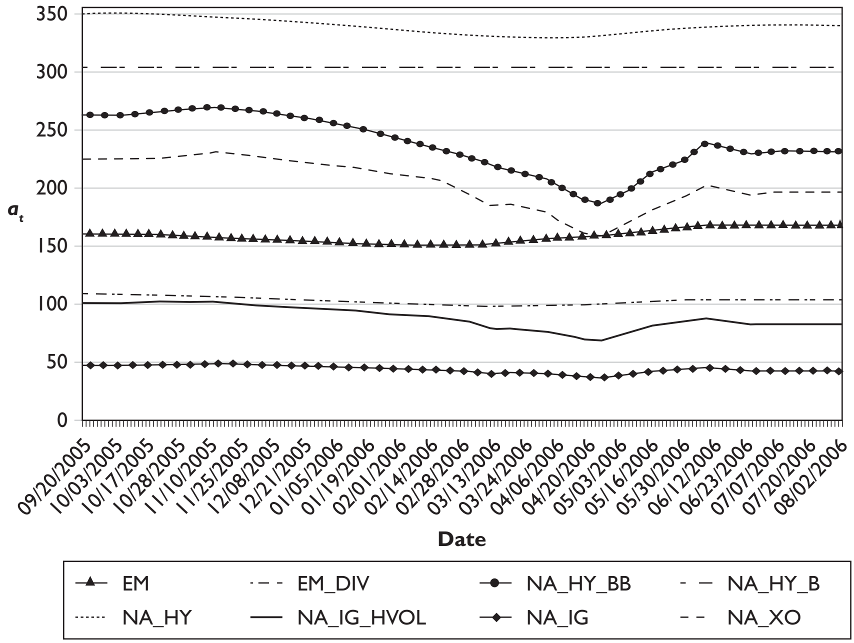


Figure 5.2 Flexible Least Squares Regression Slope Coefficient

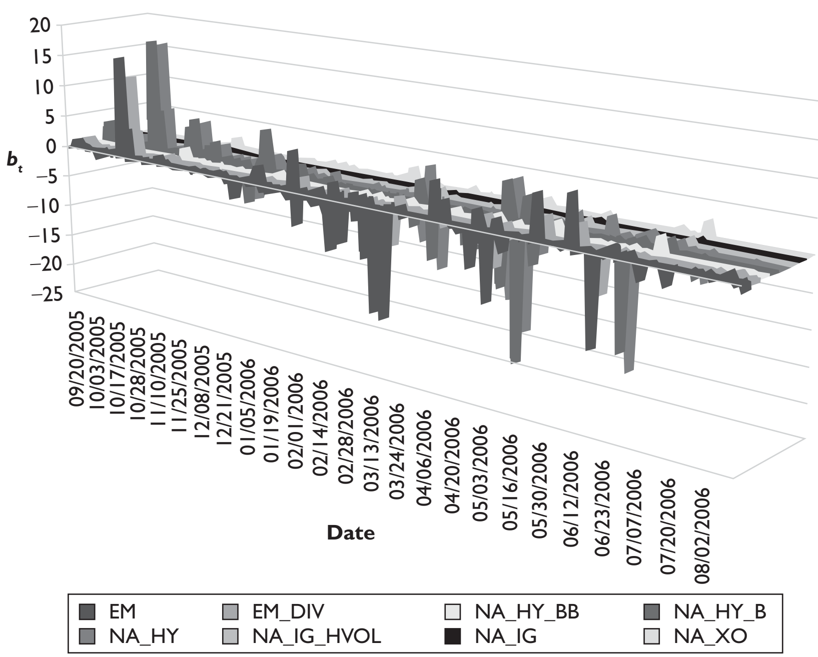


Table 5.4 Statistics About Time-varying Regression Coefficients

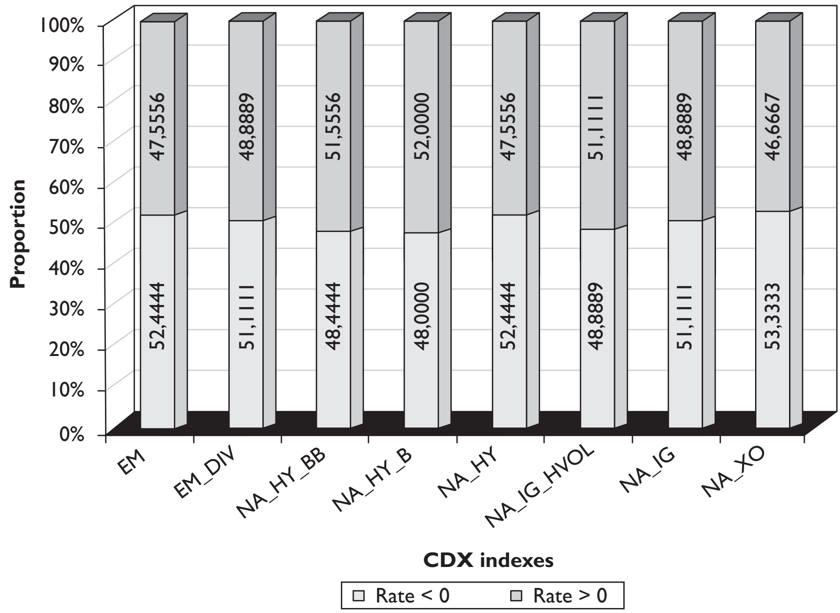
	Trend a_t			Slope b_t		
Index	**Mean**	**Median**	**Standard Deviation**	**Mean**	**Median**	**Standard Deviation**
EM	158.5191	157.9222	36.4352	−0.2333	−0.0411	7.0738
EM_DIV	102.8709	103.2003	9.7011	−0.0597	−0.0013	1.7182
NA_HY_BB	237.2983	234.3079	573.2639	−0.0117	0.0021	0.2315
NA_HY_B	303.8953	303.8389	0.0093	−0.1039	0.0107	7.2588
NA_HY	339.3182	339.6218	45.4022	−0.1846	−0.0211	7.0929
NA_IG_HVOL	88.3319	86.7457	97.3956	0.0058	0.0007	0.0407
NA_IG	43.7308	43.6557	10.5388	0.0021	−0.0005	0.0073
NA_XO	203.0801	200.7260	423.9249	−0.0039	−0.0034	0.2116

CDX spread component) is lower at the end of the studied time horizon than at the beginning. The previous stylized facts support a strong relationship between credit risk fundamentals and equity volatility (i.e., equity market). Specifically, the low, realized volatility level supports persistent and tight credit spreads over our time horizon. Possible explanations result from the firms' financial stability and propitious financing environment coupled with a low correlation risk within the stock market (Beil and Rapoport, 2008).

With regard to Figure 5.2, the prevailing link between CDX spread levels (i.e., credit risk fundamental) and DJC return level (i.e., market risk proxy) illustrates the dynamic dependence of credit risk relative to market risk. Hence, the slope coefficient b_t represents a dynamic correlation risk measure. Figure 5.2 exhibits a highly fluctuating and sign-varying instantaneous correlation risk between DJCDX spreads and DJC return. Such a stylized fact probably explains the failure of classic statistical tools to capture the (instantaneous) correlation risk prevailing between credit markets and equity markets. Targeting the joint credit risk and market risk evolution, we focus on the sign of b_t regression coefficients. Specifically, a negative b_t coefficient means that DJCDX spread increases when DJC return decreases, and the converse.[21] In the worst case, credit risk increases (through CDS spread widening) when market risk increases (through DJC return tightening). To get a view, we compute the proportion of positive and negative values of b_t coefficients over our studied time horizon for each DJCDX index (Figure 5.3).

The proportions of observed negative b_t slope coefficients lie generally above the proportions of observed positive b_t slope coefficients over the time horizon under consideration, except for the NA_HY_BB, NA_HY_B, and NA_IG_HVOL indexes. Such a stylized fact tends to

Figure 5.3 Proportions of Positive and Negative b_t Coefficient Values



support an average negative relationship between credit risk fundamentals and market risk proxy.

Unreported computations handled Spearman correlation coefficients between the first order differences of DJCDX spreads[22] (i.e., daily variations $\Delta S_t = S_t - S_{t-1}$) and the first order differences of DJC return (i.e., daily changes ΔR_DJC_t). All correlation coefficients are negative and significant at a 1 percent bilateral test level. Therefore, DJCDX spreads and DJC return tend to evolve in an opposite way over the time horizon under consideration. Moreover, previous correlation coefficients range from −0.3584 for the EM index to −0.1541 for the NA_HY_BB index. As an extension, the joint evolution of first order differences of both DJCDX spreads and DJC return was also considered (Table 5.5). Table 5.5 considers the respective signs of the first order differences of DJCDX spreads and DJC return, and summarizes the cases where those signs are identical and reverse, as compared with the total number of observed cases (i.e., a total of 224 observations for first order differences time series). The proportion of simultaneous reverse changes in both ΔS_t and ΔR_DJC_t is far above the proportion of simultaneously correlated changes in both ΔS_t and ΔR_DJC_t. The lowest and highest rate of correlated joint daily variation is 33.9286 percent and 45.5357 percent for the EM and NA_HY_BB indexes, respectively, whereas

Table 5.5 Proportions for Joint Changes in CDX Spreads and DJC Return

	Correlated behavior			Reverse behavior		
Percentage*	**CDX spreads increase and DJCI return increases**	**CDX spreads decrease and DJCI return decreases**	**Sum**	**CDX spreads increase and DJCI return decreases**	**CDX spreads decrease and DJCI return increases**	**Sum**
EM	12.5000	21.4286	33.9286	31.2500	34.3750	65.6250
EM_DIV	16.5179	25.0000	41.5179	272321	308036	58.0357
NA_HY_BB	18.7500	26.7857	45.5357	25.4464	28.1250	53.5714
NA_HY_B	15.6250	28.1250	43.7500	24.1071	31.6964	55.8036
NA_HY	15.6250	25.8929	41.5179	26.3393	31.6964	58.0357
NA_IG_HVOL	18.3036	24.1071	42.4107	28.1250	29.0179	57.1429
NA_IG	15.6250	25.0000	40.6250	27.2321	31.2500	58.4821
NA_XO	16.9643	24.5536	41.5179	28.1250	29.9107	58.0357

* All probabilities do not sum to 100% since CDX spreads remain stable in less than 1% of cases.

the respective lowest and highest rate of converse joint daily variation is 53.5714 percent and 65.6250 percent for the NA_HY_BB and EM indexes, respectively. Such feature confirms the general worst case-joint trend for credit and market risks over the studied time horizon. Therefore, market risk tends to impair credit risk over the time horizon under consideration.

In conclusion, it is possible to decompose CDS spreads into a pure default (i.e., nonsystematic) component and a systematic component, which is linked to the equity market. The default CDS spread component evolves in a stable way over the studied time horizon but exhibits a reversal during the first half of 2006. In an opposite way, the systematic CDS spread component is highly fluctuating. Moreover, the instantaneous correlation risk, as represented by the slope coefficient, is also highly volatile and exhibits frequent sign changes. Such features suggest to model credit spreads while accounting for a stable time-varying trend, which undergoes deviations subsequent to shocks arising from moves in the equity market. Hence, the correlation between the credit market and the equity market needs to be taken into account in a dynamic and volatile way. Finally, equity market moves tend generally to impact the credit market in a negative way. The negative equity market influence dominates while driving the credit market. The previous results are significant for credit risk and market risk management prospects such as value-at-risk implementations. Related implications are twofold. First, the impact of equity markets has to be taken into account. Second, the correlation between equity and credit markets needs to be soundly measured under a time-varying setting. Failing to account for those two features will engender estimation biases while assessing credit risk under the Basel II landscape. Indeed, the Basel II setting (Basel Committee on Banking Supervision, 2006, 2009a,b,c) sheds light on the main standards for sound risk management practices (i.e., regulatory capital framework). In particular, three cornerstones called pillars I, II, and III describe such a management process. Pillar I focuses on the minimum capital requirements implied by the different risk exposures. In this view, sound risk measurement practices are needed so as to assess fairly and accurately the risks under consideration as well as the related potential economic expected losses. Moreover, the potential correlation between asset classes and within asset classes needs to be seriously considered in both risk mitigation and fair risk assessment prospects. For this purpose, model risk and model validation control processes have to be undertaken (e.g., back-testing). Pillar II strengthens such needs since it focuses on the supervisory review process. Under this pillar, banks and financial institutions among others apply their homemade internal processes to assess their specific capital requirements and to monitor their respective capital adequacy. Basically, the economic capital modeling and corresponding expected losses rely on stress-testing methods (i.e., computing expected losses under various future market or portfolio scenarios[23]). At this stage, the robustness property and the stability of the selected model are very important (e.g., model stability, efficiency, and sensitivity to various market moves or extreme scenarios). Finally, the risk assessment results inferred from the settings in pillars I and II are displayed along with pillar III, which requires regular information disclosure about capital adequacy and risk level assessment (i.e., transparency rules).

CONCLUSION

Under the Basel II setting, model risk concern has gained a significant place among the risk management community. Such issue requires using appropriate quantitative tools to soundly assess risk. Such tool appropriateness depends strongly on the statistical profile exhibited by asset prices over time. Therefore, the statistical profile has to be drawn before running any analysis in order to emphasize the observed real world and to drive the relevance of the chosen model. Indeed, the stylized features and properties of price and return time series often invalidate the use of classic risk measurement tools. Such an issue applies also to the relationship between CDS spreads and stock market benchmark returns, which are far from being Gaussian.

Resorting to FLS regression method, we exhibited the dynamic link prevailing between DJCDX spreads and DJC return. Indeed, the time-varying trend coefficient illustrates idiosyncratic DJCDX spread components, whereas the time-varying slope coefficient reflects the instantaneous correlation risk between credit risk and market risk over time (i.e., market volatility risk). We also quantified the joint evolution of credit risk and market risk over our studied time horizon. First, we found stable positive unsystematic/idiosyncratic DJCDX spread components over time. Second, the link between DJCDX spreads and DJC return was extremely volatile and exhibited frequent sign changes over time (i.e., unstable correlation risk). Therefore, the dependence structure between credit risk (i.e., DJCDX spreads) and market risk (i.e., DJC return) is proved to be time-varying and highly volatile. Finally, we found a general negative link between credit risk and market risk over the time horizon under consideration (i.e., aggregate static view). Further extension should however study such a dependence structure in a two-dimension setting (i.e., more accurate bivariate setting assessing the simultaneous correlation risk) as well as in a nonlinear framework (e.g., nonlinear Kalman filter methodology). Such an assessment task could easily be handled, for example, in a multivariate distribution setting or a copula-based modeling framework (Cherubini, Luciano, and Vecchiato, 2004; Gatfaoui, 2007).

Finally, the noticeable reversal in the general trend of DJCDX spreads during the first half of 2006 may be considered ex post as a signal about the future heat of credit markets. Consequently, the FLS approach may assist in building useful advanced indicators of credit conditions and credit market trend. However, such a useful task requires further investigation.

REFERENCES

Abid, F. and Naifar, N. (2006) The Determinants of Credit-Default Swap Rates: An Explanatory Study. *International Journal of Theoretical and Applied Finance*, 9(1): 23–42.

Amromin, G. and Sharpe, S.A. (2005) From the Horse's Mouth: Gauging Conditional Expected Stock Returns from Investor Survey. Finance and Economics Discussion Series, Working paper 2005–26. Washington, DC: Federal Reserve Board, Divisions of Research & Statistics and Monetary Affairs.

Anderson, T.W. and Darling, D.A. (1954) A Test of Goodness of Fit. *Journal of the American Statistical Association*, 49(268): 765–769.

Barberis, N., Shleifer, A., and Wurgler, J. (2005) Comovement. *Journal of Financial Economics*, 75(2): 283–318.

Basel Committee on Banking Supervision. (2006) International Convergence of Capital Measurement and Capital Standards: A Revised Framework, Comprehensive Version. Basel, Switzerland: Bank for International Settlements.

Basel Committee on Banking Supervision. (2009a) Guidelines for Computing Capital for Incremental Risk in the Trading Book. Consultative document. Basel, Switzerland: Bank for International Settlements.

Basel Committee on Banking Supervision. (2009b) Proposed Enhancements to the Basel II Framework. Consultative document. Basel, Switzerland: Bank for International Settlements.

Basel Committee on Banking Supervision. (2009c) Revisions to the Basel II Market Risk Framework. Consultative document. Basel, Switzerland: Bank for International Settlements.

Beil, D. J. and Rapoport, P. (2008) The Equity Volatility-Credit Link. *Risk*, April: 42–43.

Black, K.H. (2006) Improving the Hedge Fund Risk Exposures by Hedging Equity Market Volatility, or How the VIX Strategy Ate My Kurtosis. *Journal of Trading*, 1(2): 6–15.

Blanco, R., Brennan, S., and Marsh, I.W. (2005) An Empirical Analysis of the Dynamic Relation Between Investment-Grade Bonds and Credit Default Swaps. *Journal of Finance*, 60(5): 2255 –2281.

Campbell, J.Y., Lettau, M., Malkiel, B.G., and Xu, X. (2001) Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk. *Journal of Finance*, 56(1): 1–43.

Cherubini U., Luciano E., and Vecchiato, W. (2004) *Copula Methods in Finance*. Chichester, UK: John Wiley & Sons.

Collin-Dufresne P., Goldstein R.S., and Martin, J.S. (2001) The Determinants of Credit Spread Changes. *Journal of Finance*, 56(6): 2177–2208.

Cont, R. (2006) Model Uncertainty and Its Impact on the Pricing of Derivative Instruments. *Mathematical Finance*, 16(3): 519–547.

Das, S. (2009) Credit Default Swaps: Quantum Hedges—Parts I and II. Available at: http://www.eurointelligence.com/Article3.1018+M5719dcf0252.0.html and http://www.eurointelligence.com/Article3.1018+M5f85214502c.0.html. Accessed September 2009.

Derman, E. (1996) Model Risk. Quantitative Strategies Research Notes, Goldman Sachs.

Dungey, M., Fry, R., González-Hermosillo, B., and Martin, V. L. (2006) Contagion in International Bond Markets During the Russian and LTCM Crises. *Journal of Financial Stability*, 2(1): 1–27.

Ericsson, J. and Renault, O. (2006) Liquidity and Credit Risk. *Journal of Finance*, 61(5): 2219–2250.

Ericsson, J., Jacobs, K., and Oviedo, R. A. (2004) The Determinants of Credit Default Swap Premia. Working paper, CIRANO, Montreal, 2004s–55. *Journal of Financial and Quantitative Analysis* (forthcoming).

Fender, I. and Kiff, J. (2004) CDO Rating Methodology: Some Thoughts on Model Risk and Its Implications. BIS Working paper 163, Monetary and Economic Department.

Fisher, L. (1959) Determinants of Risk Premiums on Corporate Bonds. *Journal of Political Economy*, 67(3): 217–237.

Frey, R. and McNeil, A.J. (2003) Dependent Defaults in Models of Portfolio Credit Risk. *Journal of Risk*, 6(1): 59–92.

Gatfaoui, H. (2005) How Does Systematic Risk Impact U.S. Credit Spreads? A Copula Study. *Revue Banque & Marchés*, 77(July–August): 5–16.

Gatfaoui, H. (2007) Credit Default Swap Spreads and U.S. Financial Market: Investigating Some Dependence Structure. Conference on Copulae and Multivariate Probability Distributions in Finance—Theory, Applications, Opportunities and Problems, Warwick Business School.

Gatfaoui, H. (2008) Investigating the Link Between Credit Default Swap Spreads and U.S. Financial Market. In: Greg N. Gregoriou (ed.), *The Credit Derivatives Handbook: Global Perspectives, Innovations and Market Drivers* (Chapter 9, pp. 183–200). New York: McGraw-Hill.

Gordy, B. (2000) A Comparative Anatomy of Credit Risk Models. *Journal of Banking and Finance*, 24(1): 119–149.

Hull, J. and Suo, W. (2002) A Methodology for Assessing Model Risk and Its Application to the Implied Volatility Function Model. *Journal of Financial and Quantitative Analysis*, 37(2): 297–318.

Kalaba, R. and Tesfatsion, L. (1988) The Flexible Least Squares Approach to Time varying Linear Regression. *Journal of Economic Dynamics and Control*, 12(1): 43–48.

Kalaba, R. and Tesfatsion, L. (1989) Time-Varying Linear Regression via Flexible Least Squares. *Computers and Mathematics with Applications*, 17(8–9): 1215–1245.

Kalaba, R. and Tesfatsion, L. (1990) Flexible Least Squares for Approximately Linear Systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 20(5): 978–989.

Kalaba, R. and Tesfatsion, L. (1996) A Multicriteria Approach to Model Specification and Estimation. *Computational Statistics and Data Analysis*, 21(2): 193–214.

Kato, T. and Yoshiba, T. (2000) Model Risk and Its Control. *Monetary and Economic Studies*, 18(2): 129–157.

Longstaff, F.A., Mithal, S., and Neis, E. (2005) Corporate Yield Spreads: Default Risk or Liquidity? New Evidence from the Credit Default Swap Market. *Journal of Finance*, 60(5): 2213–2253.

Markowitz, H.W. (1952) Portfolio Selection. *Journal of Finance*, 7(1): 77–91.

Markowitz, H.W. (1959) *Portfolio Selection: Efficient Diversification of Investment*. New York: John Wiley.

Merton, R.C. (1974) On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *Journal of Finance*, 29(2): 449–470.

Nalholm, M. and Poulsen, R. (2006) Static Hedging and Model Risk for Barrier Options. *Journal of Futures Markets*, 26(5): 449–463.

Peters, G.W., Shevchenko, P.V., and Wüthrich, M.V. (2007) Model Risk in Claims Reserving Within Tweedie's Compound Poisson Models. Preprint CMIS 2609, CSIRO.

Phillips, P.C.B. and Perron, P. (1988) Testing for a Unit Root in Time Series Regression. *Biometrika*, 75(2): 335–346.

Sharpe, W.F. (1963) A Simplified Model for Portfolio Analysis. *Management Science*, 9(2): 499–510.

Taleb, N. (2007) *The Black Swan: The Impact of the Highly Improbable*. New York: Penguin.

Zhu, H. (2006) An Empirical Comparison of Credit Spreads Between the Bond Market and the Credit Default Swap Market. *Journal of Financial Services Research*, 29(3): 211–235.

NOTES

1. We would like to thank participants at the 4th AFE-QASS conference (July 2007, Samos, Greece) and the GdRE financial risks day (November 2007, Créteil, France), as well as Carsten Wehn for their interesting questions and comments. We apologize for any missing references on the topic. A summarized draft of the paper was formerly titled "Are credit default swap spreads market driven?" The usual disclaimer applies.
2. Model inadequacy refers to the adequacy with the targeted dynamic rather than the related model's degree of complexity. For example, estimating a specific dynamic over a past horizon of observed relevant data to assess a future risk exposure assumes implicitly that the past observed dynamic will prevail on the future time horizon under consideration. Such stability considerations need to be carefully handled in terms of risk features, and specifically when the risk position is sensitive to market and economic reversals (e.g., structural breaks). Of course,

the role of the considered horizon's length depends on the risk structure stability which is assumed (e.g., local risk exposure on a short-term horizon or global risk exposure over a medium or long-term horizon). A striking example comes from the asset returns? normality assumption. Such assumption is valid as long as returns are highly and homogeneously distributed around their related historical mean so that extreme return values exhibit a low probability of occurrence.

3. Equivalence is valid under specific assumptions so as to yield an equilibrium relationship between CDS prices and credit spreads (e.g., stable market conditions). During disturbed market times such as the subprime mortgage crisis of 2007–2008, the correlation between credit spreads and CDS prices drops since CDS prices incorporate additional information on corporate bond liquidity and credit spread volatility among others (Das, 2009). However, these two fundamentals encompass much common relevant information about credit risk (i.e., high correlation).
4. Studies usually consider regressions of credit risk indicators on market risk determinants among others. Hence, the impact of market/systematic factors on credit risk determinants is studied (Abid and Naifar, 2006; Collin-Dufresne, Goldstein, and Martin, 2001; Ericsson, Jacobs, and Oviedo, 2004; Gatfaoui, 2008).
5. At least, such an index should encompass relevant market information, be it on an implicit basis (e.g., latent information content).
6. We'll notice later in this chapter that previous history may encompass insights of future history.
7. Dow Jones Composite Average index encompasses at least implicitly relevant information about the global financial market's trend.
8. Credit risk becomes sometimes worse due to a degradation of liquidity conditions. For example, Ericsson and Renault (2006) exhibited the liquidity component in U.S. corporate credit spreads during the 1990s. They underlined the decreasing term structure of the liquidity component in credit spreads. Indeed, the shorter the credit spread maturity is, the higher the liquidity component in credit spreads becomes. Specifically, Longstaff et al. (2005) split the nondefault component in credit spreads into both a microeconomic component representative of bond-specific liquidity and a macroeconomic component representative of corporate market-based liquidity.
9. All data analyzed in this applied research paper were initially extracted from the Dow Jones Corporation's Web site. Since January 2007, data have been provided by Markit Corporation.
10. One percent test level according to Phillips and Perron (1988).

11. Such a feature may result from the frequency of data, which probably generates some noise without forgetting the one-lag dependency in DJCDX spread data. Moreover, the volatile behavior observed from the daily data renders the correlation analysis more complex.
12. An Anderson-Darling test at a 5 percent level yields generally the rejection of the normal probability distribution assumption (Anderson and Darling, 1954). The Gaussian distribution assumption is however validated for the Dow Jones Composite Average index return (R_DJC).
13. "In particular, the units in which the regressor variables are measured should be chosen so that the regressors are approximately of the same order of magnitude." Kalaba and Tesfatsion (1989).
14. The slope coefficient represents the link with the equity market, and generally with the broad financial market.
15. The case where the sums of squared dynamic specification errors are zero corresponds to the ordinary least squares setting. Indeed, such sums are zero when we require the regression coefficients to be constant.
16. Time-variation in regression parameters is weakly penalized, whereas it is strongly penalized in the converse case.
17. "Without additional prior information, restricting attention to any proper subset of the FLS estimate is an arbitrary decision." Kalaba and Tesfatsion (1996).
18. The level of regression residuals has to be balanced with the observed level of default swap spreads, which is globally measured in hundreds of basis points.
19. "The standard deviation of the FLS kth coefficient estimates about their average value provides a summary measure of the extent to which these estimates deviate from constancy." Kalaba and Tesfatsion (1989).
20. It is convenient to assume white noise residuals when the econometric relationship captures the most relevant features (i.e., economic soundness and statistical validity). In our case, such a pattern is more than welcome since our methodology accounts for time-variation in parameters as well as corresponding structural breaks. Indeed, even if markets are described by nonlinear dynamics, it is always powerful and true to assume a local time-updating linear dynamic. In the end, the dynamic is locally linear (i.e., over very short-term windows) but nonlinear over a larger time scale. Moreover, the time-varying coefficients also handle volatility changes and clustering to some extent.
21. A negative slope coefficient illustrates an opposite simultaneous evolution of DJCDX spreads and DJC return.

22. Recall that DJCDX spreads are based on credit default swap spreads, which are computed against corresponding LIBOR rates. Therefore, DJCDX spreads are relative returns per se.
23. Value-at-risk or expected shortfall techniques are often employed. However, the assumptions, the theoretical and practical limitations, and the economic as well as financial anomalies describing and surrounding the selected model require special attention. Previous features are key determinants of the model?s efficiency, robustness, and validity.

This page intentionally left blank

CHAPTER 6

PRICE TRANSMISSIONS AND MARKET RISK IN FINANCIAL MARKETS

Viviana Fernandez

ABSTRACT

Decomposing a time series into its high- and low-frequency components has been the object of study in various fields of knowledge since the nineteenth century. Such decomposition has gained ground in the finance field in recent years motivated by the existence of heterogeneous investment horizons. A mathematical tool developed in the early 1990s, denominated as wavelets, has become increasingly popular to characterize the short- and long-term behavior of financial indexes. In particular, in an article published in 2005, Connor and Rossiter provide an appealing interpretation of wavelet analysis by pointing out that long-term traders focus on price fundamentals which drive overall trends, while short-term traders react to incoming information within a short-term horizon. Such heterogeneity, in Connor and Rossiter's (2005) view, can be modeled by means of wavelet analysis.

In this chapter, we illustrate the usefulness of wavelets in gauging asset return co-movement, structural variance shifts, and the overall risk of a portfolio as measured by its value at risk. Specifically, we concentrate on the nine Standard & Poor's Depository Receipts sectors—i.e., consumer discretionary, consumer staples, energy, financial, health care, industrial, materials, technology, and utilities—along the period 1999 to 2007.

INTRODUCTION

The extant literature on price transmission has resorted to several statistical techniques to gauge spillovers and tail dependency. To date, the most popular ones are vector autoregressive regression (VAR) systems, multivariate generalized autoregressive conditional heteroscedasticity (GARCH) models in their various forms, heteroscedasticity-robust correlation coefficients, and extreme value theory. An alternative approach, which is relatively new to the finance field, is wavelet analysis. This is a refinement of Fourier analysis whose origin dates back to the late 1980s, and which offers a powerful methodology for the decomposition of a time series into orthogonal components at different frequencies. Such frequencies are associated with short- and long-term fluctuations, which are characterized by their time location. Put simply, wavelets make it possible to trace the evolution of a return series within a one-week period, for instance.

Wavelets have several advantages over traditional statistical tools commonly utilized in finance. First, given that they enable us to decompose a time series into its low- and high-frequency components, we can characterize a financial index into its trend and the deviations from it. A potential application of such characterization is cointegration analysis and the quantification of risk diversification. Second, wavelets make it possible to carry out a variance decomposition of a time series, which yields us information about the most important contributors to time-series variability. Such decomposition allows us to test for the presence of variance structural breaks, for instance. Similarly, a wavelet-based covariance can be computed for paired time series. The combination of variance and covariance decompositions provides us with multiple applications, such as the obtainment of wavelet-based betas and value at risk (VaR), which are dealt with in this chapter. Finally, wavelets can be also utilized in forecasting and business-cycle analysis, given their ability to distinguish among seasonal, cyclical, and trend components of a time series.

Wavelet analysis has gained more ground in the finance field from the mid–1990s onward. The main reason for such success is that wavelets represent an alternative way of looking at different sources of risk (i.e., on a scaled basis), which is key under the worldwide turbulence experienced by international financial markets in the past decade.

Recent contributions in this area have dealt with the relation between futures and spot prices, the estimation of systematic risk of an asset in the context of the domestic and international version of the capital asset pricing model (CAPM), heterogeneous trading in commodity markets, selection of an optimal hedge ratio for a commodities portfolio, structural breakpoints in volatility, volatility persistence, business-cycle decomposition, and wavelet-based VaR computation, among other themes. See, for instance, Lin and Stevenson (2001), Gençay, Whitcher, Selçuk (2003, 2005), Connor and Rossiter (2005), Karuppiah and Los (2005), In and Kim (2006), Kyaw, Los, and Zong (2006), Fernandez (2005, 2006a, 2006b, 2007, 2008), Fernandez and Lucey (2007), Lien and Shrestha (2007), and Yogo (2008). A thorough discussion on the use of wavelets in finance can be found in the textbook by Gençay, Selçuk, and Whitcher (2002).

In this chapter, we illustrate several applications of wavelets that are of interest to practitioners, such as computing time-varying correlations, asset betas, and a portfolio VaR, and testing for volatility shifts. A wavelet-based decomposition is mathematically complex to carry out, but nowadays there exist canned routines built into mathematical languages, which enormously facilitate its implementation.[1]

This chapter is organized as follows. The second section of this chapter briefly introduces the concept of wavelets and various applications. The third is devoted to an empirical application illustrating the various uses of wavelets referred to in the second section, by resorting to the nine sectors of the Standard & Poor's Depository Receipts (SPDR, called "spiders")—consumer discretionary, consumer staples, energy, financial, health care, industrial, materials, technology, and utilities—along the period 1999 to 2007. A brief description of these data series also is presented in the third section in this chapter, along with discussions on quantifying the degree of co-movement of the cyclical components of the sampled spiders, testing for the presence of variance shifts in such cyclical components, and computing a VaR estimate of an equally weighted portfolio made up of the nine spiders. This chapter ends with a summary of our main findings.

Overall, our results have two important implications for risk modeling. First, as our estimation shows, the degree of co-movement of financial series depends on the timescale under consideration. This finding suggests that portfolio diversification gains will vary across individuals, depending on their investment horizons. Second, similarly to an economic indicator, a financial index can be decomposed into a cyclical and a trend component. The cyclical component, which is associated with medium-term fluctuations, may be subjected to variance shifts over time. The presence of such breakpoints must be taken into consideration when computing a portfolio potential loss in a medium-term horizon.

WAVELETS IN A NUTSHELL

Connor and Rossiter (2005) provide an appealing interpretation of wavelet analysis by pointing out that long-term traders focus on price fundamentals that drive overall trends, while short-term traders react to incoming information within a short-term horizon. Hence, market dynamics in the aggregate is the outcome of the interaction of agents with heterogeneous time horizons. Such heterogeneity, in Connor and Rossiter's view, can be modeled by means of wavelet analysis, a mathematical tool developed in the early 1990s.

Wavelets enable us to decompose a time series into high- and low-frequency components (see, for instance, Percival and Walden, 2000). High-frequency components describe the short-term dynamics, whereas low-frequency components represent the long-term behavior of the series. Wavelets are classified into father and mother wavelets. Father wavelets capture the smooth and low-frequency parts of a time series, whereas mother wavelets describe its detailed and high-frequency parts.

In addition, wavelet-variance analysis makes it possible to decompose the variance of a time series into components that are associated with different time scales (i.e., time horizon). That is, this methodology enables us to conclude which scales are important contributors to the overall variability of a time series at a given time horizon. A concrete application of wavelet-variance analysis is the computation of a wavelet-based VaR. As discussed in Fernandez (2006a) and Fernandez and Lucey (2007), one can easily derive an expression for the VaR of a portfolio based on an empirical representation of CAPM. And, from such a representation a wavelet-based VaR can be readily obtained. The usefulness of a wavelet-based VaR is that it yields an estimate of the potential loss at a particular time horizon, by relying on a frequency decomposition of the data. In particular, the more volatile components of the returns series will contribute more to the overall portfolio risk. Indeed, such components will be associated with the high-frequency components of the returns data.

Another application of wavelets we provide in the empirical section deals with extracting the cyclical component from a time series. In particular, such procedure allows us to study the co-movement of paired time series along the business cycle. In other words, once the trend component of each series has been extracted, we are able to gauge the correlation between the two at the short and medium-term. (The trend components of the two series can be also obtained from the above-mentioned wavelet decomposition. Both trend components will be in turn associated with the long-term behavior of the series.)

Wavelet analysis also allows testing for the presence of variance shifts in the data. This subject is of particular interest to risk gauging. As we know, failing to account for variance breaks will lead to biased estimates of market risk (e.g., Fernandez and Lucey, 2007).

In the next section, we illustrate these alternative applications of wavelet analysis to sector SPDRs. First, we concentrate on the degree of co-movement of the cyclical components of some selected spiders, and study how they have evolved over time. Second, we explore the presence of variance shifts along the business cycle in order to test whether the war in Iraq has turned out to be a source of instability across economic sectors. Third, we focus on a portfolio made up of the nine above-mentioned spiders and construct a rolling VaR measure.

EMPIRICAL RESULTS

The Data

Our data set comprises nine years (January 1999–December 2007) of daily data of the sector SPDRs, which are exchange-traded funds (ETFs) that divide the S&P 500 into nine sector index funds. Spiders are traded like regular stocks on the American Stock Exchange under the ticker symbol SPY. Each share of a spider contains one-tenth of the S&P index and trades at approximately one-tenth of the dollar-value level of the S&P 500. Spiders allow large institutions and traders to keep track of the overall direction of the U.S. main economic sectors. They are also appealing to individual investors engaged in passive management.[2]

The SPDRs include consumer discretionary (automobiles and components, consumer durables, apparel, hotels, restaurants, leisure, media, and retailing); consumer staples (food and drug retailing, beverages, food products, tobacco, household products, and personal products); energy (development and production of crude oil and natural gas, drilling, and other energy-related services); financial (business lines ranging from investment management to commercial and investment banking); health care (health care equipment and supplies, health care providers and services, biotechnology, and pharmaceuticals industries); industrials (aerospace and defense, building products, construction and engineering, electrical equipment, conglomerates, machinery, commercial services and supplies, air freight and logistics, airlines, marine, road and rail, and transportation infrastructure); materials (chemicals, construction materials, containers and packaging, metals and mining, and paper and forest products); technology (semiconductor equipment and products, computers and peripherals, diversified telecommunication services and wireless telecommunication services); and, utilities (water and electrical power and natural gas distribution). In addition, our data set includes the S&P index as a proxy for the market portfolio.

Cyclical Co-Movement over Time

As mentioned in the Introduction section, wavelets enable us to decompose a time series into its cycle and trend. To illustrate, Figure 6.1 and Figure 6.2 show scatter plots of the cycles obtained for the health and energy spiders by assuming a zero-mean random walk and taking scales 1 to 4 from the wavelet decomposition of each index. For both sectors, the Pearson correlation between the random walk and wavelet-based cyclical components is around 0.4.[3]

The cyclical decomposition of the sampled indexes makes it possible in turn to analyze the co-movement of the high-frequency components of the data across economic sectors. In other words, the unit-root component associated with the low-frequency component of the data is removed so that correlation coefficients are free from spuriousness.

For the sake of brevity, we compute rolling wavelet-based correlation coefficients for four paired cyclical components: consumer discretionary/consumer staples, energy/industrials, financial/technology, and health

Figure 6.1 Cyclical Components of Health Care Select SPDR Sector

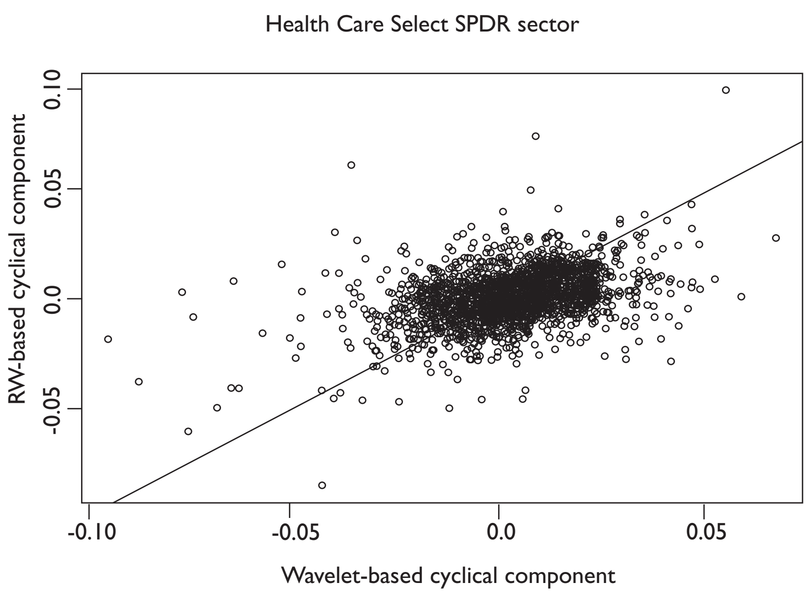
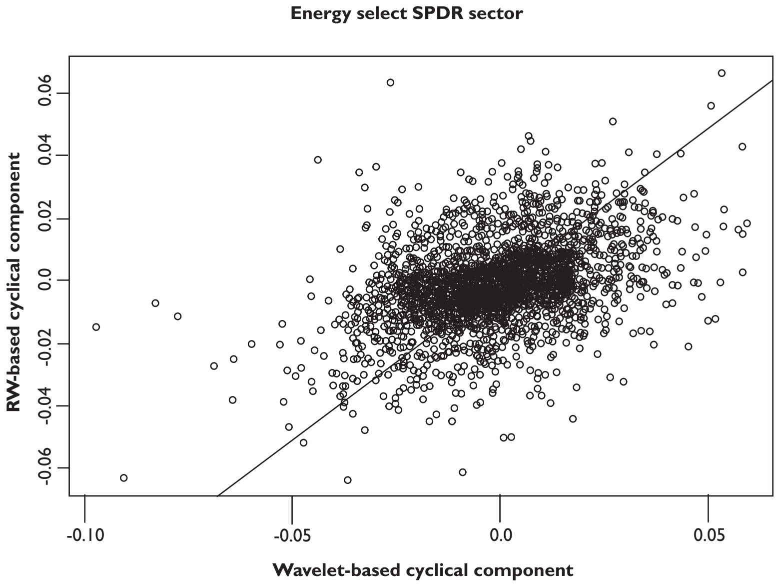


Figure 6.2 Cyclical Components of Energy Select SPDR Sector



care/technology, by taking a rolling window of 450 observations and focusing on scales 2 and 4 (i.e., 4–8 day and 16–32 day dynamics, respectively) of the paired series. As a benchmark, a rolling-correlation coefficient is computed for the raw data.

A couple of remarks prior to the discussion of our results seem relevant. First, the window width suffices to make the computation of rolling wavelet-based correlations numerically feasible. In general, shorter windows tend to cause numerical problems. On the other hand, by choosing a window as short as possible, our correlation estimates become less sensitive to the potential existence of structural breaks along the sample period. Second, gauging the co-movement of paired cyclical components at different timescales boils down to describing the degree of synchronicity of paired de-trended series at alternative time horizons.

Our results obtained for the period 2001 to 2007 are depicted in Figures 6.3 to 6.6. As we see, the behavior of the rolling correlations is similar in the raw data and at its second scale. This comes as no surprise because each cycle already summarizes the high-frequency parts of each corresponding sector spider. By contrast, more noticeable differences arise when comparing the co-movement dynamics in the raw data and at its fourth scale.

For instance, when focusing on Figure 6.6, we see that, while the correlation of health care and technology reached a peak of 0.66 in the raw data

Figure 6.3 Rolling Wavelet-based Correlation of Cyclical Components of Consumer Discretionary and Consumer Staples Select SPDR Sectors

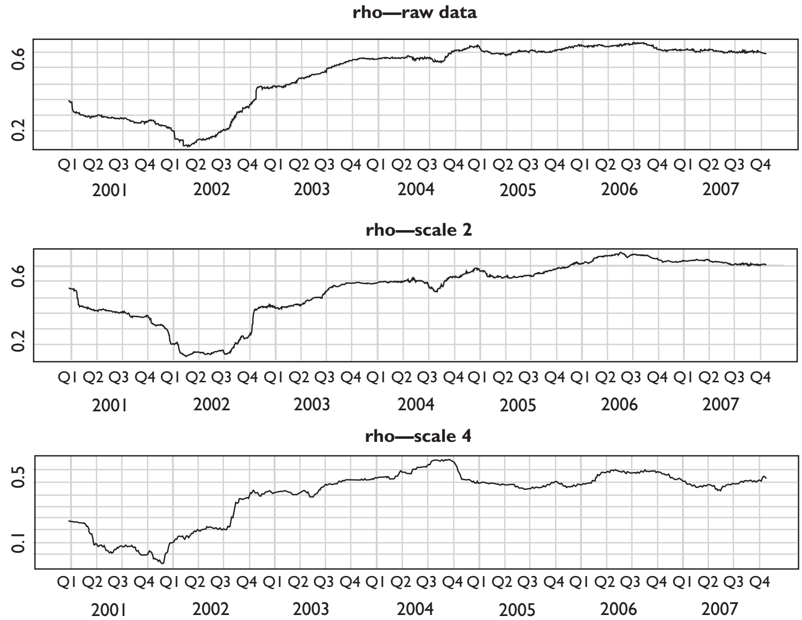


Figure 6.4 Rolling Wavelet-based Correlation of Cyclical Components of Energy and Industrial Select SPDR Sectors

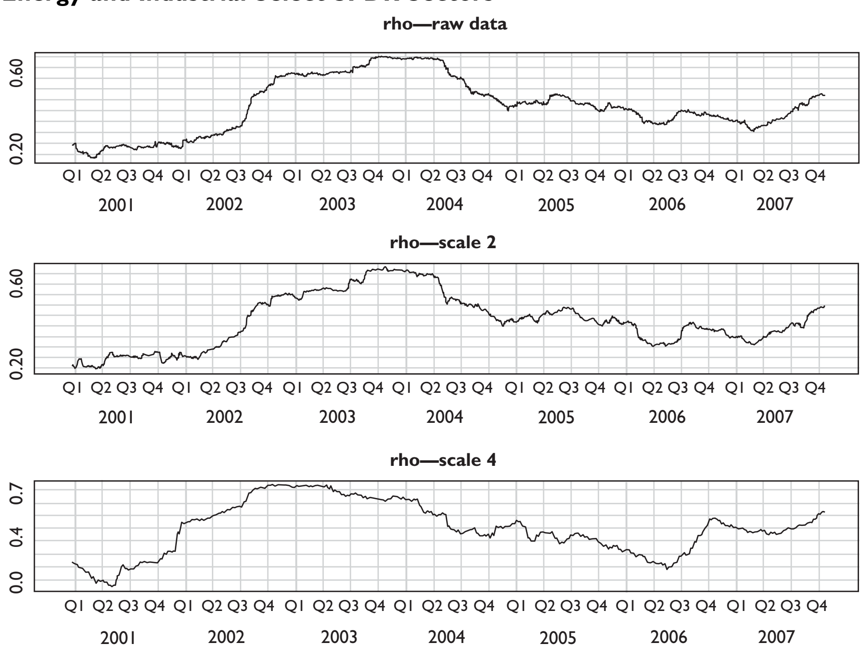


Figure 6.5 Rolling Wavelet-based Correlation of Cyclical Components of Financial and Technology Select SPDR Sectors

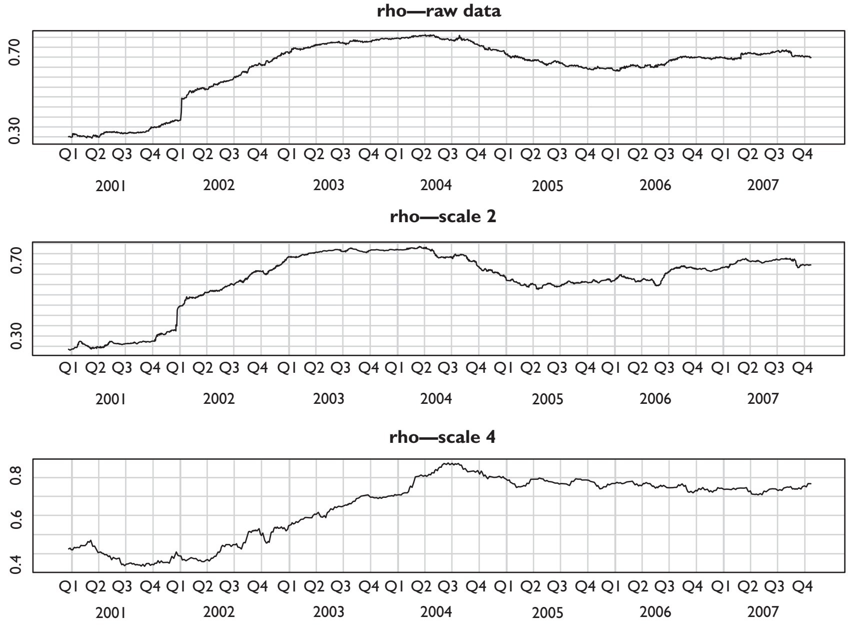
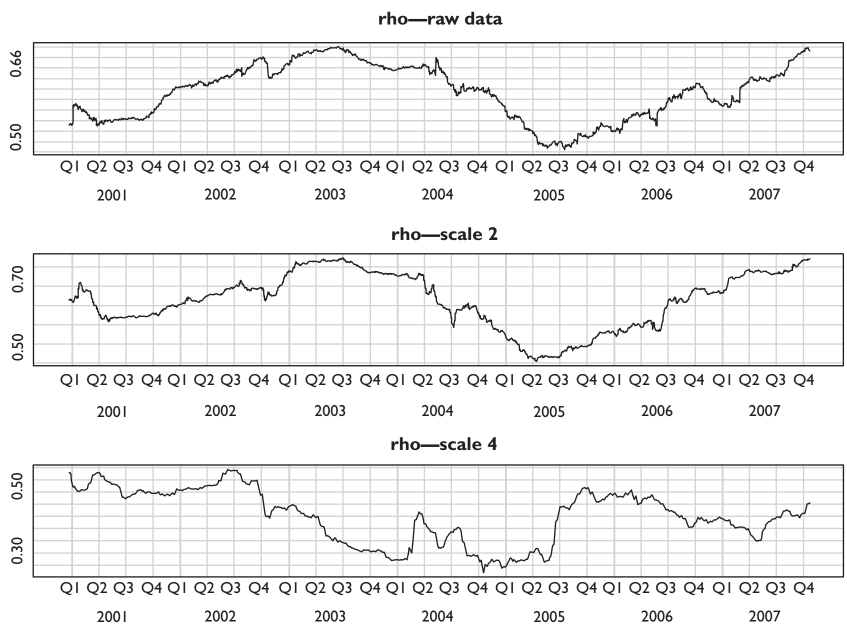


Figure 6.6 Rolling Wavelet-based Correlation of Cyclical Components of Health Care and Technology Select SPDR Sectors



and of 0.7 at scale 2 during the second and third quarters of 2003, such correlation exhibited a decreasing trend at scale 4 by dropping from around 0.5 to 0.4 from the second to the third quarter of 2003.

With the exception of the consumer discretionary/consumer staples pair, we observe a greater degree of co-movement in the raw data and at its second scale during 2003 to 2004. In particular, the industrial sector appears to have been more sensitive to the behavior of the energy sector during that time period. It is possible that the beginning of the war in Iraq may have been a driving factor behind such a pattern.

On the other hand, it is worth noticing that the cycles of both the health care and financial sectors were positively and highly correlated with that of the technology sector for the most part of the sample period. Specifically, toward the end of the sample, the correlation exhibited by the two pairs was more than 0.6. This finding may be indicative of a nonnegligible dependence of the financial and health care sectors on technological issues.

Variance Shifts in SPDR Cycles

Given that the time period we are examining includes two major political events, namely, 9/11 and the ongoing war in Iraq, it is likely that the sampled series experienced variance shifts. Testing for the existence of such breakpoints is relevant to risk management as quantifying financial risk may be quite sensitive to the assumption of variance constancy.

We again concentrate on cyclical components and take a rolling window of 450 observations to compute the so-called D statistic (e.g., Percival and Walden, 2000, chapter 9). Specifically, we construct a rolling D statistic during 2001 to 2007 for the second scale (i.e., 4–8 daily dynamics) of the de-trended consumer staples, health care, technology, and utilities spiders. A violation of the null hypothesis of variance constancy is encountered when the D statistic exceeds the critical value represented by the dashed horizontal line at the selected significance value (5 percent in this case).

As illustrated in Figure 6.7 and Figure 6.8, the null hypothesis is strongly rejected for health care and technology during 2001 to 2004 when taking a window length of about two years. This finding seems to suggest that during this time period the cyclical variability of these two spider sectors most probably exhibited regime shifts that lasted less than two years. A similar conclusion is drawn for consumer staples and utilities, although in these two cases the evidence against the null is not so strong for 2001.

Figure 6.7 Rolling Wavelet-based Variance Shift Test Applied to Cyclical Components of Health Care Select SPDR Sector

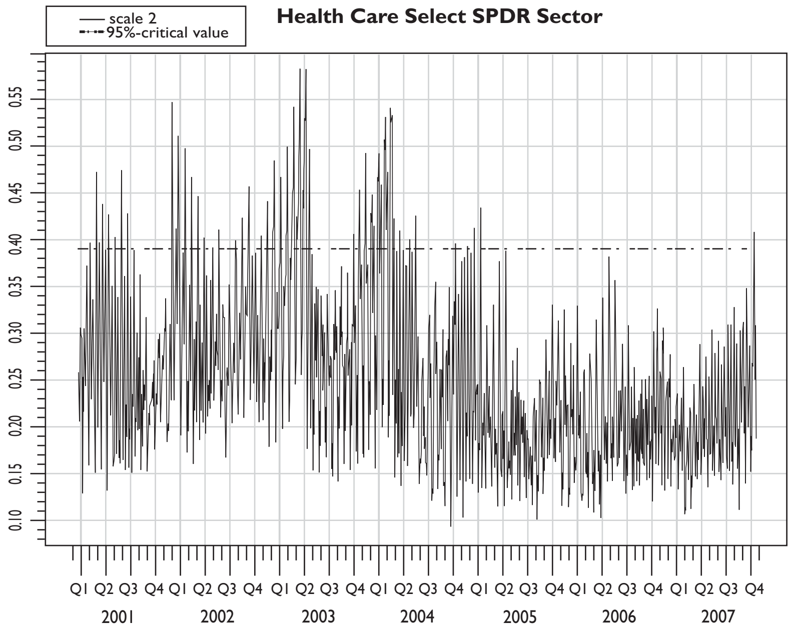
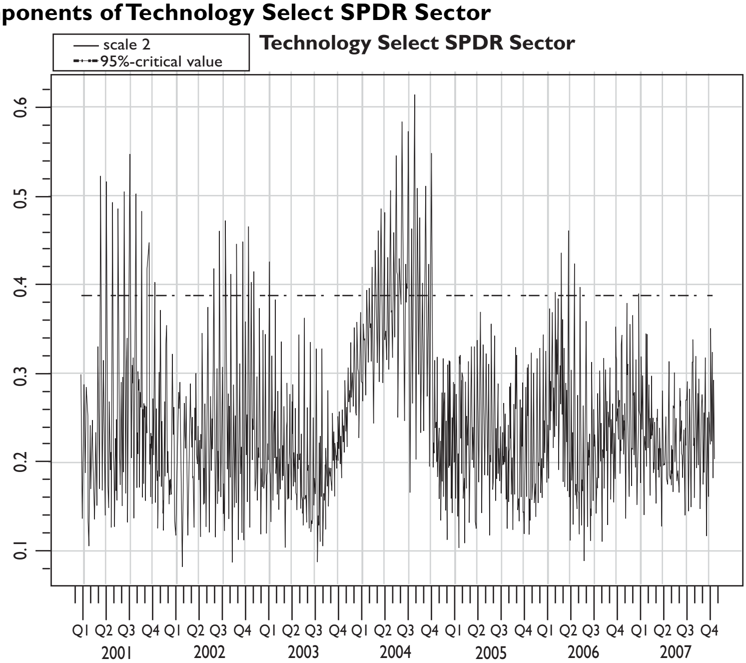


Figure 6.8 Rolling Wavelet-based Variance Shift Test Applied to Cyclical Components of Technology Select SPDR Sector



When focusing on 2005 to 2007, we find that the evidence against variance constancy is much less compelling for the four spiders under consideration. In particular, the cyclical component of the health care spider did not exhibited any variance shifts during that time period when considering the dynamics at 4 to 8 day horizon.

Rolling Wavelet-based Value at Risk of a SPDR Portfolio

As previously discussed, 2001 to 2004 was a particularly volatile time period during which the de-trended sampled indexes experienced various regimens. We further investigate this issue by computing a rolling wavelet-based estimate of the VaR of an equally weighted portfolio made up of the nine SPDR sectors.

Specifically, by taking successive rolling-windows of 500 observations (i.e., about two years of data), we obtain series of 1,763 observation of wavelet-based betas for each of the nine return indexes at alternative timescales. The market portfolio is approximated by the S&P 500 index. Descriptive statistics of the rolling betas at scales 1 to 6 are provided in Table 6.1. The energy and utility sectors are the ones less subject to systematic risk, while the opposite holds for the consumer staples and industrial sectors. As reported in previous studies, e.g., Gençay, Whitcher, Selçuk (2005) and Fernandez (2005, 2006a), the CAPM tends to have more predictive power at a medium-term horizon (i.e., at the intermediate scales 3 and 4, in this case), as the computed betas tends to reach a maximum at such scales.

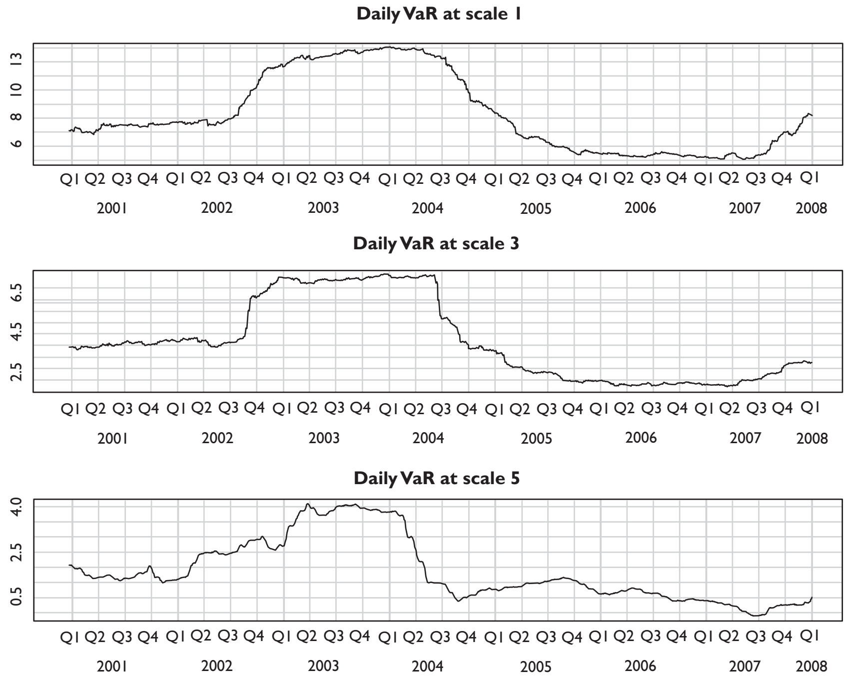
Ninety-five percent VaR estimates for a $1,000 equally weighted portfolio at scales 1, 3, and 5 (i.e., 2–4 day, 8–16 day, and 32–64 day dynamics, respectively) are depicted in Figure 6.4. As we see, the portfolio VaR remained approximately constant during 2001 and the first two quarters of 2002. From mid–2002 onward, it exhibited an increasing trend to reach ultimately a maximum and remain relatively stable over 2003 to 2004. Specifically, the greatest potential portfolio loss over the sample period was $13, $6.5, and $4 per day at scales 1, 3, and 5, respectively, at a 95 percent confidence level. Afterward, the portfolio VaR displayed a decreasing trend at the three reported scales during 2005 to 2007, to end up at a smaller level than that observed in 2001. (At the first scale, however, the VaR tends to equal that observed in 2001 toward the end of 2007.)

Table 6.1 Descriptive Statistics of Wavelet-based Rolling Betas

SPDR Sector	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6
(a) Mean						
Consumer discretionary	0.680	0.683	0.679	0.677	0.675	0.678
Consumer staples	0.757	0.767	0.726	0.782	0.708	0.772
Energy	0.348	0.351	0.326	0.355	0.351	0.362
Financial	0.683	0.670	0.706	0.698	0.715	0.723
Health care	0.732	0.761	0.704	0.721	0.681	0.633
Industrial	0.775	0.779	0.781	0.769	0.775	0.722
Materials	0.530	0.522	0.546	0.553	0.503	0.526
Technology	0.548	0.563	0.539	0.534	0.525	0.516
Utilities	0.489	0.497	0.465	0.539	0.460	0.506
(b) 1st quartile						
Consumer discretionary	0.618	0.606	0.634	0.639	0.570	0.660
Consumer staples	0.576	0.559	0.469	0.716	0.527	0.610
Energy	0.212	0.219	0.202	0.228	0.224	0.218
Financial	0.667	0.657	0.701	0.689	0.698	0.669
Health care	0.663	0.703	0.648	0.653	0.602	0.542
Industrial	0.739	0.745	0.737	0.712	0.719	0.659
Materials	0.468	0.472	0.482	0.445	0.431	0.381
Technology	0.443	0.448	0.429	0.459	0.440	0.426
Utilities	0.461	0.405	0.437	0.427	0.405	0.392
(c) 3rd quartile						
Consumer discretionary	0.756	0.767	0.746	0.751	0.755	0.777
Consumer staples	0.926	0.943	0.899	0.975	0.899	1.045
Energy	0.520	0.539	0.465	0.522	0.511	0.511
Financial	0.788	0.789	0.798	0.767	0.778	0.850
Health care	0.803	0.863	0.752	0.806	0.752	0.760
Industrial	0.805	0.822	0.815	0.809	0.841	0.769
Materials	0.668	0.655	0.670	0.716	0.624	0.569
Technology	0.603	0.618	0.610	0.600	0.613	0.589
Utilities	0.550	0.551	0.535	0.646	0.504	0.605

The wavelet-based rolling betas are computed by taking a window length of 500 observations and by using the S&P 500 index as the market portfolio.

Figure 6.9 95 Percent Rolling Value at Risk of an Equally Weighted Portfolio Composed of SPDR Sectors



A wavelet-based value at risk is obtained by taking rolling windows of 500 observations. The portfolio value is $1,000.

As discussed by Fernandez (2005, 2006a) and by Fernandez and Lucey (2007), and as can be also observed from Figure 6.9, the wavelet-based portfolio VaR tends to decrease as we move to the upper scales of the return data. The reason is that the least volatile parts of the sampled series (i.e., their lowest-frequency components) are precisely located at the upper scales, which translates into lower portfolio risk.

CONCLUSION

Wavelet analysis has become increasingly popular in the finance field as a tool to characterize the short- and long-term behavior of financial series. Such characterization is possible due to the time localization of the high- and low-frequency components of a time series yielded by the wavelet technique. A simple interpretation of the wavelet-based decomposition of a time series is that its high- and low-frequency parts represent, respectively, short- and long-term investors' viewpoints. Specifically, long-term traders focus on price fundamentals that drive overall trends (i.e., low-frequency components), while short-term traders react to incoming information within a short-term horizon (i.e., high-frequency components).

This chapter illustrated the usage of the wavelet technique for gauging asset return co-movement, structural variance shifts, and VaR at alternative timescales. Specifically, we concentrated on the nine SPDR sectors—consumer discretionary, consumer staples, energy, financial, health care, industrial, materials, technology, and utilities—along the period 1999 to 2007.

Our estimation results suggest that, in general, a greater degree of co-movement in the raw data and at its second scale was observed during 2003 to 2004. In addition, we conclude that 2001 to 2004 was a particularly volatile period over which the cyclical components of the sampled U.S. sector indexes experienced several variance shifts. The existence of such variance breakpoints had an impact on the overall risk of an equally-weighted portfolio composed of the sampled indexes. Indeed, when computing a rolling estimate of the VaR of such a portfolio, we find that along the sample period VaR reached its maximum during 2003 to 2004.

From these findings, we conclude that portfolio diversification gains will be contingent on agents' investment horizon, and that the presence of variance breakpoints in the cyclical component of a financial index must be taken into account when computing VaR.

REFERENCES

Connor, J. and Rossiter, R. (2005) Wavelet Transforms and Commodity Prices. *Studies in Nonlinear Dynamics & Econometrics*, 9(1): article 6.

Fernandez, V. (2005) The International CAPM and a Wavelet-based Decomposition of Value at Risk. *Studies of Nonlinear Dynamics & Econometrics*, 9(4): article 4.

Fernandez, V. (2006a) The CAPM and Value at Risk at Different Time Scales. *International Review of Financial Analysis*, 15(3): 203–219.

Fernandez, V. (2006b) Does Domestic Cooperation Lead to Business-cycle Convergence and Financial Linkages? *The Quarterly Review of Economics and Finance*, 46(3): 369–396. [Special issue on Real and financial aspects of financial integration.]

Fernandez, V. (2007) Stock Markets Turmoil: Worldwide Effects of Middle East Conflicts. *Emerging Markets Finance and Trade*, 43(3): 61–105.

Fernandez, V. (2008) Multi-period Hedge Ratios for a Multi-asset Portfolio When Accounting for Returns Co-movement. *The Journal of Futures Markets*, 28(2): 182–207.

Fernandez, V. and Lucey, B. (2007) Portfolio Management under Sudden Changes in Volatility and Heterogeneous Investment Horizons. *Physica A: Statistical Mechanics and its Applications*, 375(2): 612–624.

Gençay R., Selçuk F., and Whitcher B. (2002) *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. San Diego, CA: Academic Press.

Gençay, R., Whitcher, B., and Selçuk, F. (2003) Systematic Risk and Time Scales. *Quantitative Finance*, 3(2): 108–116.

Gençay, R., Whitcher, B., and Selçuk, F. (2005) Multiscale Systematic Risk. *Journal of International Money and Finance*, 24(1): 55–70.

In, F. and Kim, S. (2006) The Hedge Ratio and the Empirical Relationship between the Stock and Futures Markets: A New Approach Using Wavelet Analysis. *Journal of Business*, 79(2): 799–820.

Karuppiah, J. and Los, C. (2005) Wavelet multi-resolution analysis of high-frequency Asian FX rates, Summer 1997. *International Review of Financial Analysis* 14(2), 211–246. Quantitative Issues in Finance.

Kyaw, N., Los, C., and Zong, S. (2006) Persistence Characteristics of Latin American Financial Markets. *Journal of Multinational Financial Management*, 16(3): 269–290.

Lien, D. and Shrestha, K. (2007) An Empirical Analysis of the Relationship between Hedge Ratio and Hedging Horizon Using Wavelet Analysis. *The Journal of Futures Markets*, 27(2): 127–150.

Lin, S. and Stevenson, M. (2001) Wavelet Analysis of the Cost-of-carry Model. *Studies in Nonlinear Dynamics & Econometrics*, 5(1): article 7.

Percival, D. and Walden, A. (2000) *Wavelets Analysis for Time Series Analysis*. Cambridge, U.K.: Cambridge University Press.

Yogo, M. (2008) Measuring Business Cycles: A Wavelet Analysis of Economic Time Series. *Economics Letters*, 100: 208–212.

NOTES

1. In particular, Gençay, Whitcher, and Selçuk's 2002 textbook contains various applications of wavelets written in R, a programming language and software environment for statistical computing and graphics. A comprehensive list of wavelet software can be found at http://www.amara.com/current/wavesoft.html. In this study, we utilize S-Plus 8.0 and its library S+ Wavelets 2.0.

2. Daily observations of spiders can be downloaded freely at finance.yahoo.com.
3. Given that wavelet filtering performs poorly at the extremes of a data series, the first and last 50 observations of each reconstructed cycle are discarded.

This page intentionally left blank

CHAPTER 7

VOLATILITY ASYMMETRY AND LEVERAGE
Some U.S. Evidence

Emawtee Bissoondoyal-Bheenick and Robert Brooks

ABSTRACT

Financial markets have become more interconnected and hence there is a greater need to understand the volatility of stock returns, given its important role in portfolio selection and asset management. An important finding is that when price drops, the volatility of its return typically rises. There has been extensive research work undertaken which concerns the apparent asymmetry in the relationship between equity market returns and volatility. Hence, quite a few explanations have been given for the observed asymmetry, namely, financial leverage, operating leverage, and risk premium. This chapter focuses on the analysis of the modeling of the volatility of the U.S. market. It aims to answer important questions about the determinants of volatility asymmetry at the individual stock level, in particular by considering whether volatility asymmetry is a function of debt-equity ratio for stocks listed in the U.S. market. The analysis uses the APARCH model, which has proven to be very successful on index data for a range of countries to construct the measures of volatility asymmetry. The evidence is not consistent with the stylized arguments of a leverage effect in that the debt ratio does not seem to be a major variable that explains volatility asymmetry.

INTRODUCTION

In current market conditions a rigorous approach to risk measurement is essential. Financial markets have become more interconnected and hence there is a greater need to understand the volatility of stock returns, given its important role in portfolio selection and asset management. An important finding is that when price drops, the volatility of its return typically rises. Extensive research work has been undertaken that concerns the apparent asymmetry in the relationship between equity market returns and volatility. Hence, there are quite a few explanations given for the observed asymmetry, namely, financial leverage, operating leverage, and risk premium. The first explanation is that as the price of stock goes down, a firm's financial leverage goes up which explains the higher volatility of stock returns. The operating leverage argument is that as a firm forecasts a lower cash flow level, the stock price tends to fall and stock becomes more volatile. The identification of the determinants of volatility is a model specification issue. Thus, given the key role that volatility plays in the models of the risk-return trade-off, this creates a potential model risk problem. That is, we might draw inaccurate conclusions about the nature of the risk-return trade-off if we do not model the determinants of volatility accurately.

Early literature on firm level volatility asymmetry has been undertaken by Black (1976), Christie (1982), Cheung and Ng (1992), and Duffee (1992), and most of these papers have attributed the asymmetric return-volatility relationship to changes in financial leverage. However, it can be argued that the magnitude of the decline in stock prices on future volatilities seems too large to be explained only by the leverage effect. The more recent literature tends to focus mostly on index-level volatility attributes which provide another leading explanation given in the literature that is the risk premium effect; see, for example, French, Schwert, and Stambaugh (1987), Glosten, Jaganathan, and Runkle (1993), Bekeart and Wu (2000), and Wu (2001). They argue that an increase in unexpected volatility will increase the expected future volatility. Then the risk premium will also increase causing prices to drop. Wu (2001) finds evidence that both financial leverage and risk premium are important determinants of volatility asymmetry. While most of the previous studies have focused on a country index level, the aim of this chapter is to extend the analysis at the individual firm level with a particular focus on the U.S. market. In addition, this chapter employs the autoregressive conditional heteroscedasticity (ARCH) family of models introduced by Engle (1982) which in fact plays a key role for studies in volatility. Following this model, there have been a number of improved models developed to better capture volatility asymmetry. A popular model, which captures the asymmetric impact of good and bad news and a leverage effect, is the asymmetric power ARCH (APARCH) model introduced by Ding, Granger, and Engle (1993). They demonstrate how the simple ARCH model is nested within the general APARCH model. The applicability of this model has been very successful in developed stock markets (United Kingdom, Japan, Hong Kong, New Zealand, Germany, France, Singapore, Canada, and Australia); see Brooks et al. (2000). The focus in the literature has been mostly the application of the APARCH model at the country index level. This chapter will make a significant extension to the previous literature via the analysis of volatility asymmetry at the individual firm level using the APARCH model to model risk.

DATA AND MODELING FRAMEWORK

In terms of the data utilized in this study, individual stocks in the U.S. market have been considered with a particular focus on daily returns for each of the companies for a period of five years, 2002 to 2006, which was collected from DataStream database. The number of companies that were initially downloaded was 984 firms. As far as the debt ratios are concerned, these were compiled from the OSIRIS database where information about the gearing level of the company, long-term debt, short-term debt, and net assets is available. The data were then cleaned for the five-year period and a new database created with the returns of each firm and the associated level of gearing for each of the firms. This process eliminated a number of firms as there was a mismatch of the firms listed in both of the databases. The final set of data that was used in the study hence included 710 firms with daily return data for the five years. In addition to the debt level, some other variables which are typical of a firm's characteristics have also been collected, that is, the size and volume data. The proxy that was used for size is market value of the firm and these were the average for the five-year period. The modeling framework will be as follows in the following equation:

$$R_{it} = \rho_j R_{it-j} + \ldots + \rho_j R_{it-k} + \varepsilon_{it} \tag{7.1}$$

$$\varepsilon_{it} \sim (0, \sigma^2_{it}) \tag{7.2}$$

$$\sigma_t^d = \lambda + \sum_{i=1}^{p} \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^d + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^d \tag{7.3}$$

The model can be estimated at the individual stock level using daily returns data in the Eviews package. In Equation (7.1) the returns on the individual stock index (R_{it}) follow an autoregressive process and autoregressive lags are included in response to significant lags in the partial autocorrelation function for each individual series. In Equation (7.2) the conditional errors, ε_{it} are allowed to follow either a normal or a t distribution, with a time varying conditional variance (σ^2_{it}). The conditional variance is modeled following the APARCH model introduced by Ding et al. (1993). The power term (d) in Equation (7.3) captures both the conditional standard deviation ($d = 1$) and conditional variance ($d = 2$) as special cases. The γ and d parameters can be used to construct a measure of the degree of volatility asymmetry at the individual stock level. The volatility asymmetry measure is then calculated. This measure is used by Jayasuriya, Shambora, and Rossiter (2005) and Brooks (2007) to compare the degree of volatility asymmetry across countries. The following formula was applied to calculate volatility:

$$\text{Volatility Asymmetry} = ((1 + \gamma)/(1 - \gamma))^d \tag{7.4}$$

Once the volatility asymmetry was calculated, this was mapped on the following equation to test for the relationship between volatility and gearing level of each firm.

$$\text{Volatility Asymmetry} = f(\text{Debt, Volume, Size, Beta}) \tag{7.5}$$

This cross-sectional analysis explores the relationship between firm leverage as measured by the gearing ratio and the estimated volatility asymmetry after adjustment for other firm level variables such as firm size, volume, and beta. The results of this analysis will provide information on the determinants of volatility asymmetry and identify possible model risk issues that flow from a focus on the relationship between financial leverage and volatility asymmetry.

RESULTS

The market model was estimated for all the firms to provide an estimate of the values of α and β. For estimation of the model the lagged ARCH terms (αi) are set to 1 ($p = 1$), and then lagged GARCH terms (βj) are set to 1 ($q = 1$). The asymmetry in the model is captured via the parameter γ and the power term d is also estimated. In contrast to the current literature across countries, this study finds there are around 325 firms for which the power parameter is significantly different from unity. Out of these firms, there are around 228 where the power parameter ranges between 1.3 to 2 and around 40 firms for which the power term is significantly greater than 2. To estimate Equation (7.2), the normal conditional errors were used.

The volatility asymmetry is then calculated using the metric presented in Equation (7.4). Of the whole sample of 710 firms, 286 firms have a volatility asymmetry measure of greater than 3. Equation (7.5) is being run using white standard error to provide correct estimates of the coefficient covariances in the presence of heteroscedasticity. The results are reported in Table 7.1. Table 7.1 reports the results of the regression in Equation (7.5). The tests have been conducted for all firms and given that the volatility measure includes some high estimates the test is repeated by trimming extreme deciles, followed by trimming extreme quintiles, and finally trimming the extreme quartiles. Contrary to the results in the current literature whereby the asymmetric return-volatility relationship is explained by changes in financial leverage, the results indicate that in the case of the U.S. firms, the gearing level of the firm does not provide much explanation of the volatility-return relationship. The debt level is statistically significant at the 10 percent level for all the firms. However, the sample includes some high value volatility asymmetry measure that may be driving the results, hence by trimming the extreme deciles, quintiles, and quartiles, the sample excludes these large estimates. However, the results differ in that in these three cases the debt level is not significant, though the volume of trading seems to be significant for all three options used and the risk measure, beta, is significant when the extreme quintiles and quartiles are excluded. This suggests that the determinants of volatility asymmetry may vary across firms which may be an important model specification and risk issue to be considered in understanding the risk-return trade-off. Table 7.1 reports the results of estimation Equation (7.5) for all firms. The analysis is also extended by excluding the extreme deciles, quintiles, and quartiles of firms by volatility asymmetry. Values in parenthesis are p values.

Table 7.1 Volatility Asymmetry: All Firms

Dependent Variable: Volatility Asymmetry

	All Firms: 1–710	**Trimming Extreme Deciles: 71–639**	**Trimming Extreme Quintiles: 142–568**	**Trimming Extreme Quartiles: 177–533**
c	105389	5.6587	2.4536	2.3669
	(0.5621)	(0.0002)	(0.0000)	(0.0000)
Volume	1.7471	−0.0002	0.0000	0.0000
	(0.8786)	(0.0922)*	(0.0024)†	(0.0124)†
Market value	−0.9605	0.0000	0.0000	0.0000
	(0.2213)	(0.2284)	(0.7103)	(0.8478)
Beta	169087	1.9816	0.4232	0.2532
	(0.3222)	(0.1590)	(0.0129)†	(0.0618)*
Gearing	−599.1774	−0.0045	−0.0003	0.0001
	(0.0654)*	(0.2878)	(0.5959)	(0.8680)

* Denotes statistical significance at 10% level.

† Denotes statistical significance at 5% level.

Following the results in Table 7.1, the analysis was repeated by considering each decile of firms in the sample. The results are summarized in Table 7.2. The results however do not improve much and are consistent with the previous analysis, that is, the leverage of the firm still does not explain volatility asymmetry. Table 7.2 reports the results of Equation (7.5), with the analysis extended to analyze firms for each decile by volatility asymmetry. Values in parenthesis are *p* values.

The analysis was repeated to include the industry classification as a dummy variable. The industries were obtained from the NYSE. The classifications are as follows: basic materials (42 firms), consumer goods (70 firms), consumer services (105 firms), financials (67 firms), health care (67 firms), industrials (144 firms), oil and gas (55 firms), technology (89 firms), telecommunications (17 firms), and utilities (50 firms). The results are summarized in Table 7.3, which reports the results of Equation (7.5), with the analysis extended to include industry dummies and the firm by excluding the extreme deciles, quintiles, and quartiles. Values in parenthesis are *p* values.

Consistent with previous results obtained in this study, debt is statistically significant when the analysis is done using the whole sample of firms at the 10 percent level. The volume of trading is significant when the extreme values are excluded from the sample and the risk measure is significant when the extreme quintiles and quartiles are excluded. However, none of the industries' dummies add any value to this analysis. The regression is repeated for each decile as in the previous cases, but the results are not reported here as there is no statistical significance in any of the results obtained. The analysis was then further extended and the regression was run for each of the industries separately. The results obtained are still consistent with the general finding of this study in that the level of debt does not provide the explanation for volatility; hence the results are not reported in this chapter.

CONCLUSION

Asset return volatilities are central to finance, whether in asset pricing, portfolio allocation, or market risk measurement. Hence there has been much focus on time-varying volatility and associated tools for its measurement, modeling, and forecasting of risk. While most of the previous studies have focused on a country index level, this chapter has explored the applicability of the APARCH model in the context of modeling the volatility of individual firms in the U.S. market for a period of five years, 2002 to 2006. The analysis has been conducted by repeating the models using different

Table 7.2 Volatility Asymmetry—Deciles Analysis

Decile Analysis

	D1: 1–71	**D2: 72–142**	**D3: 143–213**	**D4: 214–284**	**D5: 285–355**	**D6: 356–426**	**D7: 427–497**	**D8: 498–568**	**D9: 569–639**	**D10: 640–710**
c	0.7565	1.1876	1.3801	1.6992	2.1279	2.6585	3.5163	5.5698	84.6113	3764985
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0065)	(0.3575)
Volume	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0008	33.3124
	(0.2624)	(0.0056)†	(0.9393)	(0.0000)†	(0.1611)	(0.1072)*	(0.2702)	(0.8908)	(0.3552)	(0.8631)
Market value	(0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	−0.0003	−3.3739
	(0.4359)	(0.0002)†	(0.7011)	(0.9778)	(0.1238)	(0.1316)	(0.4879)	(0.7523)	(0.1950)	(0.7946)
Beta	0.0345	−0.0288	−0.0025	−0.0124	0.0024	−0.0329	−0.1313	−0.1381	−36.4701	−976865.8
	(0.6254)	(0.0208)†	(0.8883)	(0.5211)	(0.9672)	(0.6259)	(0.1338)	(0.6420)	(0.1488)	(0.7496)
Gearing	−0.0001	0.0000	0.0000	0.0000	−0.0002	0.0001	0.0003	−0.0009	−0.0520	−9169
	(0.4171)	(0.7288)	(0.7696)	(0.7607)	(0.3983)	(0.6846)	(0.4019)	(0.1678)	(0.0591)*	(0.2038)

* Denotes statistical significance at 10% level.

† Denotes statistical significance at 5 % level.

Table 7.3 Volatility Asymmetry and Industry Effects

Dependent Variable: Volatility Asymmetry

	All Firms: 1–710	Trimming Extreme Deciles: 71–639	Trimming Extreme Quintiles: 142–568	Trimming Extreme Quartiles: 177–533
c	−90818	7.0499	2.5847	2.5542
	(0.5463)	(0.0864)	(0.0000)	(0.0000)
Volume	−0.1924	−0.0002	0.0000	0.0000
	(0.9871)	(0.0500)†	(0.0106)*	(0.0286)†
Market value	−1.0189	0.0001	0.0000	0.0000
	(0.3480)	(0.1549)	(0.7403)	(0.9015)
Beta	163190	1.9142	0.4493	0.2714
	(0.3443)	(0.1596)	(0.0093)†	(0.046)†
Gearing	−653.7862	−0.0070	−0.0003	0.0001
	(0.0756)*	(0.1460)	(0.6555)	(0.7846)
Basic materials	50611	−4.1044	0.2054	−0.0181
	(0.3728)	(0.3159)	(0.7579)	(0.9735)
Consumer goods	−2903	−5.2638	−0.2617	−0.0727
	(0.9574)	(0.1939)	(0.6574)	(0.8823)
Consumer services	363659	−3.9264	−0.0725	−0.2544
	(0.1574)	(0.3353)	(0.9032)	(0.6020)
Financials	32459.35	−4.3633	−0.3249	−0.3992
	(0.5483)	(0.2885)	(0.5836)	(0.4134)
Health care	630820.1	0.8481	−0.1895	−0.2092
	(0.3245)	(0.8861)	(0.7496)	(0.6752)
Industrials	17849.53	2.0359	−0.1890	−0.3395
	(0.7560)	(0.7023)	(0.7461)	(0.4773)
Oil and gas	326860.1	1.8606	−0.2262	−0.1509
	(0.2987)	(0.7622)	(0.7051)	(0.7667)
Technology	459159.7	1.9737	−0.3425	−0.1892
	(0.0901)	(0.7098)	(0.5545)	(0.6955)
Utilities	65344.93	−2.7156	0.2404	0.0623
	(0.3111)	(0.5110)	(0.7015)	(0.9070)

* Denotes statistical significance at 10% level.

† Denotes statistical significance at 5 % level.

subsamples from the whole data set as well as by including the industry classification in the models. The results obtained from this study are inconsistent with what has been argued in the literature in this area. While most studies have attributed the asymmetric return-volatility relationship to changes in financial leverage, the results obtained for this volatility asymmetry measure for this time period in the case of U.S. individual stock is different. The gearing level of the firm does not seem to be an important driver of the return-volatility relationship. This suggests that studies that focus on leverage as the prime determinant of volatility might be exposed to model risk issues.

REFERENCES

Bekeart, G. and Wu, G. (2000) Asymmetric Volatility and Risk in Equity Markets. *Review of Financial Studies*, 13: 1–42.

Black, F. (1976) Studies of Stock Price Volatility Changes. *Proceedings of the Business and Economic Statistics of the American Statistical Association*: 177–181.

Brooks, R. (2007) Power ARCH Modelling of the Volatility of Emerging Equity Markets. *Emerging Markets Review*, 8: 124–133.

Brooks, R., Faff, R., McKenzie, M., and Mitchell, H. (2000) A Multi-country Study of Power ARCH models and National Stock Market Returns. *Journal of International Money and Finance*, 19: 377–397.

Cheung, Y. and Ng, L. (1992) Stock Price Dynamics and Firm Size; An empirical investigation. *Journal of Finance*, 47: 1985–1997.

Christie, A. (1982) The Stochastic Behaviour of Common Stock variances: Values, Leverage and Interest Rates Effects. *Journal of Financial Economics*, 10: 407–432.

Ding, Z., Granger, C., and Engle, R. (1993) A Long Memory Property of Stock Market Returns and a New Model. *Journal of Empirical Finance*, 1:83–106.

Duffee, G. (1992) Stock Returns and Volatility: A Firm Level Analysis. *Journal of Financial Economics*, 37: 399–420.

Engle, R. (1982) Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50: 987–1007.

French, K.R., Schwert G.W., and Stambaugh R.F. (1987) Expected Stock Returns and Volatility. *Journal of Financial Economics*, 19: 3–30.

Glosten, L., Jaganathan, R., and Runkle, D. (1993) On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance*, 68: 1779–1801.

Jayasuriya, S., Shambora, W., and Rossiter, R. (2005) Asymmetric Volatility in Mature and Emerging Markets. Working paper, Ohio University, Toledo, OH.

Wu, G. (2001) The Determinants of Asymmetric Volatility. *Review of Financial Studies*, 14: 837–859.

This page intentionally left blank

CHAPTER 8

THE EFFECTS OF DIFFERENT PARAMETER ESTIMATION METHODS ON OPTION PRICING

An Empirical Analysis

Zeynep Iltuzer Samur,
Cumhur Ekinci, and Oktay Tas

ABSTRACT

In this study, we aim to determine the extent to which the GARCH option pricing model developed by Duan (1995) is robust in determining varying parameter estimation methods and in capturing the effects of using different objective functions such as implied volatility error, pricing error, percentage pricing error, and absolute pricing error on overall and relative pricing performance. Hence, with an application to S&P 500 index options, we attempt to discover which parameter estimation method minimizes the model risk.

INTRODUCTION

Derivative instruments have reached remarkable trading volumes in the last three decades. Parallel with this remarkable growth, many pricing models have been developed. Banks and other financial institutions use these mathematical models to determine the prices of securities for either speculating or hedging. As a consequence, they have become increasingly exposed to some risks such as applying the wrong model, implementing the model improperly, or using inappropriate data. Known as "model risk," this notion occupies more and more the work in the field. For example, in its second accord, the Basel Committee has expanded the scope of risks that financial institutions are exposed to in such a way as to cover the operational risk, one important component of which is the model risk.

There are a number of reasons why building a model can engender some risks. Derman (1996) provides a comprehensive analysis of the assumptions of models and the list of their consequent risk. Figlewski (1998) discussed the reasons that give rise to model risk and groups them as follows:

1. Although a correct model is applied, an incorrect solution may be obtained due to technical mistakes. For example, the renowned Black-Scholes option pricing model assumes that the underlying asset's volatility is constant, but this simplification does not match with the observed volatility in the real world. The same argument holds for the distributional assumptions of the underlying asset. (Probability distributions are assumed symmetrical in the model, whereas they can be asymmetrical in reality.)
2. The model may be applied correctly, but not in the way its developers intended.
3. Due to the errors in numerical solution, badly approximated answers may be obtained for a correctly formulated problem.
4. Estimations from certain historical data that the model requires may not satisfy the properties of the variable or parameter. For instance, there are a number of ways for obtaining volatility estimates such as finding the variance of historical data, implied volatility, and generalized autoregressive conditional heteroscedasticity (GARCH) methods. Each of these methods gives good estimations for different conditions, some of which have been examined in detail by Figlewski (1997). For example, while estimating historical volatility with daily data improves the accuracy for short-term horizons, it does not give good results for long-term horizons, estimation with monthly data being more suited (see Figlewski, 1997).

For every model, model risk is an inevitable consequence since each model is just a theoretical conceptualization of the real phenomena. However, model risk is more striking in derivative pricing due to the use of highly sophisticated mathematical techniques in this field. The Black-Scholes (1973) option pricing model is a base for almost all subsequent derivative pricing research. Nevertheless, the weaknesses and unrealistic assumptions of this model have been thoroughly discussed and reported in Macbeth and Meville (1979). Many papers such as Merton (1975), Hull and White (1987), Naik (1993), Amin and Ng (1993), Scott (1997), and Duan (1995), have developed models which take into account and deal with the shortcomings of the Black-Scholes model, such as constant volatility, constant interest rate, and lognormality assumptions. Merton (1975) questioned the continuity assumption in the Black-Scholes model and proposed a jump diffusion process for stocks returns. Hull and White (1987) modeled stock returns as a generalized Wiener process. Naik (1993) tried to incorporate discontinuity and changing volatility in the same model but with a different approach than Merton (1975) and Hull and White (1987). Amin and Ng (1993) tried to develop a more realistic model by incorporating stochastic volatility and interest rate. Scott (1997) took into account all of the deviations, stochastic variance, stochastic interest rate, and discontinuity topics, which were also addressed by previous researchers. Duan (1995), in his turn, added the *variable* variance by introducing the GARCH models. However, while researchers have tried to approximate the models to real dynamics of option prices, they have had to use more complex mathematical techniques which in turn increase the likelihood of being subject to model risk.

Inadequately used or applied models may give rise to many problems for those who take part in financial markets. As pointed out by Figlewski (1999), option writers may sell contracts at very low prices or investors may buy contracts at very high prices, which may lead to bad hedging strategies or cause market risk and credit risk measures to be significantly erroneous.

As we mentioned previously, model risk arises from different sources and exist at different levels. Therefore, one can examine a model and its risk from different perspectives. In this chapter, we mainly focus on the model risk arising from model specification. We aim to evaluate Duan's (1995) GARCH option pricing model for different parameter estimation methods and determine the sensitivity of the model to different parameter estimation procedures. The following section briefly introduces the GARCH option pricing model. The third and fourth sections describe the data and methodology for the analysis. The final section in this chapter presents the results and the last section concludes the study.

GARCH OPTION PRICING MODEL

GARCH type models have gained widespread acceptance in modeling volatility of stock returns in finance literature. Duan (1995) developed an option pricing model by using this strong econometric model of volatility.

According to GARCH in mean volatility model, stocks are assumed to follow the below dynamic:

$$\ln\frac{S_t}{S_{t-1}} = r + \lambda\sqrt{h_t} - \frac{1}{2}h_t + \varepsilon_t \tag{8.1}$$

$$\varepsilon_t \big| \phi_{t-1} \sim N(0, h_t) \tag{8.2}$$

$$h_t = \alpha_0 + \sum_{i=1}^{q}\alpha_i\varepsilon^2_{t-i} + \sum_{i=1}^{p}\beta_i h_{t-i} \tag{8.3}$$

where *ln* is the natural logarithm, S_t is the stock price at time t, λ is the risk premium, ε_t is a process with zero mean and unit variance, ϕ_t is the information set, and h_t is the conditional variance. To price options, Duan used the local risk-neutral valuation relationship that allows a valuation under changing variance. This is different than the traditional risk-neutral valuation which assumes a constant variance. The stock price dynamic under the locally risk-neutralized measure takes the following form:

$$\ln\frac{S_t}{S_{t-1}} = r - \frac{1}{2}h_t + \xi_t \tag{8.4}$$

$$\xi_t \big| \phi_{t-1} \sim N(0, h_t) \tag{8.5}$$

$$h_t = \alpha_0 + \sum_{i=1}^{q}\alpha_i(\xi_{t-i} - \lambda\sqrt{h_{t-i}})^2 + \sum_{i=1}^{p}\beta_i h_{t-i} \tag{8.6}$$

After the determination of the process of underlying asset with Equation (8.4), one can find the option price with strike K and maturity T as the discounted value of conditional expected value of the end of maturity payment (S_t − K) under the risk-neutral measure Q, mathematically denoted as follows:

$$C_t = e^{-r(T-t)}E^Q[\max(S_t - K, 0)] \tag{8.7}$$

Simulation methods are a natural tool for approximating this expectation. For the estimation of Equation (8.7), there are a variety of numerical methods detailed by Boyle, Broadie, and Glasserman (1997). However, a drawback of using simulation methods for option pricing is that the simulated prices may violate the rational option pricing bounds, hence may result in unreasonable price estimates. Therefore, we prefer the simulation procedure of the empirical martingale simulation developed by Duan and Simonato (1998) and also used in Duan and Zhang (2001). This method incorporates a simple correction to the standard procedure by ensuring that the simulated paths together are martingales empirically.

DATA

The empirical analysis in this study is based on the data that include quotes covering the S&P 500 European index options traded on the Chicago Board Options Exchange from August 22, 2007 to February 27, 2008. We choose S&P 500 index options for several reasons as many other researchers do (Hsieh and Ritchken, 2005; Rubinstein, 1994; Bakshi, Cao, and Chen, 1997; Bates, 1991; Dumas, Fleming and Whaley, 1998). These options are the most actively traded European-style contracts and provide a wide range of actively traded contracts with different strikes and maturities. Since there are a variety of contracts, it is possible to construct the prices of options with constant moneyness and maturity.

We construct weekly data sets using the official settlement prices on Wednesday. The general tendency in the literature is the use of Wednesday data to perform empirical analysis because Wednesday is the day of the week with the least coincidence with holidays. Although the results may ignore the seasonal dynamics like the day-of-the-week effect in stock returns (see Lakonishok and Smidt, 1988; and French, 1980), this allows for a comparative study. Hence, in line with the general tendency, we construct weekly data sets using the official settlement prices on Wednesday and end up with 28 sets, each set containing data for all the available maturities. Approximately 500 options data are used in the analysis. To avoid liquidity biases, we exclude from the sample all the options having less than 15 days to maturity. For the same purpose, we exclude deep-in-the-money and deep-out-of-the-money options since these are thinly traded and may distort the information on volatilities. As a result, we only include the options data within the 0.9 and 1.1 moneyness regions.

Table 8.1 details the number of call options contracts in our empirical analysis. It shows the groups and their usage in the analysis. As can be seen, the most prevailing group in moneyness dimension is at-the-money options, while in maturity dimension the options having more than 60 days to expiry are the most numerous.

METHODOLOGY

As mentioned in the previous section, we use sets of weekly cross-sectional option prices, i.e., the data of options with different strikes and maturities. Although it is possible to employ the time series of S&P 500 index data instead of option prices to determine the parameters of the model, we

Table 8.1 Number of S&P 500 Index Call Option Contracts by Moneyness and Maturity

Moneyness		Days-to-Expiration			Subtotal
K/S		<30	30–60	>60	
ITM	>1.1	30	43	33	(106)
ATM	0.97–1.02	60	78	98	(236)
OTM	<0.97	46	60	70	(176)
Subtotal		(136)	(181)	(201)	(518)

K, strike price; *S*, spot S&P 500 index level; ITM, in-the-money option; ATM, at-the-money option; OTM out-of-the-money option.

prefer cross-sectional option prices because this method takes account of the information content of the option prices complementing the time series of the underlying asset.

Parameter estimation requires an optimization procedure to discover the values of parameters from the data sets. In the literature, it is common to take the pricing error as the objective function for the corresponding option pricing model. For instance, Huang and Wu (2004) used the pricing error but with little difference by giving different weights to pairs of theoretical and market prices. Hsieh and Ritchken (2005) used the pricing error for their model as did Heston and Nandi (1997). Bakshi, Cao, and Chen (1997) examined the alternative option pricing models from different perspectives such as hedging and out-of-sample pricing. They also use the pricing error as the objective function for the parameter estimation. Jackwerth and Rubinstein (1996) derived the risk-neutral probabilities of S&P 500 European index options by using the sum of squared differences of prior from posterior probabilities, which follows the same logic as that of the pricing error. They also use *goodness-of-fit*, *absolute difference*, and *maximum entropy* functions as the objective function of the optimization problem.

These examples show that the pricing error function predominates over other objective functions such as absolute pricing error, percentage pricing error, and implied volatility error functions. Each objective function gives different weights to different error terms. The pricing error function punishes large errors in comparison to small errors due to the quadratic form, while the percentage pricing error function attributes a higher importance to small and high percentage errors. These differences lead to discrepancies in the optimization procedure. For instance, the pricing error may attribute too much importance to expensive options, e.g., in-the-money and long-term options, while giving less weight to cheap options, e.g., out-of-the-money and short-term options. On the other hand, the percentage pricing may give more weight to cheap options (out-of-the-money and short-term options) than expensive ones (in-the-money and long-term options).

In this study, we examine how the results of Duan (1995) GARCH option pricing model would differ when different objective functions are implemented in the optimization process. We perform an analysis with four different objective functions.

1. *The implied volatility error function:* Let $\sigma_{m.t}^{\text{mod}}$ denote the Black-Scholes implied volatility under GARCH option pricing model for the option with exercise price K_m and maturity T_t, and $\sigma_{m.t}^{mkt}$ denote the corresponding implied volatility based on the market price. Then, the objective function for the parameter estimation is

$$Objective1 = \min \left[\frac{\sum_{t=1}^{T} \sum_{m=1}^{M_t} (\sigma_{m.t}^{\text{mod}} - \sigma_{m.t}^{mkt})^2}{N} \right]$$

Here, T shows the number of maturities available at a certain week, and M_t represents the number of exercise prices for the t^{th} maturity. Consequently, $N = \sum_{t}^{T} M_t$ is the total number of options used in the optimization for a given week.

2. *The pricing error function:* Let $P_{m.t}^{\text{mod}}$ denote the price obtained by the GARCH option pricing model for the option with exercise price K_m and maturity T_t, and $P_{m.t}^{mkt}$ denote the market price. Then, the objective functions of pricing error, absolute pricing error, and percentage pricing error are as follows:

$$Objective2 = \min \left[\frac{\sum_{t=1}^{T} \sum_{m=1}^{M_t} (p_{m.t}^{\text{mod}} - p_{m.t}^{mkt})^2}{N} \right]$$

3. *The absolute pricing error function:*

$$Objective3 = \min \left[\frac{\sum_{t=1}^{T} \sum_{m=1}^{M_t} \left| (p_{m.t}^{\text{mod}} - p_{m.t}^{mkt}) \right|}{N} \right]$$

4. *The percentage pricing error function:*

$$Objective4 = \min \left[\frac{\sum_{t=1}^{T} \sum_{m=1}^{M_t} \left| (p_{m.t}^{\text{mod}} - p_{m.t}^{mkt}) / p_{m.t}^{mkt} \right|}{N} \right]$$

The parameters α_0, α_1, β_1, λ and h_t in Equation (8.6) are estimated by minimizing the above objective functions. We employ the quasi Newton-Raphson method with Golden Section Search and quadratic interpolation for the unconstrained optimization problem as in Duan and Zhang (2001). We begin the iteration with an initial guess of parameter values and then obtain the option prices by the empirical martingale simulation with 5,000 sample paths. Parameters are updated until the objective functions reach their minimum value. The simulation is seeded with the same random numbers to ensure the continuity of the optimization problem. Therefore, the optimized parameters and indirectly the performance of the GARCH model will show discrepancies. If the differences are significant, then users of the GARCH option pricing model are faced with a model risk due to the choice of the objective function. For evaluating the performance of the models, we compare them with the ad hoc version of the Black-Scholes model (used as a benchmark). For each set of data, we update the volatility parameter in the Black-Scholes model. Then, the relative performance is measured by the logarithm of the ratio of the root-mean-square errors (RMSE) that were obtained from the corresponding objective function $\ln(RMSE_{GARCH}/RMSE_{BS})$ where $RMSE_{GARCH}$denotes the root-mean-square error obtained from the GARCH model and $RMSE_{BS}$ denotes the root-mean-square error obtained from the Black-Scholes model. Negative (positive) values of this log ratio indicate the outperformance (underperformance) of the GARCH model and when this value is close to zero, both models perform the same.

EMPIRICAL ANALYSIS

This section presents the results of the empirical analysis we performed on the S&P 500 index options. Table 8.2 reports the RMSEs of the GARCH model and its benchmark (the Black-Scholes model) as well as the performance of the GARCH model in comparison with this benchmark separately for each objective function. A negative value in the Performance column signifies that the GARCH model outperforms the Black-Scholes model since the performance is measured by $\ln(RMSE_{GARCH}/RMSE_{BS})$.

Table 8.2 Root-Mean-Square Errors and Performance Ratios

OBJ. FN.	Implied Volatility Error			Pricing Error			Absolute Pricing Error			Percentage Pricing Error		
WEEK	$RMSE_{GARCH}$	$RMSE_{BS}$	Performance	$RMSE_{GARCH}$	$RMSE_{BS}$	Performance	$RMSE_{GARCH}$	$RMSE_{BS}$	Performance	$RMSE_{GARCH}$	$RMSE_{BS}$	Performance
1	0.02	0.03	−0.56	2.23	6.29	−1.03	1.80	2.23	−0.22	0.32	0.58	−0.60
2	0.01	0.03	−0.88	2.39	7.05	−1.08	1.58	2.29	−0.37	0.28	0.55	−0.66
3	0.01	0.03	−1.75	4.95	4.70	*0.05*	2.12	1.98	*0.07*	0.36	0.56	−0.45
4	0.01	0.03	−1.28	4.42	8.05	−0.60	2.12	2.39	−0.12	0.25	0.44	−0.57
5	0.03	0.05	−0.41	4.56	8.45	−0.62	1.88	2.47	−0.28	0.29	0.43	−0.38
6	0.02	0.03	−0.31	1.65	45.99	−3.33	1.47	5.07	−1.24	0.24	0.61	−0.93
7	0.01	0.04	−1.76	0.97	23.45	−3.19	1.36	3.23	−0.86	0.31	0.51	−0.48
8	0.01	0.02	−0.79	1.39	3.63	−0.96	1.25	1.67	−0.29	0.19	0.27	−0.35
9	0.01	0.02	−0.84	2.41	26.56	−2.40	1.42	3.09	−0.77	0.22	0.40	−0.61
10	0.04	0.10	−1.03	9.17	14.33	−0.45	2.43	2.75	−0.13	0.38	0.54	−0.34
11	0.01	0.03	−1.48	2.12	39.83	−2.93	1.42	4.02	−1.04	0.29	0.63	−0.76
12	0.01	0.05	−1.86	3.75	9.26	−0.90	1.72	2.62	−0.42	0.40	0.70	−0.57
13	0.02	0.05	−0.95	3.25	8.38	−0.95	2.00	2.46	−0.20	0.37	0.91	−0.90
14	0.03	0.04	−0.18	1.90	8.34	−1.48	2.52	2.45	*0.03*	0.37	0.53	−0.36
15	0.01	0.02	−0.63	2.48	3.94	−0.46	1.98	1.81	*0.09*	0.34	0.42	−0.20
16	0.02	0.02	−0.39	4.50	29.70	−1.89	1.90	3.11	−0.49	0.35	0.40	−0.13
17	0.04	0.06	−0.56	7.36	39.03	−1.67	2.53	4.35	−0.54	0.35	0.66	−0.64
18	0.27	0.44	−0.49	22.6	39.92	−0.57	3.71	5.03	−0.31	0.54	0.66	−0.21
19	0.01	0.04	−1.62	1.17	7.55	−1.87	1.19	2.35	−0.68	0.35	0.70	−0.68
20	0.03	0.05	−0.49	4.31	15.97	−1.31	2.05	3.09	−0.41	0.49	0.57	−0.15
21	0.01	0.02	−0.54	2.61	5.12	−0.67	1.52	1.99	−0.27	0.28	0.50	−0.60
22	0.02	0.02	−0.03	3.12	4.57	−0.38	4.26	4.99	−0.16	0.59	0.60	−0.02

(Continued)

Table 8.2 (Continued)

OBJ. FN.	Implied Volatility Error			Pricing Error			Absolute Pricing Error			Percentage Pricing Error		
WEEK	$RMSE_{GARCH}$	$RMSE_{BS}$	Performance	$RMSE_{GARCH}$	$RMSE_{BS}$	Performance	$RMSE_{GARCH}$	$RMSE_{BS}$	Performance	$RMSE_{GARCH}$	$RMSE_{BS}$	Performance
23	0.12	0.12	−0.05	15.5	39.08	−0.92	3.42	4.64	−0.31	0.46	0.81	−0.57
24	0.05	0.08	−0.51	6.70	11.73	−0.56	2.84	2.98	−0.05	0.69	0.84	−0.20
25	0.02	0.08	−1.43	2.14	16.87	−2.06	1.68	3.57	−0.75	0.92	1.46	−0.46
26	0.01	0.03	−1.05	2.49	4.71	−0.64	2.14	2.01	*0.06*	0.36	0.44	−0.20
27	0.01	0.02	−0.47	2.84	4.26	−0.41	2.22	1.85	*0.18*	0.30	0.46	−0.42
28	0.03	0.04	0.39	7.33	8.49	0.15	2.67	2.43	*0.10*	0.41	0.61	0.40

The table depicts for various objective functions the root-mean-square errors of the GARCH and the Black-Scholes models ($RMSE_{GARCH}$ and $RMSE_{BS}$) as well as the performance of the GARCH model relative to the Black-Scholes model. The Performance column represents the logarithm of the ratio of RMSEs, i.e. $\ln(RMSE_{GARCH}/RMSE_{BS})$. The calibration is carried out by minimizing the corresponding objective function. Positive values of Performance (which means the Black-Scholes model outperforms the GARCH model) are in italic.

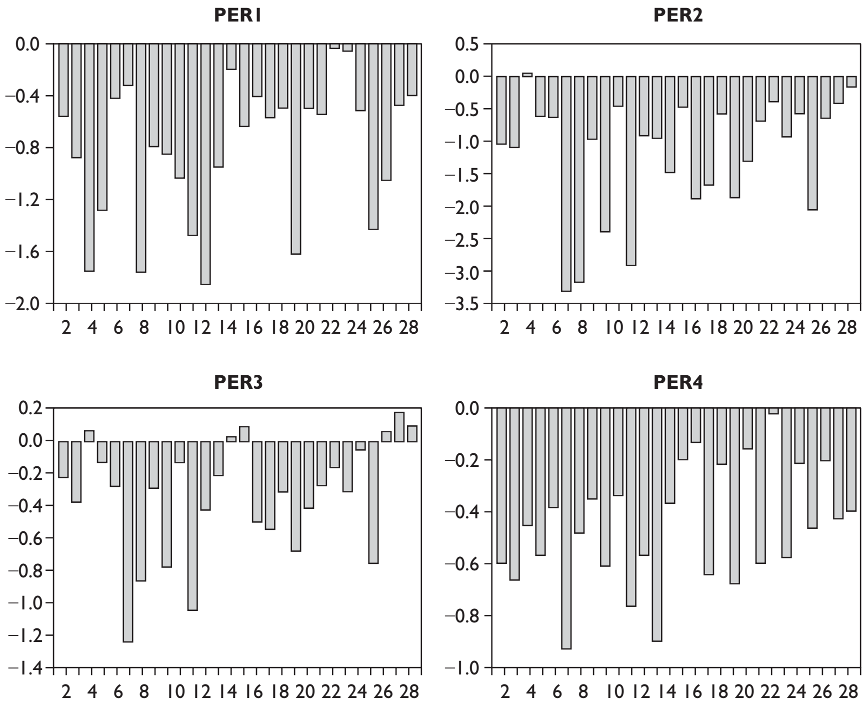
From Table 8.2, one can easily notice the superior performance of the GARCH option pricing model over Black-Scholes model, obvious by the dominance of negative values for each objective function. In fact, this was expected since GARCH uses more parameters than Black-Scholes. However, the focus of this study is not the comparison of these two models, but the effect of the different objective functions on the performance of the GARCH model in relation to its benchmark. Therefore, we analyze the Performance columns of the models with different objective functions.

From Table 8.2, the parameter estimations with implied volatility error and percentage pricing error have the same pattern. Thus the choice between these two objective functions makes no difference in terms of the number of weeks that GARCH outperforms the benchmark. In all 28 weeks, the GARCH model shows better performance relative to the benchmark. However, the parameter estimations with pricing error and absolute pricing error reveal some differences. While the GARCH model outperforms relative to the benchmark for all 28 weeks when parameterized by the implied volatility error and percentage pricing error functions, it underperforms for one week when parameterized by the pricing error function and for five weeks when parameterized by the absolute pricing error function.

The differences are more visible on the histograms in Figure 8.1 that gives the performance of the GARCH model relative to the benchmark separately for each objective function. In Figure 8.1, PER2 and PER3, which represent the relative performances of the GARCH model calibrated with pricing error and absolute pricing error, respectively, take both negative and positive values; while PER1 and PER4, which represent the relative performances of the GARCH model calibrated with implied volatility error and percentage pricing error, respectively, take only negative values. In this figure, negative values indicate the outperformance of the GARCH model and positive values indicate the outperformance of the benchmark.

In order to determine whether the discrepancies between the models with different objective functions are significant, we run t tests on the mean performances (see the last row of panel A in Table 8.3) in addition to descriptive statistics. As t test requires the data to be normally distributed and this is only rarely satisfied in financial time series, we additionally perform nonparametric tests (Table 8.3, panel B). Although one should interpret parametric test results with caution due to this eventual problem with t test, Tables 8.1 and 8.3 provide strong evidence about the dominance of negative values over positive ones. Therefore, we conveniently conclude that the outperformance of the GARCH model relative to its benchmark is independent from the choice of the objective function (Table 8.3, panel A). The Kruskal-Wallis test in panel B (Table 8.3) also shows that the average

Figure 8.1 Relative Performances



The figures depict the distribution of performances of the GARCH model relative to the Black-Sholes model (the benchmark). PER1 represents the performance of the GARCH model calibrated with the implied volatility error function; PER2, PER3, and PER4 represent the same for the pricing error, the absolute pricing error, and the percentage pricing error functions, respectively.

Table 8.3 Descriptive Statistics and Equality Tests for Performances

PANEL A. Descriptive Statistics

	Performance 1	Performance 2	Performance 3	Performance 4
Minimum	−1.86	−3.33	−1.24	−0.93
Maximum	−0.03	0.05	0.18	−0.02
Mean	−0.81*	−1.19*	−0.33*	−0.46*

PANEL B. Equality of Means and Variances

	Value	Probability
Kruskal-Wallis	29.04	0.0000
Barlett	50.96	0.0000
Levene	14.68	0.0000
Brown-Forsythe	7.93	0.0001

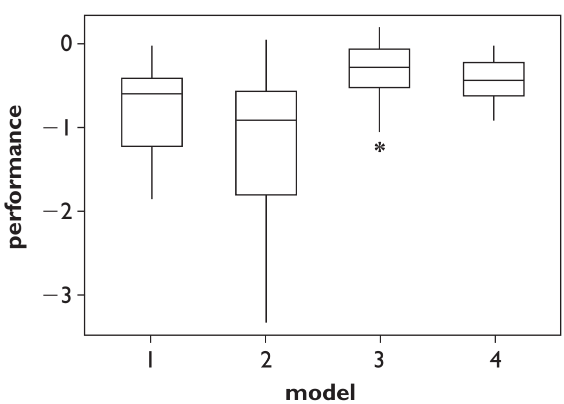
* The corresponding mean values are significantly different from zero at 99% level.

performances of the models with different objective functions are significantly different from each other at 99% level. That is, different objective functions affect the performance of the GARCH option pricing model, although it always outperforms the benchmark. Barlett, Levene, and Brown-Forsythe tests results reveal that the variances of the models with different objective functions are also significantly different. This implies that for some models the relative performance fluctuates in a narrow band and for some others in a wide range.

For a more detailed examination, we take the advantage of a box plot of performance estimations.

With the help of Figure 8.2, one can make a deeper insight into the discrepancies originated from the use of different parameter estimation procedures. In what follows, we can deduce that the use of pricing error function improves the performance of the GARCH model relative to the benchmark. In other words, on average, the GARCH option pricing model shows its best performance relative to the benchmark when parameterized by the pricing error function. However, when absolute error function is used, the performance of the GARCH model becomes close to the benchmark. Another difference can be observed in the spreads of the models. Pricing error function again differs from the others strikingly. It has a very large spread changing approximately between 0 and −3.5, which means that compared with the parameter estimations with other error functions, the parameter estimation with the pricing error function results in a more varying degree of outperformance of the GARCH model. Hence, the low

Figure 8.2 The Box Plot of the Performance Estimation for the Models with Different Objective Functions



The numbers in the x-axis represent the models. 1, 2, 3, and 4 are for the model with implied volatility error, pricing error, absolute pricing error, and percentage pricing error functions, respectively.

and negative values of this spread makes the GARCH model more appropriate compared with the benchmark. The second objective function that most improves the performance of the GARCH model is the implied volatility error function. Absolute pricing error and percentage pricing error functions are relatively more consistent in their context. That is, the performances of the option pricing models characterized by these two objective functions do not change significantly. In general, it is not wrong to say that the success of the model depends on the choice of the objective function in the optimization procedure.

CONCLUSION

The aim of this study is to highlight the sensitivity of mathematically sophisticated option pricing models to different parameter estimation methods. We use four different objective functions to estimate the parameters in the model and try to determine whether they affect the model performance. Our fundamental finding is that different parameter estimation procedures affect the model performance and obviously this brings about a risk for the model users. For some objective functions, the outperformance of the GARCH option pricing model is beyond dispute for all 28 sets of weekly data, while for some others, the benchmark outperforms the GARCH model in some weeks. Changing the objective function in the optimization procedure affects the performance of the GARCH model in terms of both mean level and variance. This implies that these models price options with varying success rates and the spread of the success rates may be large depending on the choice of objective function. In summary, changes made in the objective function that is used in the parameter estimation significantly affects both the performance of the GARCH option pricing model and its performance relative to the benchmark.

REFERENCES

Amin, K.I. and Ng, V.K. (1993) Option Valuation with Systematic Stochastic Volatility. *Journal of Finance*, 48(3): 881–910.

Bakshi, G., Cao, C., and Chen, Z. (1997) Empirical Performance of Alternative Option Pricing Models. *Journal of Finance*, 52(5): 2003–2049.

Bates, D. (1991) The Crash of 87: Was It Expected? The Evidence from Options Markets. *Journal of Finance*, 46(3): 1009–1044.

Black, F. and Scholes, M. (1973) The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81(3): 637–654.

Boyle, P., Broadie, M., and P. Glasserman. (1997) Monte Carlo Methods for Security Pricing. *Journal of Dynamics and Control*, 21(8–9): 1267–1321.

Derman, E. (1996) Model Risk. Quantitative Strategies Research Notes, Goldman Sachs.

Duan, J. (1995) The Garch Option Pricing Model. *Mathematical Finance*, 5(1): 13–32.

Duan, J. and Simonato, J. (1998) Empirical Martingale Simulation for Asset Prices. *Management Science*, 44(9): 1218–1233.

Duan, J. and Zhang, H. (2001) Pricing Hang Seng Index Options Around the Asian Financial Crisis—A GARCH Approach. *Journal of Banking & Finance*, 25(11): 1989–2014.

Dumas, B., Fleming, J., and R. Whaley. (1998) Implied Volatility Function: Empirical Tests. *Journal of Finance*, 53(6): 2059–2106.

Figlewski, S. (1997) Forecasting Volatility. *Financial Markets, Institutions and Instruments*, 6(1): 1–88.

Figlewski, S. (1998) Derivative Risks, Old and New. Department of Finance, Working Paper Series, NYU Stern, FIN–98–033.

Figlewski, S. (1999) Market Risk and Model Risk for a Financial Institution Writing Options. *Journal of Finance*, 54(4): 1465–1499.

French, K.R. (1980) Stock Returns and Weekend Effect. *Journal of Financial Economics*, 8(1): 55–70.

Heston, S.L. and Nandi, S. (1997) A Closed-Form GARCH Option Pricing Model. Working Paper 97–9, Federal Reserve Bank of Atlanta.

Hsieh, K.C. and Ritchken, P. (2005) An Empirical Comparison of GARCH Option Pricing Models. *Review of Derivatives Research*, 8(3): 129–150.

Huang, J.-Z. and Wu, L. (2004) Specification Analysis of Option Pricing Models Based on Time-Changed Levy Processes. *Journal of Finance*, 59(3): 1405–1439.

Hull, J.C. and White, A. (1987) The Pricing of Options on Assets with Stochastic Volatilities. *Journal of Finance*, 42(2): 281–300.

Jackwerth, J.C. and Rubinstein, M. (1996) Recovering Probability Distributions from Option Prices. *Journal of Finance*, 51(5): 1611–1631.

Lakonishok, J. and Smidt, S. (1988) Are Seasonal Anomalies Real? A Ninety-Year Perspective. *Review of Financial Studies*, 1(4): 403–425.

Macbeth, J.D. and Merville, L.J. (1979) An Emprical Examination of Black-Scholes Call Option Pricing Model. *Journal of Finance*, 34(5): 1173–1186.

Merton, R.C. (1975) Option Pricing When Underlying Stock Returns Are Discontinuous. *Journal of Financial Economics*, 3(1–2): 125–44.

Naik, V. (1993) Option Valuation and Hedging Strategies With Jumps In The Volatility of Asset Returns. *Journal of Finance*, 48(5): 1969–1984.

Rubinstein, M. (1994) Implied Binomial Trees. *Journal of Finance*, 49(3): 771–818.

Scott, L.O. (1997) Pricing Stock Options in a Jump-Diffusion Model with Stochastic Volatility and Interest Rates: Applications of Fourier Inversion Methods. *Mathematical Finance*, 7(4): 413–424.

CHAPTER 9

EFFECT OF BENCHMARK MISSPECIFICATION ON RISK-ADJUSTED PERFORMANCE MEASURES

Laurent Bodson and Georges Hübner

ABSTRACT

This chapter examines the existence, magnitude, and significance of benchmark misspecification effect on risk-adjusted performance measures. We analyze three multifactor performance measures, namely, Jensen's alpha, the information ratio (IR) and we propose a generalization of the traditional Modigliani-Modigliani measure (noted GM^2). We focus our analysis on the monthly returns of 5,012 distinct mutual funds from January 1996 to December 2006. Our results reveal that the GM^2 measure is more stable and more persistent to model specification changes than IR and even more compared with Jensen's alpha.

INTRODUCTION

In the funds industry, a large literature is devoted to the different performance measures. Since a portfolio manager can increase the expected return by increasing the systematic risk of the portfolio, the assessment of performance must integrate the notion of risk. Therefore, a traditional performance reporting generally computes various risk-adjusted performance measures that assess the reward with some adjustments for risks. Currently, these risk-adjusted metrics are generally used to define portfolio allocations and performance fees. These measures do not treat the risk (and consequently the adjustment applied to the reward) in the same way, but they all have the same objective of performance assessment. For these reasons, the stability and persistence of such measures is fundamental.

Nowadays, the majority of the professionals are using performance measures that are adjusted to a single risk factor. However, there is no doubt that a unique risk factor is not sufficient to capture the variety of systematic risk sources. Therefore, we decide to focus our analysis on risk-adjusted performance measures designed for several risk factors.

We limit our analysis to three performance measures: namely Jensen's alpha (α), the information ratio (**IR**), and we propose a generalized form of the Modigliani-Modigliani measure that we note ($\mathbf{GM}^2$). We study these measures because they are designed for a multifactor market model and they are the most famous metrics used by practitioners.

We explore in this chapter the impact of model specification changes on the three performance measures values and rankings. Indeed, model validation plays a key role in finance modeling and influences in different ways the behavior of portfolio managers. In that way, we evaluate the model risk linked to the benchmark selection.

This chapter has two main objectives. First, we analyze the empirical stability of the three measures between different model specifications. Second, we assess the persistence of the rankings from one model specification to another. Our analysis is then restricted to the study of the impact of model specification (or benchmark (mis)specification) on the performance measures (values or rankings) over a given period of time. We do not assess the sensitivity of the measures to other parameters such as the length of the computation window or the return frequency used. Furthermore, we do not analyze the stability or persistence through time.

The empirical part of this chapter is based on a large sample of 5,012 mutual funds from the database of the Center for Research in Security Prices (CRSP). The period of time covered begins in January 1996 and finishes in December 2006.

The chapter is organized as follows. The second section in this chapter depicts the risk-adjusted performance measures used in our study. In this chapter's third section, we present the empirical method and the data sources. We report in the fourth section the results of the different tests and correlation analysis realized between the series of values or the series of rankings. The last section provides conclusive remarks and suggestions for future research and practices.

DESCRIPTION OF THE RISK-ADJUSTED PERFORMANCE MEASURES

We present in this section the three risk-adjusted performance measures that we compare in the empirical part of the chapter. Their common particularity is that they are all designed for multifactor models.

We assume that the traditional Fama and French (1993) market model resulting from the following empirical regression (for all t of the analyzed period) with n risk factors is confirmed: $R_{it} = \alpha_i + \sum_{j=1}^{n} \beta_{ij} * F_{jt} + e_{it}$, where R_{it} is the return of the fund i in excess of the risk-free rate at time t, α_i is the intercept of the linear regression for the fund i, β_{ij} is the risk loading related to the risk factor j for the fund i, F_{jt} is the premium (i.e., expressed in excess of the risk-free rate) of the risk factor j at time t and e_{it} is the residual of the regression for the fund i at time t.

We assume no constraint on the risk loadings and the optimal values are obtained minimizing the sum of the squared residual terms (i.e., the method of ordinary least squares).

Our first performance measure is Jensen's alpha (1967), noted $\boldsymbol{\alpha}$, defined as the intercept of the empirical capital asset pricing model (CAPM) developed by Sharpe in 1964. The Jensen measure represents the excess return that a fund/portfolio generates over its expected return (defined by the benchmark portfolio return, i.e., the systematic risk factors).

Our second performance measure is the information ratio (***IR***) computed as the alpha divided by the standard deviation of the empirical regression residual terms: $IR_i = \frac{\alpha_i}{\sigma(e_{i.})}$

In other words, the ***IR*** is the expected active return (the difference between the fund return and its benchmark return) divided by the tracking error. This metric detects the manager-specific skills and autonomy. The higher the ***IR*** the better the manager is. Note that the Sharpe ratio (1964 and 1966) is similar to the information ratio where there is a unique risk factor, the risk-free asset. Furthermore, a strong but tenable assumption is that the specific risk of the security i is assessed by the standard deviation of the error term series.

The third performance measure that we propose is a generalization of the Modigliani-Modigliani measure (***GM²***) developed in 1997. Indeed, we are working in a multifactor market model framework which requires an adaptation of the $\boldsymbol{M^2}$ measure. The traditional $\boldsymbol{M^2}$ measure is computed multiplying the fund's average excess return by the quotient of the standard deviation of the index excess return by the standard deviation of the fund's excess return. Knowing that we do not have a single index, we need to replace the one-factor index by the estimated composite index series $\sum_{j=1}^{n} \beta_{ij} * F_{jt}$ for each value of t. We have therefore the following formulation of our generalization of the Modigliani-Modigliani measure:

$$GM^2_i = E[R_{i.}] * \frac{\sigma\left(\sum_{j=1}^{n} \beta_{ij} * F_{j.}\right)}{\sigma(R_{i.})}$$

, where $R_i.$ is the return series of the fund i in excess of the risk-free rate, β_{ij} is the risk loading related to the risk factor j for the fund i, $F_j.$ is the risk premium series (i.e. expressed in excess of the risk-free rate) of the risk factor j and σ is the standard deviation of the series between the brackets.

Thus, this measure expresses the expected return of the fund weighted by its relative risk compared with the (composite) index risk. We can note that the risk is considered here to be the standard deviation of the returns and does not take into consideration the higher order moments of the statistical distribution (notably the skewness and the kurtosis). Therefore, this measure is more relevant to investors who invest in normally distributed financial assets (Scholz and Wilkens, 2005).

METHODOLOGY AND DATA SOURCES

Our CRSP mutual fund database covers the monthly returns of 5,012 mutual funds over the period from January 1996 to December 2006 (a total of 132 months). We have expressed all mutual funds returns in excess of the risk-free rate (using the risk-free series provided on Kenneth R. French's Web site, http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/).

To test the impact of model specification changes, we choose eight different common risk factors and we test all potential combinations of ***n*** factors selected among this set of regressors. In this way, we are composing different sets of benchmarks to compare the sensitivity of each performance measure to a model specification change.

The eight risk factors considered are the following: the traditional Fama and French three factors (RMRF, SMB, and HML) (1993), the Carhart momentum effect (UMD) (1997) (see Kenneth R. French's Web site), a liquidity risk premium (LIQ), and three multimoment premiums (COV, SKE, and KUR).

Hereafter, we look at each potential combination of factors for a given number ***n*** of selected factors (***n*** going from one to seven). Obviously, we do not consider ***n*** equals to 8 because there is, in this specific case, only one possible combination of factors and no model specification change possibility.

For a given number of factors (from one to seven over eight available alternatives), for each possible factor combination, we compute the values of the three analyzed performance measures and the corresponding series of rankings between the 5,012 mutual funds. We study these series of values and rakings two by two by calculating the rank correlations for each performance measure and each number of selected factors.

EMPIRICAL RESULTS

On the one hand, the sensitivity between the series of values (i.e., between two different model specifications) is measured by the Spearman correlation which is a nonparametric correlation measure more reliable than the Pearson correlation coefficient (which is linear and more sensitive to outliers). On the other hand, we use the Kendall correlation to compute the dependence between the series of rankings. Indeed, Kendall correlation is also a nonparametric statistic but specially designed to compute the degree of correspondence between two series of rankings.

We present in Table 9.1 the number of model specifications, the mean R^2, and the mean adjusted R^2 for each number of selected factors (from one to seven factors selected among the eight factors proposed). The number of model specifications is the number of **n** possible combinations of elements taken once from eight elements. The mean (adjusted) R^2 is computed as the average (adjusted) R^2 of each combination of model specification and mutual fund.

Table 9.1 shows simply that the fitting is improving when the number of selected factors increases (even if the R^2 is adjusted). It is important to note that the average (adjusted or not) R^2 is quite poor when there is only one risk factor.

First, we report below some correlation statistics (almost all significant at the 1 percent level) for the different performance measures without considering the model specification individually. In other terms, we aggregate the information without comparing the performance measures for an identical model specification change.

Table 9.1 Number of Model Specifications, Mean R^2, and Mean Adjusted R^2 for Each Number of Selected Factors

	Number of Selected Factors						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Number of Model Specifications	8	28	56	70	56	28	8
Mean R^2	0.1004	0.1787	0.2403	0.2905	0.3331	0.3707	0.4047
Mean-adjusted R^2	0.0935	0.1660	0.2225	0.2682	0.3067	0.3405	0.3711

We compute in Table 9.2 the mean correlations between the values of each performance measure for all the potential combinations of factors. We observe that ***GM²*** mean correlations between the series of values are always strictly higher than the mean correlations of the two other performance measures. More precisely, the difference tends to be higher when the number of different model specifications is high (i.e., when the model misspecification likelihood is high). For each number of selected factors, the alpha is always the poorest measure in terms of stability to model specification changes.

We also calculate in Table 9.2 the mean correlations between the series of rankings of each performance measure for all the potential combinations of factors. These results confirm, the ones obtained with the correlations between the values, that dominates the two other measures providing a better persistence to model specification changes. Again, quite intuitively, the difference between the **GM²** and the two other measures is emphasized for high number of model specification possibilities. Except for a unique selected factor, the alpha is the worst persistent measure.

If we look at the standard deviation of the correlation coefficients (between the series of values or the series of rankings), we note that, beyond the higher mean correlations, there is also clearly less disparity for the ***GM²*** stability and persistence around its mean value. In each scenario, i.e., each number of selected risk factors, the standard deviation of the ***GM²*** correlations is lower than the standard deviation of the ***IR***, which is itself lower than the standard deviation of the α.

Second, we compare the correlation statistics (almost all significant at the 1 percent level) between the series (of values or rankings) for an identical change in the model instead of averaging the values.

Table 9.3 presents the ranking between the three measures for each correlation computed between two series of values (given the model specification change). More precisely, for a given model specification change, we compute the correlation between the series of values for each performance measure. Next, we simply rank the three computed correlations to define the highest, the middle, and the lowest correlations between the three measures. We count the number of times that a given performance measure is the highest, the middle, and the lowest correlation, and we convert these figures in percentages.

We note that, except for the case where a unique factor is chosen, the ***GM²*** overperformed the two other measures exhibiting a higher stability (correlation between the series of values) for an identical model specification change. Globally, ***IR*** is at the second place and the alpha is the worst performer in terms of stability to model changes.

Table 9.2 Mean and Standard Deviation of the Correlation Metrics for Each Number of Selected Factors Considering All Possible Model Specification Changes

	Number of Selected Factors 1	2	3	4	5	6	7
Mean Correlation (SD)							
Alpha	0.8145	0.7068	0.6205	0.5654	0.5550	0.6042	0.7215
	(0.2299)	(0.2636)	(0.2893)	(0.2963)	(0.2897)	(0.2783)	(0.2586)
IR	0.8463	0.7901	0.7567	0.7476	0.7667	0.8152	0.8866
	(0.1485)	(0.1826)	(0.1932)	(0.1887)	(0.1742)	(0.1474)	(0.0976)
GM^2	0.8980	0.9743	0.9870	0.9918	0.9946	0.9965	0.9981
	(0.0593)	(0.0164)	(0.0076)	(0.0051)	(0.0039)	(0.0031)	(0.0024)
Mean Correlation (SD)							
Alpha ranking	0.7383	0.6372	0.5640	0.5169	0.5038	0.5395	0.6409
	(0.2094)	(0.2276)	(0.2434)	(0.2480)	(0.2418)	(0.2342)	(0.2287)
IR ranking	0.7230	0.6707	0.6388	0.6281	0.6435	0.6900	0.7701
	(0.1579)	(0.1854)	(0.1934)	(0.1885)	(0.1754)	(0.1562)	(0.1289)
GM^2 ranking	0.7497	0.8812	0.9183	0.9377	0.9514	0.9634	0.9758
	(0.0781)	(0.0410)	(0.0279)	(0.0235)	(0.0217)	(0.0204)	(0.0182)

SD, standard deviation; IR, information ratio; GM^2, generalized Modigliani-Modigliani measure.

Table 9.3 Comparison of the Correlation Between the Series of Values for an Identical Model Specification Change

	Number of Selected Factors 1	2	3	4	5	6	7
% Highest Correlation							
Alpha	42.86	7.67	2.27	1.16	0.19	0	0
IR	17.86	7.67	2.27	0.87	0.45	0.26	0
GM^2	39.29	84.66	95.45	97.97	99.35	99.74	1
% Median Correlation							
Alpha	25.00	33.33	23.64	16.11	10.58	5.03	0
IR	57.14	57.14	73.96	82.86	89.03	94.71	1
GM^2	17.86	9.52	2.40	1.04	0.39	0.26	0
% Lowest Correlation							
Alpha	32.14	58.99	74.09	82.73	89.22	94.97	1
IR	25.00	35.19	23.77	16.27	10.52	5.03	0
GM^2	42.86	5.82	2.14	0.99	0.26	0	0

IR, information ratio; GM^2, generalized Modigliani-Modigliani measure.

We compute exactly the same figures for the series of rankings in Table 9.4. We arrive at equivalent conclusions for the persistence of each performance measure. The persistence of the ***GM***^{2} is clearly higher than the persistence of the two other measures (excluding, again, the case of a single-factor model).

Thus, the impact of a model specification change is generally lower for the ***GM***^{2} than for the two other measures analyzed. In other words, a misspecification of the model used to compute the performance measure has a lower impact on the stability and the persistence of the ***GM***^{2} compared with the two other frequently used measures.

From this set of evidence, we can definitely conclude that the new performance measure derived from the original **M**^{2} measure outperforms the more traditional metrics adopted by financial practice, namely the alpha and the information ratio. In the context of model uncertainty regarding the actual systematic risk exposures adopted by active portfolio managers, it is important to be able to trust a performance measure, and it appears that the traditional view of computing the alpha generates significant instability of rankings across asset pricing specifications.

The **GM**^{2} cumulates the advantages of yielding an interpretation of abnormal return, just as the original measure and the alpha, while providing the level of confidence that is sought for in parametric performance evaluation.

Table 9.4 Comparison of the Correlation Between the Series of Rankings for an Identical Model Specification Change

	Number of Selected Factors						
	1	**2**	**3**	**4**	**5**	**6**	**7**
% Highest Correlation							
Alpha ranking	57.14	10.05	3.25	1.78	0. 32	0	0
IR ranking	7.14	6.61	1.36	0. 41	0. 19	0	0
GM^2 ranking	35.71	83.33	95.39	97.81	99.48	1	1
% Median Correlation							
Alpha ranking	14.29	36.24	29.55	21.04	14.29	6.88	0
IR ranking	67.86	52.91	67.99	77.68	85.45	93.12	1
GM^2 ranking	17.86	10.85	2.47	1.28	0.26	0	0
% Lowest Correlation							
Alpha ranking	28.57	53.70	67.21	77.18	85.39	93.12	1
IR ranking	25.00	40.48	30.65	21.90	14.35	6.88	0
GM^2 ranking	46.43	5.82	2.14	0.91	0.26	0	0

IR information ratio; GM^2, generalized Modigliani-Modigliani measure.

CONCLUSION

Although there exist plenty of performance measures (Cogneau and Hübner, 2009a, 2009b, report more than 100 different ways to measure performance), to date only the alpha is really popular in the context of multifactor models. Our findings indicate that the choice of such a way to measure the skills of portfolio managers is significantly prone to the modeling choice for the measurement of required returns. Such an exposure to specification risk can prove to be particularly delicate when the remuneration and, possibly, the hiring (or firing) decision of a manager is based on such a metric.

Of course, one may try to refine the asset pricing model itself in order to come up with the highest reliability of the regression constant, and thus of the performance attributed to the manager. The perspective taken in this paper is different, as we recognize that such a perfect specification does not always exist. Rather, we identify that the other popular absolute performance measure, the $\mathbf{M^2}$, has not really been exploited in a multifactor context. Obviously, our results indicate that there would be significant value added in considering this new, yet very simple, indicator in the context of performance evaluation. Of course, much remain to be done to refine the measurement and interpretation of such a measure in a multifactor context, but we view this chapter as a step in this direction.

REFERENCES

Carhart, M.M. (1997) On persistence in Mutual Fund Performance. *Journal of Finance*, 52(1): 57–82.

Cogneau, P. and Hübner, G. (2009a) The (more than) 100 Ways to Measure Portfolio Performance. Part 1: Standardized Risk-Adjusted Measures. *Journal of Performance Measurement*, forthcoming.

Cogneau, P. and Hübner, G. (2009b) The (more than) 100 Ways to Measure Portfolio Performance. Part 2: Special Measures and Comparison. *Journal of Performance Measurement*, forthcoming.

Fama, E.F. and French, K. (1993) Common Risk Factors in the Returns of Stocks and Bonds. *Journal of Financial Economics*, 33(1): 3–56.

Jensen, M.J. (1967) The Performance of Mutual Funds in the Period 1945–1964. *Journal of Finance*, 23(2): 389–416.

Modigliani, F. and Modigliani, L. (1997) Risk-Adjusted Performance. *The Journal of Portfolio Management*, 23(2): 45–54.

Scholz, H. and Wilkens, M. (2005) A Jigsaw Puzzle of Basic Risk-adjusted Performance Measures. *Journal of Performance Measurement*, 9(3): 57–64.

Sharpe, W.F. (1964) Capital Asset Prices: A Theory for Market Equilibrium Under Conditions of Risk. *Journal of Finance*, 19(3): 425–442.
Sharpe, W.F. (1966) Mutual Fund Performance. *Journal of Business*, 39(1): 119–138.

PART III

MODEL RISK RELATED TO CREDIT AND CREDIT DERIVATIVES INVESTMENTS

This page intentionally left blank

CHAPTER 10

CARRY TRADE STRATEGIES AND THE INFORMATION CONTENT OF CREDIT DEFAULT SWAPS

Raphael W. Lam and Marco Rossi[1]

ABSTRACT

This chapter analyzes the relationship between the term structure of sovereign credit default swaps (CDS) spreads and the carry trade in foreign exchange markets. The term structure of sovereign CDS spreads proxies sovereign default probability and recovery rate on a high-frequency basis. These variables yield important information on exchange rate risk, and therefore help predict returns and inform trading activities in the carry trade. The chapter uses daily data on sovereign CDS and exchange rates and documents the extent that sovereign CDS are useful in predicting carry trade activities and returns.

INTRODUCTION

A currency carry trade refers to the strategy of borrowing low-yielding currencies to purchase high-yielding currencies, at the risk to have to pay back the borrowed currencies at higher cost if exchange rates move adversely. In theory, the uncovered interest rate parity would imply that carry trades should not yield predictable excess returns as exchange rates would be expected to move to offset interest rate differentials. If markets were efficient and investors risk-neutral,[2] forward rates would be unbiased forecasts of future spot rates, and, therefore, a strategy that would, for a given interest rate differential, sell forward currencies that are at a forward premium and buy forward currencies that are at a forward discount would not systematically produce profits.

The vast amount of empirical literature has rarely supported the theoretically appealing purchasing power parity and market efficiency under rational expectations hypotheses.[3] The failure of the forward exchange rate to be an unbiased forecast of the future spot rate leave scope for effecting profitable carry trade strategies. While considering also the role of fundamentals, this chapter focuses on one possible factor behind short-term exchange rate movements, potentially explaining the forward exchange rate puzzle: credit risk.

Credit default swaps (CDS) are credit protection contracts that require one party, in exchange for a periodic premium, to make a contingent payment to another party in the case of a defined credit event. Sovereign CDS are financial contracts whose payoffs are linked to changes in the credit quality of underlying sovereign securities. Under certain conditions, the premium on CDS (in basis points) is closely correlated to the spread on a sovereign bond of the same maturity, reflecting a fairly close cross-sectional relationship with credit risk of the underlying security as measured by credit rating agencies. The term structure of sovereign CDS spreads reflects the probability of defaults and possible recovery rates on sovereign securities (Pan and Singleton, 2008).

According to Packer and Suthiphongchai (2003) and Zhu (2004), the CDS market has grown more than double both in number of trades and quotes between 1997 and 2003, of which the sovereign CDS accounts for about 15 percent of the total market capitalization. The fast-developing market for sovereign CDS provides a unique opportunity to investigate investors' risk-neutral probabilities of a credit event. In addition, unlike sovereign bonds, CDS contracts are standardized across maturities, which has contributed to a surge in sovereign CDS trading to the point that these derivatives have become far more liquid than the underlying sovereign bonds for a wide range of maturities. CDS spreads are therefore useful proxies to infer a full term structure of default probabilities.

This chapter analyzes the relationship between the term structure of sovereign credit default swaps (CDS) spreads and the carry trade in foreign exchange markets. The term structure of sovereign CDS spreads proxies sovereign default probability and recovery rate on a high-frequency basis. These variables yield important information on exchange rate risk, and therefore help predict returns and inform trading activities in carry trade.

The chapter uses daily data on sovereign CDS and exchange rates and documents the extent to which sovereign CDS are useful in predicting carry trade activities and returns.

The chapter is organized as follows. The second section presents the carry trade strategy and introduces CDS spreads as factors determining carry trade payoffs. The third section discusses the data and data sources; and the fourth introduces the empirical methodology to estimate carry trade returns controlling for fundamentals and, more specifically, credit risks. This chapter's fifth section reports on the empirical findings; its final section concludes.

THE CARRY TRADE STRATEGY

Burnside, Eichenbaum, and Rebelo (2007) report three main empirical findings with regard to a carry trade strategy: (1) it generates payoffs that are on average large and uncorrelated with traditional risk factors (for instance, the excess risk-adjusted return on carry trade strategy is fairly high but uncorrelated with the returns on US stocks); (2) the Sharpe ratio associated with carry trade substantially increases when emerging markets currencies are included in the portfolio; and (3) the bid-ask spreads are two to four times larger in emerging markets than in developed countries and that large positive Sharpe ratios emerge only if these bid-ask spreads are taken into account. These puzzling results suggest that carry trade strategies generate large excess returns even after market risk is factored in.

We use a carry trade strategy that exploits the forward premium and discount, while incorporating bid-ask spreads, as in Burnside et al (2006). Let S_a and S_b denote ask and bid spot exchange rates, F_a and F_b denote ask and bid forward exchange rates at time t for forward contract maturing at time $t + 1$. All exchange rates represent foreign currency units per U.S. dollar. The strategy involves selling X dollars (normalized to 1) forward according to the rule:

$$X_t = \begin{cases} 1 \text{ if } F_t^b / S_t^a > 1 \\ -1 \text{ if } F_t^a / S_t^b < 1 \\ 0 \text{ otherwise} \end{cases}$$

This means that the investor will have a short position (selling forward) if the bid forward rate is at a discount compared to the ask spot rate and a long position if ask forward rate is at a premium compared to the bid spot rate. The investor will have neutral position if the bid forward rate is neither at discount or at premium. The strategy involves settling existing open positions and taking new open positions at time t. The strategy is optimal under risk-neutrality with respect to nominal payoffs and under the assumption that exchange rates are martingales. Also, no collateralization requirement for the short position is assumed. The carry trade strategy takes a long/short position in the same currency, although it would be possible to use cross-currencies bid/ask forward and spot rates. The strategy will yield the following payoff:

$$R_{t+1} = \begin{cases} X_t \,(F_t^b / S_t^a - 1) \text{ if } X_t > 0 \\ X_t \,(F_t^a / S_t^b - 1) \text{ if } X_t < 0 \\ 0 \quad \text{if } X_t = 0 \end{cases}$$

This strategy takes into account the potential transaction cost involved in carry trade strategies. The carry trade return (R_t+1) also incorporates a liquidity cost as reflected in bid-ask spreads. The accumulated return of the carry trade strategy is simply the sum of R_t+1 over a certain period. The returns are not normally distributed but concentrated on the lower end, implying that carry trade returns are generally small but positive.

Carry trade payoffs are clearly linked to exchange rate movements, with the latter reflecting, inter alia, investors' perception of a country's fundamentals and, ultimately, the country's credit risk.[4] The term structure of risk would provide useful information about the likelihood that a certain carry trade strategy yield excess returns. As a proxy for credit risk and potential recovery rate in a certain country across maturities, we look at the term structure of sovereign CDS spreads. The existing evidence indicates that sovereign CDS spreads at different maturities tend to move closely together and that sovereign CDS spreads across emerging markets are highly correlated (Chan-Lau and Kim, 2004).[5]

We estimate the term structure of CDS spreads on a rolling basis with a two-month window for each country in the sample. This provides parametric estimates of the slope and curvature of the term structure of the CDS spreads. The intercept of the term structure would correspond to the short maturity of CDS spreads. In our case, it refers to the one-year sovereign CDS spreads. Most countries exhibit a positive term structure on CDS spreads in most periods, with the exception of the period after mid–2008, where a few crisis countries have inverted U-shaped term structure, i.e., one-year CDS spreads remarkably higher than five-year spreads, which, in turn, are lower than 10-year spreads.

THE DATA

Data on exchange rates across countries are obtained from Datastream, covering the period from January 2001 to April 2009 on a daily and weekly basis. Reliable data are available for a selection of 46 countries. The data structure is an unbalanced panel as countries are included as data become available. The original source of data is the Reuters system (daily quotes at closing). The exchange rates include spot and forward rates (one week and one month) with both bid and ask rates. Each exchange rate is quoted as foreign currency (C_f) per U.S. dollar. Daily data are converted to weekly data.

Data on sovereign CDS are provided by CMA, available from Datastream, one of the major trading platforms for credit derivatives. Data are available for 46 countries. It contains various quotes for sovereign CDS spreads. Our sample consists of daily quotes for CDS contracts with maturities of 1, 2, 3, 5, and 10 years. The sample covers the period January 2001 to April 2009. Throughout the sample period, countries' sovereign rating would vary as would the corresponding CDS spreads.

Stock market returns and relative returns on financial and commodity sector assets are obtained from the stock exchanges in individual countries. Stock market returns are based on percentage changes in aggregate indexes. Relative sector returns are calculated as the percentage change in the sector indices relative to the stock market index (normalized to 100 as of January 1, 2007). The returns are then annualized. Relative sector returns are positive if returns on the sector exceed that of the aggregate stock market. The financial sector relative return would capture the degree of financial development relative to other sectors in a country. The commodity sector relative return would indicate the degree of reliance and exposure of a country to movements in commodity prices. These two factors would serve as control variables in examining the impact of sovereign CDS spreads on carry trade activity.

Reflecting the drastic fluctuations in asset prices during the crisis, the sample is divided into five periods (Table 10.1), broadly matching the development of the crisis (Mody 2009). The first "pre-crisis" period spans from the start of the sample period to mid-July 2007, when the turmoil in the U.S. market for asset-backed securities first emerged. The second period starts in mid-July 2007 and ends on March 7, 2008, at the time of Bear Stearns' collapse. The third period ranges between the collapse of Bear Stearns and May 20, 2008, at the time banks in the United Kingdom were

Table 10.1 Timetable for the Collapse of Lehman Brothers

Dummy	Sample Period
D1	Before July 13, 2007
D2	Between July 13, 2007 and March 7, 2008
D3	Between March 7, 2008 and May 16, 2008
D4	Between May 16, 2008 and September 3, 2008
D5	Since September 3, 2008

intervened. The fourth period goes from May 2008 to September 5, 2008, i.e., the time of Lehman Brothers' collapse. Finally, the last period is from the collapse of Lehman Brothers until the end of the sample period at end-April 2009. Dummy variables are used to control for these periods.

Each country is identified as developed or emerging market and whether it is experiencing a crisis or not. The classification between developed and emerging countries are based on the World Economic Outlook. The classification of whether a country is experiencing a crisis is based on whether the country has an active program with the International Monetary Fund as of end-April 2009 (see Appendix for list of countries and their classification).

EMPIRICAL FACTS ON FORWARD EXCHANGE PREMIUM AND SOVEREIGN CDS SPREADS

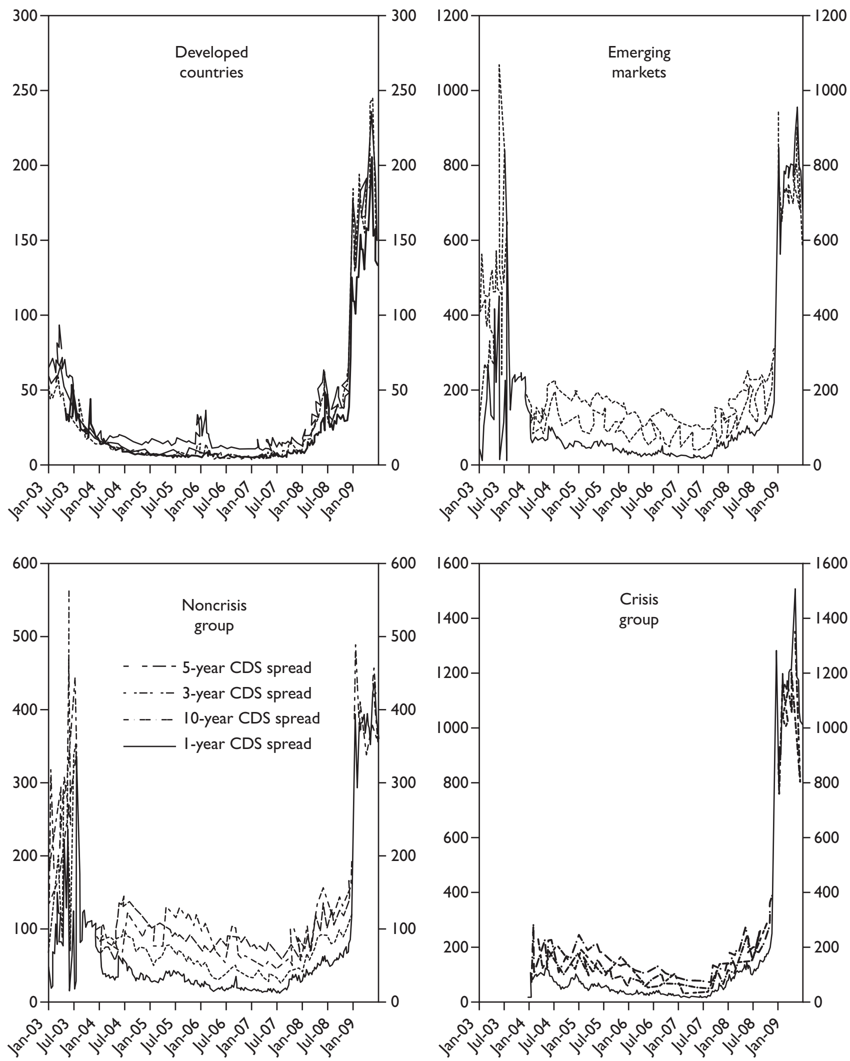
We construct four portfolios of carry trade strategies with different combinations of countries in the portfolio. The first portfolio consists of all emerging markets, the second of all developed countries, the third of all countries currently in crisis, and the last portfolio includes countries that do not have a Fund program as of end-April 2009. Each portfolio assigns equal weight at each point in time for each currency if data are available and the position from the carry trade strategy is not neutral.

Several stylized facts emerge from the summary statistics of the five portfolios (Figure 10.1 and Table 10.2). The table reports the mean, median, standard deviation, and Sharpe ratio of the carry trade returns and the sovereign CDS spreads at different maturities. Returns on the stock market and sector returns are also reported. These summary statistics are reported for each of the five sample subperiods.

Exchange Rate Premium

1. Forward rates move closely with spot rates—most foreign currencies appreciated between 2004 and 2007, but have depreciated sharply since the start of the current crisis.

Figure 10.1 Sovereign Credit Default Swap Spreads (basis points)



Source: CMA, Datastream

2. All five portfolios show positive monthly returns, ranging from 0.13 to 0.44 percent (annualized to be 6.7 and 22.8 percent), with standard deviation of about 4.0 percent. Moreover, Sharpe ratios are significantly higher than the stock market in the U.S.
3. The annualized Sharpe ratios and returns vary significantly across sample periods but remain consistently positive and higher than the Sharpe ratios for stocks, similar to Burnside et al. (2007).

Table 10.2 Summary Statistics

	A. Emerging market economies				B. Developed countries			
	Mean	Standard deviation	Sharpe ratio	Obs.	Mean	Standard deviation	Sharpe ratio	Obs.
Forward exchange rate premium								
Period 1[1]	0.0018	0.0048	0.3680	4616	0.0015	0.0016	0.9331	2368
Period 2	0.0018	0.0048	0.3680	884	0.0016	0.0021	0.7628	442
Period 3	0.0026	0.0028	0.9362	260	0.0016	0.0018	0.9230	130
Period 4	0.0029	0.0029	0.9950	416	0.0020	0.0020	0.9990	208
Period 5	0.0058	0.0117	0.4956	832	0.0013	0.0019	0.6695	416
1-year Sovereign CDS spread								
Period 1	40	60		4239	3	3		1874
Period 2	63	75		922	10	18		442
Period 3	88	89		281	32	74		123
Period 4	107	146		448	28	65		208
Period 5	660	1064		878	135	276		429
3-year Sovereign CDS spread								
Period 1	78	100		4239	7	8		2058
Period 2	98	105		922	16	23		476
Period 3	147	125		281	40	72		133
Period 4	158	162		448	35	62		224
Period 5	667	948		878	159	240		461
5-year Sovereign CDS spread								
Period 1	116	135		4478	9	11		2198
Period 2	128	129		922	19	26		454
Period 3	188	146		281	47	69		133
Period 4	195	175		448	40	62		224
Period 5	660	863		878	167	232		461
10-year Sovereign CDS spread								
Period 1	145	156		4239	15	16		2505
Period 2	157	157		918	24	27		476
Period 3	216	170		271	52	63		133
Period 4	220	193		432	47	59		224
Period 5	647	819		864	161	194		461
Stock market return								
Period 1	0.31	0.24	1.2663	5051	0.24	0.16	1.4736	2944
Period 2	−0.16	0.29	−0.1912	986	−0.30	0.25	−1.2324	544
Period 3	0.02	0.25	0.0906	290	0.25	0.26	0.9757	160
Period 4	−0.53	0.26	−2.0442	464	−0.39	0.19	−1.9851	256
Period 5	−0.52	0.51	−0.9637	957	−0.69	0.51	−1.3486	528
Relative return on financial sector								
Period 1	0.05	0.15	0.3475	4175	0.00	0.10	−0.0135	2890
Period 2	−0.09	0.14	−0.6302	782	−0.13	0.14	−0.9063	544
Period 3	−0.05	0.20	−0.2320	230	0.08	0.17	0.4560	160
Period 4	0.06	1.39	0.0408	368	−0.13	1.33	−0.0975	256
Period 5	−0.07	0.42	−0.1692	759	−0.16	0.39	−0.4131	528
Relative return on commodity sector								
Period 1	0.00	0.22	−0.0122	3680	0.11	0.26	0.4160	2657
Period 2	0.21	0.25	0.8157	680	0.25	0.32	0.7876	544
Period 3	0.39	0.24	1.5808	200	0.18	0.34	0.5169	160
Period 4	−0.02	0.24	−0.0864	320	0.16	0.29	0.5482	256
Period 5	0.01	0.34	0.0373	660	0.32	1.08	0.2948	528

Source: CMA, Datastream, and individual stock markets.

Note: Returns on financial markets are annualized returns. See Appendix 1 for a list of countries in different groups.

1. Period 1 represents date before July 13, 2007; period 2 from July 13, 2007, to March 7, 2008 (Asset-backed secruities to collapse of Bear Stearns; period 3 from March 7, 2008, to May 16, 2008, (collapse of Bear Stearns to UK banks); period 4 from May 23, 20 to September 5, 2008 (collapse of UK banks to collapse to Lehman Brothers); period 5 from September 5, 2008, to end-April 2009 (after collapse of Lehman Brothers).

C. Program countries				D. Nonprogram countries				E. All countries			
Mean	Standard deviation	Sharpe ratio	Obs.	Mean	Standard deviation	Sharpe ratio	Obs.	Mean	Standard deviation	Sharpe ratio	Obs.
0.0017	0.0025	0.6924	1756	0.0017	0.0044	0.3736	5228	0.0017	0.0040	0.4160	6984
0.0012	0.0021	0.5657	340	0.0016	0.0022	0.7322	986	0.0015	0.0022	0.6885	1326
0.0020	0.0020	1.0079	100	0.0024	0.0027	0.8851	290	0.0023	0.0025	0.9014	390
0.0027	0.0023	1.1307	160	0.0026	0.0028	0.9294	464	0.0026	0.0027	0.9733	624
0.0040	0.0039	1.0234	320	0.0044	0.0112	0.3927	928	0.0043	0.0098	0.4364	1248
39	53		1349	26	52		4764	29	53		6113
72	89		340	37	56		1024	46	67		1364
119	97		70	55	80		304	71	89		404
145	158		160	61	114		496	82	131		656
997	1205		320	323	734		987	488	920		1307
72	94		1349	50	87		4948	55	89		6297
102	107		340	60	88		1058	70	94		1398
175	113		100	93	118		314	113	121		414
199	167		160	91	133		512	117	149		672
981	1084		320	339	640		1019	492	816		1339
99	125		1535	76	120		5141	81	122		6676
129	131		340	80	111		1036	92	118		1376
213	121		100	120	142		314	142	143		414
235	171		160	115	151		512	143	164		672
954	992		320	344	584		1019	490	750		1339
129	151		1349	96	139		4940	103	142		6289
152	162		340	98	134		1054	111	143		1394
230	148		100	140	162		304	162	163		404
250	193		160	132	166		496	161	180		656
899	932		320	344	560		1005	478	710		1325
0.35	0.24	1.4695	1713	0.26	0.21	1.2560	6282	0.28	0.22	0.3041	7995
−0.29	0.27	−1.1077	340	−0.10	0.28	−0.3610	1190	−0.14	0.27	−0.5214	1530
−0.19	0.24	−0.8233	100	0.19	0.26	0.7314	350	0.11	0.26	0.4108	450
−0.62	0.27	−2.3368	160	−0.44	0.23	−1.9063	560	−0.48	0.24	−2.0090	720
−0.94	0.59	−1.5935	330	−0.48	0.51	−0.9365	1155	−0.58	0.53	−1.0966	1485
1.10	1.16	0.6040	1234	1.02	0.12	0.1291	5831	0.03	0.13	0.2302	7065
−0.04	0.11	−0.3943	238	−0.12	0.15	−0.8054	1088	−0.10	0.14	−0.7430	1326
−0.10	0.13	−0.7468		0.03	0.20	0.1353	320	0.00	0.19	0.0221	390
−0.25	1.14	−0.2150	112	−0.02	1.41	−0.0134	512	−0.02	1.37	−0.0146	624
−0.38	0.41	−0.9182	231	−0.05	0.41	−0.1194	1056	0.11	0.41	0.2639	1287
−0.02	0.27	−0.0791	1222	0.06	0.23	0.2630	5115	0.44	0.24	1.8689	6337
0.36	0.34	1.0639	238	0.19	0.27	0.7201	986	0.23	0.28	0.7967	1224
0.62	0.37	1.6679	70	0.21	0.27	0.7999	290	0.29	0.29	1.0060	360
0.11	0.28	0.3851	112	0.05	0.26	0.1816	464	0.06	0.26	0.2229	576
0.72	1.55	0.4651	231	0.01	0.37	0.0261	957	0.15	0.76	0.1946	1188

Credit Default Swap Spreads

1. The term structure of CDS spreads in the selected country groups has been largely positive, with the exception in the last period (period 5) since the collapse of Lehman Brothers in September 2008. It was almost flat but positive before the onset of the financial crisis, but has steepened sharply since early 2008. In particular, the difference between 10-year and one-year CDS spreads have exceeded 100 basis points for emerging markets on average. However, the term structure of CDS spreads has turned negative for the crisis countries that have programs with the International Monetary Fund.
2. CDS spreads were at their lowest level between 2005 to early 2007. Spreads for emerging markets were very close to those for developed countries. The difference has increased remarkably since the crisis, as evidenced by the big jump in CDS spreads for emerging markets, while the increase in developed countries has been mild.

The carry trade payoffs in emerging markets are related to the level and term structure of sovereign CDS spreads. This empirical fact is robust even accounting for stock market returns, financial development, and commodity price movements in individual countries. The latter are generally considered to be the factors explaining the profitable carry trade activities in the last few years. Profitability of these carry trade strategies dropped after the collapse in the financial sector and decline in commodity prices in late 2008.

There is strong (at 5 percent significance level) positive correlation between forward exchange premium and CDS spreads across all maturities and in most periods (Table 10.3). For emerging markets, correlation coefficients are larger than 0.3 before the global crisis started, and become less significant in the midst of crisis between March and September 2008 (periods 3 and 4). The correlation becomes stronger after the collapse of Lehman Brothers. A similar pattern was observed for developed countries, although the correlations are less significant and smaller. For crisis countries, the forward exchange premium tends to become significantly correlated with CDS spreads only after the collapse of Lehman Brothers. Nonprogram countries show strong positive correlations throughout the entire sample period.

Notably there is also a strong correlation between the forward exchange premium with the term structure of sovereign CDS spreads. The term structure of CDS spreads is measured by the slope parameter on a rolling basis. The strong correlation remains for most country groups across all sample periods except period 4. Moreover, correlation coefficients were positive in periods before the collapse of Lehman Brothers (periods 1 to 4) and turned negative afterward. This largely reflects the fact that the term structure of CDS spreads have turned negative, with one-year spreads at a much higher level than longer maturity, probably reflecting the short-term increase in default probabilities and lower recovery rate. A higher difference between long- and short-maturity CDS spreads suggest a steepening of the risk term structure, which likely generates a higher forward exchange premium.

The correlations of forward exchange premium with individual stock returns and sector returns on both financial development and commodity boom are negligible—less than 0.05—and insignificant. This finding is consistent with the result in Burnside et al. (2006) that U.S. stock returns could not explain excess returns on carry trades.

ESTIMATION RESULTS AND ROBUSTNESS

The estimation uses a panel regression with fixed effects across countries. All estimations have forward exchange premium as the dependent variable. The dependent variable refers to the returns on the carry trade strategy constructed using the forward and spot exchange rates with bid-ask spreads discussed in the previous section. Returns on each country are then used to construct four portfolios, namely, emerging markets, developed countries, crisis countries, and noncrisis countries, using simple average of returns across selected countries. The right-hand side variables include the lagged forward exchange premium, lagged change in sovereign CDS spreads, the term structure of sovereign CDS spreads, and various control variables. Dummy variables are included to distinguish across subperiods.

$$R_{i,t} = \alpha_i + \beta R_{i,t-\tau} + \gamma \Delta CDS_{i,t-\tau} + \lambda Term_{i,t-\tau} + \phi Z_{i,t-\tau} + \varepsilon_{i,t} \qquad (10.1)$$

where i stands for individual countries, $\tau | \geq 0$ denotes the lagged variables. R_t is the forward exchange premium according to the carry trade strategy described in this chapter's second section, using one month forward rates, *CDS* refers to the five-year sovereign CDS spreads, *Term* is the term structure of the sovereign CDS spreads calculated as rolling average with a two-month window, and Z represents the control variables including stock market returns, sector returns, and dummy variables.

These estimates suggest that sovereign CDS spreads and their term structure are important determinants of the forward exchange premium (Table 10.4). The panel regression shows robust standard-errors despite heteroscedasticity. The reported coefficients on CDS spreads are significant at five percent level for emerging markets and crisis countries. The term structure of CDS spreads is significant across all country groups. Dummy

Table 10.3 Correlation with Forward Exchange Rate Premium

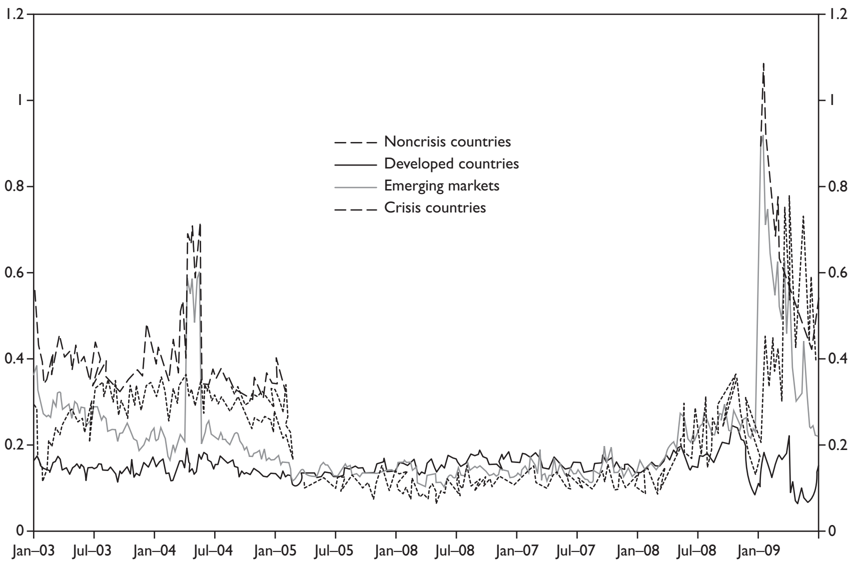
	Before July 13,2007		July 13, 2007, to March 7, 2008 (ABS crisis to Bear Stearns collapse)		March 7, 2008, to May 16, 2008 (Bear Stearns collapse to intervention in UK banks)		May 23, 2008, to September 5, 2008 (Bear Stearns to Lehman Brothers collpase)		September 5, 2008, to now (after Lehman Brothers collapse)	
A. Emerging market economies										
1-year CDS spread	0,26	*	0,19	*	0,04		0,32	*	0,42	*
3-year CDS spread	0,30	*	0,28	*	0,09		0,37	*	0,42	*
5-year CDS spread	0,32	*	0,31	*	0,13	*	0,41	*	0,42	*
10-year CDS spread	0,31	*	0,34	*	0,18	*	0,47	*	0,43	*
Term structure CDS	0,31	*	0,40	*	0,30	*	0,53	*	−0,32	*
Stock market return	0,00		−0,01		−0,03		0,04		0,03	
Financial indicators	0,03		−0,05		−0,02		0,05		−0,05	
Commodities indicators	0,00		0,00		−0,05		0,04		0,03	
B. Developed countries										
1-year CDS spread	−0,02		0,47	*	−0,14		0,35	*	0,35	*
3-year CDS spread	0,09	*	0,35	*	−0,13		0,29	*	0,27	*
5-year CDS spread	−0,10	*	0,36	*	−0,13		0,27	*	0,27	*
10-year CDS spread	−0,09	*	0,37	*	−0,12		0,28	*	0,26	*
Term structure CDS	0,05	*	0,50	*	0,21		0,07		−0,35	*
Stock market return	−0,03		−0,08		−0,04		0,05		−0,01	
Financial indicators	0,01		0,00		0,01		0,02		−0,03	
Commodities indicators	0,03		0,15	*	0,09		0,04		−0,01	

C. Crisis countries										
1-year CDS spread	−0,01		0,07		−0,20	*	0,34	*	0,21	*
3-year CDS spread	−0,03		0,05		−0,14		0,34	*	0,20	*
5-year CDS spread	−0,04		0,04		0,07		0,30	*	0,19	*
10-year CDS spread	−0,05		0,05		0,08		0,36	*	0,18	*
Term structure CDS	−0,07	*	0,03		0,31	*	0,16	*	−0,24	*
Stock market return	0,03		−0,10		−0,07		0,04		0,15	*
Financial indicators	0,00		−0,04		−0,01		−0,05		0,04	
Commodities indicators	0,03		0,24	*	−0,01		0,02		−0,07	
D. Noncrisis countries										
1-year CDS spread	0,36	*	0,34	*	0,17	*	0,37	*	0,63	*
3-year CDS spread	0,38	*	0,37	*	0,23	*	0,45	*	0,64	*
5-year CDS spread	0,39	*	0,39	*	0,26	*	0,49	*	0,66	*
10-year CDS spread	0,38	*	0,41	*	0,28	*	0,52	*	0,67	*
Term structure CDS	0,36	*	0,43	*	0,34	*	0,62	*	−0,42	*
Market return	−0,01		−0,01	*	−0,05	*	0,04		0,01	
Financial indicators	0,04	*	−0,03		−0,04		0,08		−0,05	
Commodities indicators	0,00		0,00		0,03		0,03		0,03	

Source: Datastream.

'*' denotes 5% significant level. See Appendix 1 for a list of countries included in different groups.

Figure 10.2 Monthly Return on Carry Trade Portfolios (%)



variables are all significant, indicating that CDS spreads and their term structure would impact the returns on carry trade differently across periods.

Table 10.4 (second panel) shows the result of the panel regression once control variables on the stock market, financial development, and commodity sector are introduced. Financial development, measured using the returns in the financial sector relative to the market, is significant across all country groups. However, commodity movements do not seem to affect the carry trade returns. Accounting for these control variables, the sovereign CDS spreads and their term structure are still significant at one percent level, confirming the informational content of sovereign CDS spreads in predicting carry trade returns across countries.

For robustness, these results are checked across different sample periods by incorporating dummy variables. After controlling for stock market movements, sovereign CDS spreads and their term structure remain significant (Table 10.5). The interacting dummy also suggests that the impact of sovereign CDS spreads is generally stronger in the initial periods, which are also the years in which carry trade activities were most common. Since the crisis started, the impact of sovereign spreads on carry trade returns, while remaining significant in most cases, has declined. Financial development remains an important variable in explaining the carry trade returns across all country groups.

Table 10.4 Panel Regression Estimates

Dependent variable: R(t)

	Baseline				Baseline with control variable			
VARIABLES	**Noncrisis countries**	**Crisis countries**	**Developed countries**	**Emerging markets**	**Crisis countries**	**Noncrisis countries**	**Developed countries**	**Emerging markets**
R(t-1)	0.500	0.612	0.777	0.603	0.491	0.581	0.749	0.586
	(0.0211)	(0.0240)	(0.0239)	(0.0158)	(0.0243)	(0.0293)	(0.0199)	(0.0192)
	(***)	(***)	(***)	(***)	(***)	(***)	(***)	(***)
R(t-2)	0.292	0.208	0.123	0.247	0.290	0.184	0.116	0.279
	(0.0198)	(0.0242)	(0.0257)	(0.0161)	(0.0227)	(0.0301)	(0.0228)	(0.0197)
	(***)	(***)	(***)	(***)	(***)	(***)	(***)	(***)
CDS(t-1)	0.0000	0.0239	0.0018	0.0072	0.0040	0.0186	0.0143	0.0065
	−0.00002	−0.00003	−0.00002	0.00000	−0.00002	−0.00003	−0.00003	−0.00001
	()	(***)	()	(***)	(**)	(***)	(***)	(***)
Term(t-1)	−4.51	−2.66	−0.256	2.52	−4.14	−5.92	−0.466	4.56
	−0.00779	−0.0103	−0.000981	−0.0118	−0.00889	−0.0148	−0.00106	−0.0139
	(***)	(**)	(***)	(**)	(***)	(***)	(***)	(***)
FIN(t-1)					0.000842	−0.00201	0.00108	−0.000321
					−0.00000974	−0.0000278	−0.0000194	−0.0000146
					(***)	(***)	(***)	(**)
OIL(t-1)					−0.0117	−0.0017	−0.00505	−0.00225
					−0.000869	−0.00935	−0.00501	−0.0141
					()	()	()	()
Rmkt(t-1)					−0.0139	−0.00981	0.114	−0.257
					−0.00189	−0.00733	−0.00252	−0.00254
					()	()	(***)	(***)

Table 10.4 Panel Regression Estimates (*Continued*)

Dependent variable: R(t)

	Baseline				Baseline with control variable			
VARIABLES	**Noncrisis countries**	**Crisis countries**	**Developed countries**	**Emerging markets**	**Crisis countries**	**Noncrisis countries**	**Developed countries**	**Emerging markets**
D1	0.000149	9.18e-05	9.33e-05	4.47e-05	−0.000669	0.00207	−0.000932	0.000358
	(9.97e-06)	(1.63e-05)	(7.54e-06)	(7.33e-06)	(9.14e-05)	(0.000278)	(0.000182)	(0.000142)
	(***)	(***)	(***)	(***)	(***)	(***)	(***)	(**)
D2	2.35e-05	−0.000173	−3.59e-05	2.61e-05	7.03e-05	−5.29e-05	3.95e-05	2.10e-05
	(6.68e-06)	(2.27e-05)	(7.97e-06)	(9.62e-06)	(8.81e-06)	(2.76e-05)	(1.46e-05)	(1.18e-05)
	(***)	(***)	(***)	(***)	(***)	(*)	(***)	(*)
D3	0.000133	−0.000283	6.19e-05	0.000111	0.000207	−0.000153	0.000156	8.43e-05
	(1.39e-05)	(5.63e-05)	(1.52e-05)	(1.63e-05)	(1.96e-05)	(6.69e-05)	(2.29e-05)	(2.03e-05)
	(***)	(***)	(***)	(***)	(***)	(**)	(***)	(***)
D4	0.000174	−0.000198	5.39e-05	0.000134	0.000269	2.40e-06	0.000201	6.71e-05
	(1.16e-05)	(4.69e-05)	(8.34e-06)	(1.34e-05)	(1.95e-05)	(6.01e-05)	(2.57e-05)	(1.79e-05)
	(***)	(***)	(***)	(***)	(***)	()	(***)	(***)
D5	9.25e-05	−0.000779	−0.000104	0.000198	0.000107	−0.000466	1.81e-05	0.000127
	(2.54e-05)	(0.000110)	(3.11e-05)	(2.63e-05)	(2.96e-05)	(0.000138)	(3.81e-05)	(3.00e-05)
	(***)	(***)	(***)	(***)	(***)	(***)	()	(***)
Observations	12099	2720	5559	9810	9158	1841	4789	6540
No. of countries	37	10	17	30	29	7	16	20

Table 10.5 Panel Regression Estimates with Interacting Terms

Dependent variable: R(t)				
	with interacting dummy variables			
VARIABLES	**Noncrisis countries**	**Crisis countries**	**Developed countries**	**Emerging markets**
R(t-1)	0.477	0.577	0.733	0.575
	(0.0236)	(0.0288)	(0.0198)	(0.0201)
	(***)	(***)	(***)	(***)
R(t-2)	0.296	0.177	0.135	0.267
	(0.0222)	(0.0305)	(0.0225)	(0.0204)
	(***)	(***)	(***)	(***)
CDS (t-1)	0.0104	0.0322	0.0115	0.0071
	−0.00002	−0.00009	−0.00003	−0.00001
	(***)	(***)	(***)	(***)
Term(t-1)	−2.82	−6.06	−0.233	5.6
	−0.00805	−0.0158	−0.00109	−0.0141
	(***)	(***)	(**)	(***)
FIN(t-1)	0.00138	−0.00195	0.00131	−0.00027
	−0.00001	−0.00003	−0.00002	−0.00001
	(***)	(***)	(***)	(*)
OIL(t-1)	−0.0112	−0.00121	−0.00763	−0.00528
	−0.000848	−0.00098	−0.000631	−0.00143
	()	()	()	()
Rmkt(t-1)	0.00151	−0.000349	0.13	−0.24
	−0.0018	−0.00738	−0.00242	−0.00253
	()	()	(***)	(***)
D1	−0.00121	0.00196	−0.00115	0.000313
	(0.000130)	(0.000340)	(0.000185)	(0.000143)
	(***)	(***)	(***)	(**)
D2	−0.000107	2.82e-05	−5.28e-05	−6.57e-05
	(1.39e-05)	(3.36e-05)	(1.74e-05)	(1.70e-05)
	(***)	()	(***)	(***)
D3	0.000123	4.00e-05	0.000187	2.71e-05
	(3.55e-05)	(0.000108)	(3.18e-05)	(3.96e-05)
	(***)	()	(***)	()

Table 10.5 Panel Regression Estimates with Interacting Terms (*Continued*)

Dependent variable: R(t)				
	with interacting dummy variables			
VARIABLES	Noncrisis countries	Crisis countries	Developed countries	Emerging markets
D4	0.000267	−0.000101	0.000174	7.91e-05
	(3.87e-05)	(0.000105)	(2.92e-05)	(2.67e-05)
	(***)	()	(***)	(***)
D5	0.000209	−0.000443	−3.48e-05	2.10e-05
	(5.23e-05)	(0.000164)	(4.55e-05)	(4.90e-05)
	(***)	(***)	()	()
D2*CDS(t-1)	0.0667	−0.0193	0.0332	0.021
	−0.0000396	−0.0000826	−0.0000444	−0.0000322
	(***)	(**)	(***)	(***)
D3*CDS(t-1)	0.0118	−0.0214	−0.00379	0.00708
	−0.0000444	−0.0000948	−0.0000344	−0.000033
	(***)	(**)	()	(**)
D4*CDS(t-1)	0.00899	−0.0036	0.00922	0.0018
	−0.00492	−0.00937	−0.00291	−0.0019
	(*)	()	(***)	()
D5*CDS(t-1)	−0.00793	−0.0126	0.00579	0.00512
	−0.0000327	−0.0000834	−0.0000152	−0.0000155
	(**)	()	(***)	(***)
Observations	9158	1841	4789	6228
No. of countries	29	7	16	20
R-squared	0.713	0.846	0.852	0.904
Log Lik	63794	11527	34528	41070

CONCLUSION

Carry trade activity has burgeoned over the last several years to the point to pose serious stress on financial markets in cases of unexpected shocks. Carry trade strategies rest on the empirical regularity, at odds with the uncovered interest parity, that currency that are at a forward discount tend to appreciate offering the opportunity for excess returns. This chapter looked at the excess return arising from carry trade activity and to how it is related to credit risk and other fundamentals by looking at sovereign CDS spreads and their term structure and by controlling for developments in the stock market and commodity sector.

Analyzing a large sample of both developed countries and emerging markets over the past several years, we find that the carry trade activity based on forward exchange premium on average generates an excess return after adjusting for traditional risk factors. Moreover, sovereign CDS spreads and their term structure are significant determinants of carry trade excess returns. The correlation is significant and robust after controlling for stock market returns, financial development, and commodity fluctuations. The correlations also seem to vary through time and country groups, clearly showing the impact of the current global financial crisis.

REFERENCES

Bakshi, G.S. and Naka, A. (1997) Unbiasedness of the Forward Exchange Rates. *Financial Review*, 32: 145–162.

Burnside, C., Eichenbaum, M., and Rebelo S. (2007) The Returns to Currency Speculation in Emerging Markets. *American Economic Review*, 97(2): 333–338.

Burnside, C., Eichenbaum M., Kleshchelski I., and Rebelo, S. (2006) Do Peso Problems Explain the Returns to the Carry Trade? National Bureau of Economic Research, Working paper 12489.

Cavaglia, S., Verschoor, W.F.C., and Wolff, C.C.P. (1994) On the Biasedness of Forward Foreign Exchange Rates: Irrationality or Risk Premia? *Journal of Business*, 67: 321–343.

Chan-Lau, J.A. and Kim, Y.S. (2004) Equity Prices, Credit Default Swaps, and Bond Spreads in Emerging Markets. International Monetary Fund, Working paper 04/27.

Fama, E.F. (1984) Forward and Spot Exchange Rates. *Journal of Monetary Economics*, 14: 319–338.

Geweke, J. and Feige, E. (1979) Some Joint Tests of the Efficiency of Markets for Forward Foreign Exchange. *Review of Economics and Statistics*, 61: 334–341.

Gregory, A.W. and McCurdy, T.H. (1984) Testing the Unbiasedness Hypothesis in the Forward Foreign Exchange Market: A Specification Analysis. *Journal of International Money and Finance*, 3: 357–368.

Hansen, L.P. and Hodrick, R.J. (1980) Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis. *Journal of Political Economy*, 88: 829–853.

Mody, A. (2009) From Bear Stearns to Anglo Irish: How Eurozone Sovereign Spreads Related to Financial Sector Vulnerability. International Monetary Fund Working Paper No. 09/108.

Packer F. and Suthiphongchai, C. (2003) Sovereign Credit Default Swaps. *BIS Quarterly Review*, December 2003: 79–88.

Pan, J. and Singleton, K. (2008) Default and Recovery Implicit in the Term Structure of Sovereign CDS Spreads. *Journal of Finance*, 63(5): 2345–2384.

Zhu, H. (2004) An Empirical Comparison of Credit Spreads between the Bond Market and the CDS Market. Bank of International Settlement Working Paper, No. 160.

APPENDIX 1: LIST OF COUNTRIES IN THE SAMPLE

	Country	Code	Emerging markets (1) / Developed countries (0)	Crisis countries (1) / Noncrisis countries (0)
1	Argentina	ARG	1	0
2	Australia	AUD	0	0
3	Austria	AUR	0	0
4	Brazil	BRA	1	0
5	Chile	CHE	1	0
6	Colombia	COL	1	0
7	China	CHN	1	0
8	Croatia	CRA	1	0
9	Cyprus	CYP	0	0
10	Czech	CZE	0	0
11	Denmark	DEN	0	0
12	Estonia	EST	1	0
13	Egypt	EGY	1	0
14	Finland	FIN	0	0
15	Greece	GRE	0	0
16	Hungary	HUN	1	1
17	Iceland	ICE	0	1
18	India	IND	1	0
19	Indonesia	IDO	1	0
20	Ireland	IRE	1	0
21	Korea	KOR	0	0
22	Kazakhstan	KAZ	1	0
23	Lithuania	LTH	1	0
24	Lebanon	LEB	1	0
25	Malaysia	MAL	1	0
26	Mexico	MEX	1	1
27	NewZealand	NZD	0	0
28	Norway	NOR	0	0
29	Peru	PER	1	0
30	Pakistan	PAK	1	1
31	Philippines	PHL	1	0

(Continued)

	Country	Code	Emerging markets (1) / Developed countries (0)	Crisis countries (1) / Noncrisis countries (0)
32	Poland	POL	1	1
33	Romania	ROM	1	1
34	Russia	RUS	1	0
35	Slovakia	SLK	1	0
36	Slovenia	SOL	1	0
37	South Africa	SAF	1	0
38	Spain	SPN	0	0
39	Switzerland	SWI	0	0
40	Thailand	THL	1	0
41	Turkey	TUR	1	0
42	Ukraine	UKR	1	1
43	Uruguay	URU	1	0
44	United Kingdom	UKB	0	0
45	Japan	JPY	0	0
46	Venezuela	VEN	1	0

NOTES

1. International Monetary Fund, 19th Street NW, Washington DC 20431. The views expressed herein are those of the authors and should not be attributed to the IMF, its Executive Board, or its management.
2. All available information is used rationally; the market is competitive; and there are no taxes, transaction costs, or other frictions. A risk neutral investor needs no compensation for risk and so the future spot rate may not differ from expectation.
3. See Bakshi and Naka (1997); Cavaglia, Verschoor, and Wolff (1994); Fama (1984); Geweke and Feige (1979); Gregory and McCurdy (1984); and Hansen and Hodrick (1980).
4. Part of the carry trade excess return could also be related to the financial development of a country and commodity price movements as these affect trade and capital flows, eventually impacting the exchange rate. For example, many investors borrowed the low-yield Japanese yen and held long positions in the high-yield Australian and New Zealand dollars, which benefited from a surge in commodity prices.
5. CDS markets in some countries or at some maturities may not be very liquid. However, this should not alter the key findings in this chapter since liquidity should be reflected in bid-ask spreads to a large extent.

This page intentionally left blank

CHAPTER 11

A STRATEGIC MANAGEMENT INSIGHT INTO MODEL RISK IN RATINGS

Werner Gleißner, Oliver Everling, and Don Ah Pak

ABSTRACT

In the present chapter, we propose a strategic management simulation approach to model risk in ratings. As companies prosper in developed or emerging markets, and they duplicate their successes in these areas, they have a template that deals exclusively with risk and this gives them the first mover advantage over their competitors. This simulation permits insight into forecasting sales volume, costs, and other influential factors.

The approach adopts elements of strategic management, risk tools currently in use and a performance management tool too creating a new risk management template that will attempt to be conducive to model risk in ratings. In essence the proposed template will heighten the elements in regard to model risk and look at new ways to approach it.

INTRODUCTION

Strategic management is the management art of guiding an organization to reenergizing its goals and to work in tandem, all within a company and responding to a changing environment. It is a disciplined approach to induce fundamental decisions and actions that shape and articulate the organization to focus on the future.

In order to be strategic the organization requires a clear distinction of what are their objectives and to incorporate at a conscious level their responsiveness to operate within a dynamic environment. Strategic management is a discipline, and a sense of order and rigidity needs to be applied, but also requires a certain amount of flexibility, as the dynamics of the environment changes it needs to adapt, yet still maintain a focus and productive approach.

There are issues raised that assist the strategists to examining issues, test assumptions, collate and retrieve data, from a historical point and to rationalize within an educated best guess on how the organization will be positioned in the future.

The dynamics include fundamental decisions and courses of actions, as the choices include the what, why, and how the company does things. In applying an effective strategy this includes a myriad of questions posed, and the choices made are tough, challenging, and may even bring discord, but these in reality are the challenges that need to be faced in bringing the organization to compete in the global market.

The basics of a supportive strategy should be the backbone of strategic thinking: Are we basing our decisions ethically and on soundness, do we fully understand the environment we operate in, and are there any internal and external mechanisms that hinder us to attain our goals? In short, it can be stated, that strategic management prepares and gears the organization to interact accordingly.

Since strategy is an evolving dimension and nothing in its environment remains stable, organizations need to reshape themselves as they encounter new environments and hindrances arises. This includes the willingness to be flexible, psychologically sound, and to effectively make proper judgments. There is a sense of creativity involved in strategy and the tools utilized and data analyzing is not the only ingredients in articulating a strategy, but only the rationale of the people involved. Strategy does not fly straight and deviations along its path will disrupt the process.

Twenty years and more ago the world was a comparably stable and predictable place and on the one hand, strategist could plan over several years to ascertain the goals. Now, all has changed, markets are volatile and one financial crisis can topple markets and cause a worldwide catastrophe. To critically formulate a workable strategy is at best said difficult and fundamentally only a course of actions that the organization proposes and foresees as the best alternative. The approach taken in this insight is essentially a strategic thought, but the management of how the model risk in ratings also needs to assess on the impact of the efficiency and effectiveness of the support systems that organizations use in providing accurate information and to reducing their risks.

As for the development in providing new insight, previous literature requires a thorough examination and a definite need to identifying the key points. This approach is also argued in the development of strategy and evaluation and alignment with its environment, with specific focus on a framework that develops the strategic plan. External market data and program evaluation results provide critical data to support strategy development. Without this information and insight, the organization's strategies will not be in alignment with or effective in the marketplace. The critical issues list serves as the specific focus and framework for the activities of the organization and the pattern of these activities (developing and selecting the strategies). Issues would include: Do the issues that relate to the company's overall strengths and knowledge base and how to integrate these two and propel ahead; is the approach sound; does it answer the right questions; and are we structured accordingly to meet future demands on the existing framework and, if not, what do we need to change?

The inclusion of having the right information and enabling the correct actions is a factor that cannot be underestimated and especially in ratings this is essential in making the right assessment. The need to properly assess the data, how it was acquired, and verifying and disseminating is all crucial to the overall process. This is a systematic approach and has been well documented, but there still seem to be flaws in the generic makeup and this needs to be addressed. Ultimately the strategic management insight is also to reduce the cost overlay, alleviating a high degree of uncertainty and risk, yet still adding value to the decision and creating information that is useful to third parties alike. This concept should be rational, within reason, and reflect adequately to what is being assessed and given the context of the purpose be clear of bias and partiality, and provide information.

The decisions gleaned should derive from a sound framework and has check points and if need be stop gaps where the process can be halted, if deviations occur and correction measures taken. Previously it was mentioned that ultimately people are involved in the overall decision making process and it is within this context that the decision makers would rationalize to make the correct decision, but as we have seen with the subprime crisis and eventually leading to the credit crisis, analysts grew bolder and started to take larger risks.

These decisions ultimately led to the crisis unfolding and it was quite clear that the decisions made were unsatisfactory and the focus was not from an objective and realistic point of view. With regard to these decisions and the adaptation of a strategic point of view, the following themes would be incorporated within the insight proposed: Assessments are to be tailored on the evaluation approach to the specific decision proposed and critical information easily obtainable and based on a mixture of financial and strategic decisions.

In 2008 we saw a new phenomenon occurring: Governments nationalized financial systems and a major bailout from the U.S. government in order to subvert unemployment and restore stability in order to foster a burgeoning economy; and investment banks that were common employment haunts for MBA graduates either shrunk, disappeared, or were merged with the their more robust competitors. The all-too-common energy source, oil, spiralled, while in some countries the residential properties derailed and expectations are more to come.

The uncertainty within the economy has bought some valuable lessons and this will make the proposed model sturdier and more able to withstand the knocks of this new global order that will emerge after the crisis has dissipated; lessons learned and knocks taken will not be easily repeated. Governments will intervene and regulate their national financial systems and this in turn will lead to a more robust global financial system, but it is still up to industry to ultimately be the players in the financial world and this will determine if industry will be able to avoid a decline and respond accordingly.

According to what we know:

- The financial drive within a global context is not functional.
- Governments are trying to stimulate their economies.
- Globalization has introduced trade and growth and free movement of capital.
- The 2008 U.S. domestic crisis has spread globally.
- Already in certain economies, foreign direct investment has shrunk.
- Unemployment is at an all-time high and looks to escalate further.
- The thought of nationalizing certain industries is under consideration.
- Short-term and long-term government intervention is proposed and initiatives undertaken.
- New models are proposed and undertaken.

The models proposed are being envisaged and introduced; it is within this context and time of volatility that we propose a strategic management insight into model risk in ratings.

Thus far, we have provided a basis of what is strategy and have included the elements of strategic decision making and also looked at the uncertainty that lies ahead. The template this is proposed consisted of the process of strategic making, but as stated in the abstract, it would be and includes risk tools currently in use, as this has to a degree provide the essentials to promulgate the proper choices to a certain degree. What is new is the tool of a performance management element will enhance a new risk management template and focus on the internal as well as the ultimate outcome for the decision makers.

In essence the proposed model will heighten the elements in regard to model risk and look at new ways to approach it.

DESCRIPTION OF A TYPICAL RATING METHOD IN THIS CATEGORY

Within the scope of a commissioned research project by Germany's Saxon State Ministry of Economics and Labor, which conducted the Institute for Practical Economic Research and Economic Advice together with the Technical University of Dresden, Germany as well as the advisory association WIMA GmbH, RMCE RiskCon GmbH, and the FutureValue Group AG, for the first time in the determination of a rating for about 150 Saxon companies, a simulation model as a stand-alone rating system has been introduced that can

- Directly reduce the chance of insolvency from the simulation and therewith
- Make rating prognosis possible

A simulation allows a company to forecast sales volume, costs, and other influential factors that would appear on balance sheets, for a project-leader through a planning period of over five years in consideration of the interrogated risks. With the help of simulation-based calculation methods, one can receive an allocation for the balance sheet profit and of the liquidity of the company, whereby the chance of insolvency through excessive indebtedness or liquidity over a period of five years can be directly determined. Then a rating-note or a rating-grade rank can be aligned to the subject, in such a way that the probability of failure can be determined. The initial point of derivation for this rating is the description of the probability of insolvency. The cause of insolvency is self-evident in the following situations:

- An excessive indebtedness, meaning the figure of (economic) capital is smaller than zero
- Liquidation, meaning that the payment obligations are no longer covered through liquid funds and an agreement for short-term credit is made

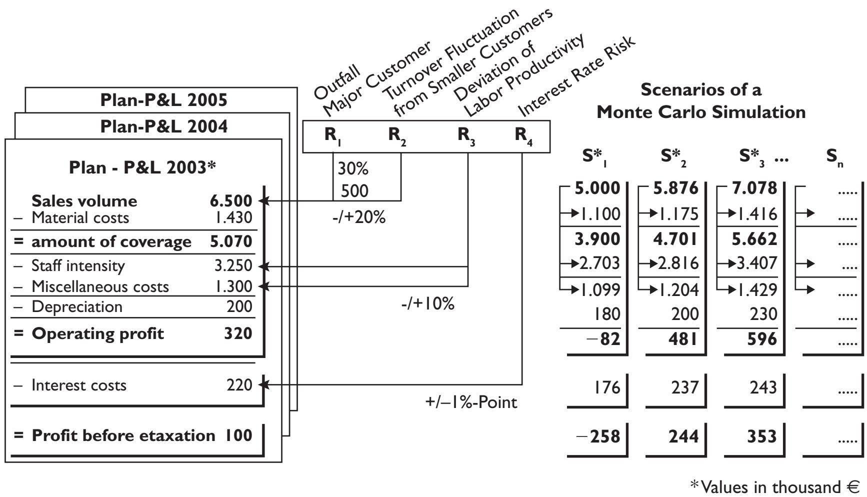
In order to be able to determine the probability for excessive indebtedness, the allocation function of the owned capital has to be determined in each period. This result defines the capital acquired over a period of time from the capital acquired in the previous period, plus the changes of the surplus saved capital and of the deposits and withdrawals, as well as of the retained balance sheet profits, whereupon an alternation of capital-similar funds such as bonds and bindings against allied companies or associations or deposits of informal associates are not included in the planning period. Then with yearly records and analysis of acquired capital, a probability of a nonpositive capital can be determined, which precisely matches the cause for insolvency: "excessive indebtedness."

However, since, as a general rule within companies, no excessive documentation of indebtedness exists, as there is concentration on the vital assets, excessive indebtedness occurs. If the companies have hidden reserves in a larger amount, the probability of default will be estimated too high. If on the other hand high derivative company values, through acquisition, are in the balance sheet, it is possible that the probability of default can be underestimated. The situation of liquidity occurs for a company, if its payments-out exceeds the payments-in plus liquid funds plus not yet exhausted credit lines. But before this situation occurs, the company still has several adaptabilities. Especially investments that can be delayed or even abandoned. Through downsizing or short-term work from personnel, a state of liquidity can also be salvaged.

Main components (regarding the company model, which was the basis of such planning) are the calculation of profit, and the calculated estimate of loss as well as the budgeted balance sheet (see Figure 11.1). The interest applies for the allocation function of capital as well as liquidation. For the aforementioned indicators of profit are modeled according to the formula "profit, is sales volume minus expenses," this means, on one hand, sales volume process and, on the other hand, costs process, and its details must be described.

In determining the probability of insolvency for the period of mid-term company planning, the company is seen through randomly determined processes, means, and affine-linear dependencies on the sales volume (detailed specifications are found in Leibbrand, 2002 and 2004, as well as Gleißner and Leibbrand, 2004a and 2004b). As randomly defined processes establish the sales volume, the material recovered paper utilization rate, the personnel costs, and the interest rates could be modeled, whereby the aforementioned processes make the base variation possible and thus the calculation of the insolvency probability. As a means for discovery, the attitude towards investing, the personnel adjustment, the distribution politics, and the setting of a credit line would be appropriate. Affine-linear dependencies of different balance sheet

Figure 11.1 Randomly Determined Company Model



and profit-and-loss positions of the sales volume define the penetrating power of turnover fluctuations on the acquired capital and liquidity.

Risks that exist within the company like machine failure or miscalculations are considered as separate stochastic processes.

At the execution of the simulation in one run, for example (see scenario S1 from Figure 11.1), it is assumed that in the company that the greatest client drops out (which is why the sales volume in Figure 11.1 drops from the estimated value of 6,500 to 5,000), a major order gets miscalculated (the other expenses rise from 20 percent of the sales volume to 22 percent) and personnel expenses rise (the personnel expenses rise from 50 percent of the sales volume to 54 percent). The accumulation of these unfortunate circumstances leads to an annual loss of €258,000.

If assuming only an acquired capital volume of €150,000 is at hand, this loss cannot be absorbed, so that insolvency on the basis of excessive indebtedness comes into existence. In another scenario, in which the sales volume rises by 9 percent, the material-recovered paper utilization rate stays constant; at constant wages, productivity growth can be reached (scenario S3 in Figure 11.1) and the company reaches an annual surplus in the volume of €353,000 before taxes and at a tax rate of 44 percent €198,000 after taxes, thus an extraordinarily good result. Ten thousand simulation runs stated that the frequency diagram from Figure 11.2 (354 simulation results can not be displayed because they are located outside of the area from −230TE and +250TE).

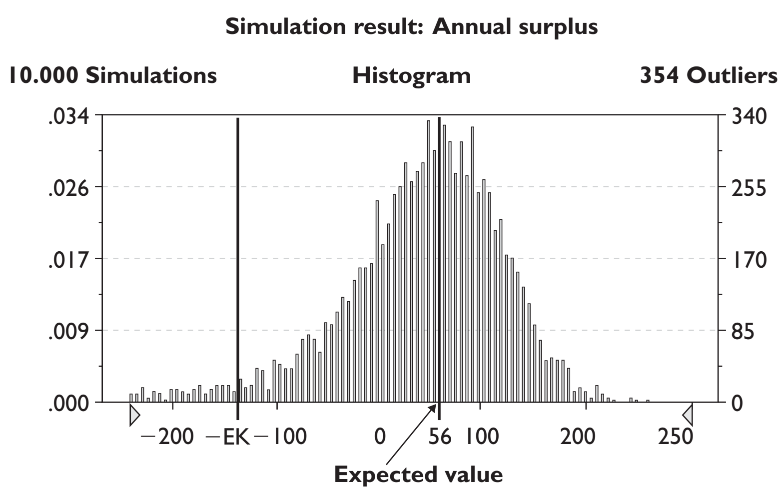
The determination of rating through randomly defined company planning has a crucial theoretical advantage compared with traditional rating systems; that fluctuations of the income level are modeled and that it is solely aimed towards the future. Theoretically, over time, the randomly defined company planning model is the optimal path, which although requires careful consideration, can prove itself as worthwhile, if, for example, it becomes laborious, complex, or even impossible for a credit institution to receive the relevant information.

CONCLUSION

Through increased simulations, the probability of insolvency can be user-defined, approached delicately, and without information being given directly at a desk. Whether the calculated probability of insolvency matches the actual, depends on two set screws: the quality of the company model and the quality of the assumed stochastic processes.

From the Saxon rating project, it is known that the modeling of the company has a crucial influence on the probability of insolvency. If, for instance, for an East German enterprise, the accelerated depreciation remains unconsidered, the probabilities of insolvency will be displayed as far too high. It is also easy to understand the modeling sensitivity for the effect of a sales collapse on the annual surplus. If, for example, no personnel are dismissed at a sales volume calibration, then as a general rule, insolvency has already occurred.

Figure 11.2 Simulation Result of the Randomly Defined Company Planning for Annual Surplus



The second big lever on the rating result comes from the modeling of the stochastic processes and their correlation structure. For instance, are the expenses for products and the staff members positively correlated, how high is the risk of flood water damage? Also, here the Saxon rating process showed that on the part of the enterpriser, the risk situation is not adequately evaluated (see also Leibbrand, 2004).

Since the prognosis of insolvency probability depends mainly on the use of randomly defined processes, the end quality comparison for finance ratings, that is, for the short term, is not at all impressive. However, it seems to be desirable that the companies start to think in density functions, so that the responsible individuals within the companies become more able to observe the randomly defined processes more precisely.

This is particularly helpful, in order to show, for example, the consequences of alternative strategies and planning in rating prognosis, and to make a contribution to crisis prevention in this way. Prognosis for insolvency probabilities through stochastic simulation are then incredibly meaningful, when for a company (for instance, because of great structural changes, growth, or distinctive coincidental effects through historical annual closing data) occurring risks for last annual closings are not representative of the future.

REFERENCES

Cogman, D. and Dobbs, R. (2008) Financial Crisis, Past and Present. *The McKinsey Quarterly*, December: 1 – 8.

Gleißner, W. (2002) Wertorientierte analyse der unternehmensplanung auf basis des risikomanagements. *Finanz betrieb*, 7–8: 417–427.

Gleißner, W. and Bemmann, M. (2008) Rating-evidenz und risikosimulation in strukturellen modellen. *Risikomanager*, 17: 6–12.

Gleißner, W. and Leibbrand, F. (2004a) Der neue ratingansatz. *Bankmagazin*, 3: 38–39.

Gleißner, W. and Leibbrand, F. (2004b), Indikatives rating und unternehmensplanung als grundlage für eine ratingstrategie. In Achleitner, A.K. and Everling, O. (eds.), *Handbuch ratingpraxis*. Wiesbaden: Gabler.

Leibbrand, F. (2002) Die Ermittlung der insolvenzwahrscheinlichkeit mittels stochastischer unternehmensplanung. Working paper.

Leibbrand, F. (2004) Klassifikation von unternehmertum als grobe richtschnur für das existenzgründerrating. In Achleitner, A.K. and Everling, O. (eds.), *Handbuch ratingpraxis*. Wiesbaden: Gabler.

Schoolman, P. (2008) Credit Crisis Lesson for Modelers. *Risk Management: The Current Financial Crisis, Lessons Learned and Future Implications*, 87–88. Society of Actuaries, the Casualty Actuarial Society, and the Canadian Institute of Actuaries. e-book.

CHAPTER 12

MODELING SECURITIZATION TRANSACTIONS UNDER BASEL II USING THE SUPERVISORY FORMULA APPROACH

Martin Knocinski

ABSTRACT

This chapter discusses various practical aspects, in particular sources of model risk, with respect to the application of the supervisory formula approach (SFA) of Basel II to securitization transactions. After a brief introduction of the different approaches that Basel II provides for the risk weighting of securitization exposures, this chapter provides an explanation of the SFA itself as well as its input parameters. The sensitivity of capital requirements calculated according to the SFA's input parameter is then discussed in detail. Thereafter, further sources of model risk resulting from the underlying pool's dynamic character as well as static regulatory minimum requirements are analyzed. At the end of the chapter, a brief summary as well as an outlook on the future role of the SFA within Basel II against the background of recent regulatory reforms is provided.

INTRODUCTION

With the introduction of Basel II, the Basel Committee for Banking Supervision (BCBS) replaced the previous set of rules, commonly known as Basel I. Probably the most important reason that made new capital requirements necessary is the substantial change that the banking industry experienced during the advent of financial engineering and the rise of securitization transactions.

The basics of Basel II are still straightforward (i.e., banks have to hold 8 percent regulatory capital for different types of risks). However, while under Basel I there have been no particular rules for securitizations, Basel II for the first time introduced a so-called securitization framework, which, unfortunately, makes determining capital requirements for securitization exposures far more complex. Hence, this chapter provides an overview of the new rules, how they are implemented and which risks—particularly model risks—arise when the securitization framework is applied.

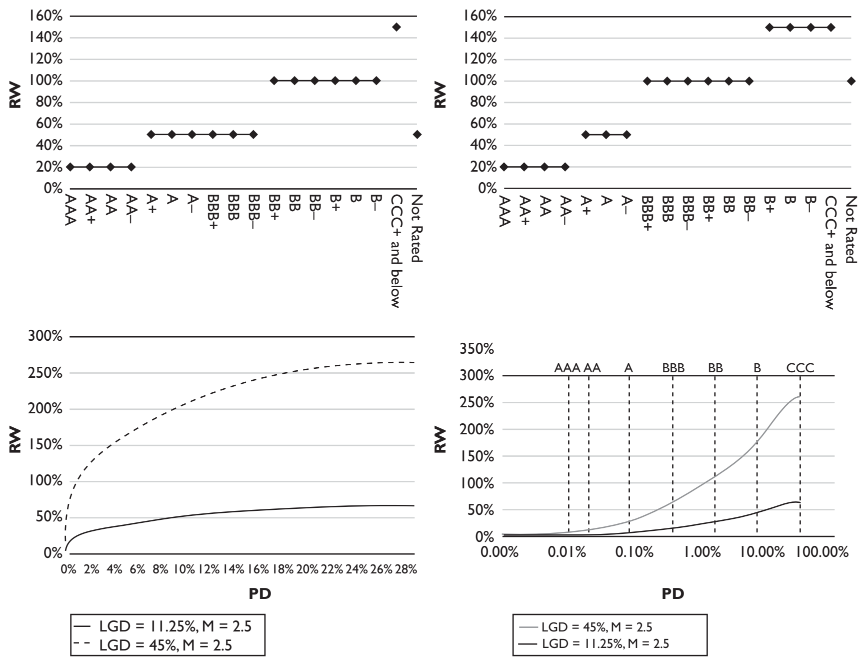
THE BASICS OF THE BASEL FRAMEWORK

According to Basel II, banks can calculate their capital requirements using one of two general types of approaches, i.e., the standardized approach (SA) and the internal ratings-based approach (IRBA). They differ with respect to their sophistication level when measuring credit risk. While the SA focuses on external ratings and is primarily an enhanced version of the Basel I approach, the IRBA does not use external ratings but instead internal ratings. Those internal ratings are derived using internally developed mathematical models based on historical data, or more precisely a *Vasicek* one-factor model based on the following functional relationship: $EL = EAD \times f(PD,M) \times LGD$.

Depending on which IRBA is used; i.e., foundation IRBA (F-IRBA) or advanced IRBA (A-IRBA), banks need to estimate as input parameter *PD* only, or the *PD—LGD* as well as *EAD*. Consequently, each possible combination of input factors provides an individual risk weight which, in addition, is unlikely to be the same for internal models that different banks have developed. In more technical terms, the IRBA risk weight function is a continuous function (see Figure 12.1), while the risk weight function used in the SA is a step or staircase function.

While applying the IRBA is more complex, it promises a more risk-sensitive calculation and thus, at least for some banks, potentially lower capital requirements. In practice, the higher risk-sensitivity as well as "soft pressure"

Figure 12.1 Risk-Weight Functions for SA and IRBA According to Basel II



from regulatory authorities encouraged larger banks to choose one of the IRBAs, while smaller banks in most cases preferred to apply the SA.

THE BASEL II SECURITIZATION FRAMEWORK

The ever-growing significance of securitization transactions as well as the recent trend of financial engineering to structure more and more complex financial instruments (which came to a most likely partial and temporary halt with the emerging credit crisis) attaches particular importance to the newly introduced rules for securitization exposures.

Different Approaches for Securitization Transactions According to Basel II

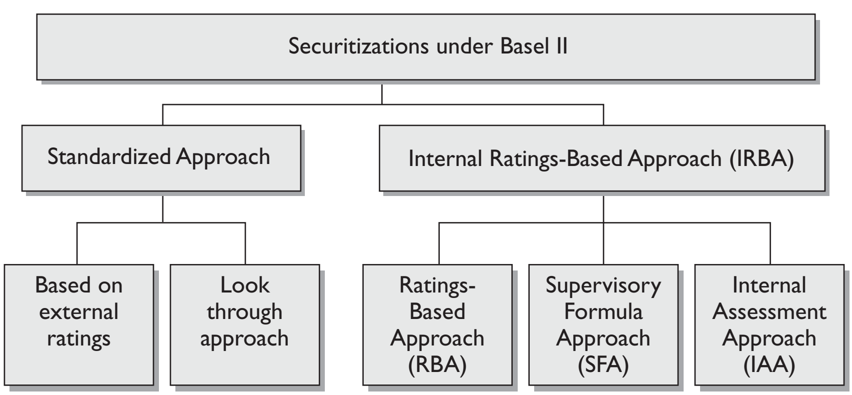
When considering the regulatory treatment of securitization exposures,[1] Basel II entails quite a number of amendments. Before the implementation of Basel II, in most cases banks applied a 100 percent risk weight to securitization exposures. The rationale behind that approach was the fact that it is usually not possible to match an exposure which was acquired in a securitization with an individual debtor. Thus, following an obviously conservative approach, the most unfavorable risk weight for standard exposures was used. As shown below in Figure 12.2, Basel II provides a broader choice of different approaches for securitization exposures.[2]

Determining the exposure's risk weight in the SA is straight forward: Exposures receive risk weights according to their individual external rating or based on the actual composition of the underlying portfolio. In contrast to the fixed 100 percent risk weighting under Basel I, the new rules obviously allow for greater risk sensitivity.

Nevertheless, the Basel Committee on Banking Supervision was aiming at further improving the risk sensitivity of capital charges for securitization exposures: If a bank applies one of the IRBAs, there are three different methods for determining the risk weights of securitization exposures. However, this does not mean that banks can choose freely between these three approaches. In fact, Basel II sets out a strict hierarchy as to which approach might be used and which might not.

If an external rating is available or if a rating can be inferred,[3] the ratings-based approach (RBA) is applied. Under the RBA, risk weights are determined in a similar way as compared with using SA for securitization exposures, i.e., the individual risk weight is based on an external or inferred credit assessment. If an external rating is not available and an external rating cannot be inferred, banks using the IRBA for securitization exposures have to apply either the internal assessment approach (IAA) or the SFA. The IAA may only be used for liquidity facilities, credit enhancements as well as other securitization exposures which were extended to a qualifying asset-backed commercial paper program.

Figure 12.2 Approaches to Measure Credit Risk Arising from Securitization Exposures According to Basel II



If an external/inferred rating is not available and the IAA cannot be used, IRBA-banks are required to determine the risk weights of their securitization exposures using the SFA or have to risk weight the respective exposure at 1,250 percent. Thus, in order to avoid a full capital deduction of the exposures, the capital charge must be calculated based on the functional relationship of several input parameters using the "supervisory formula" that follows:

$$S[L] = \begin{cases} L & \text{when } L \le K_{IRBA} \\ K_{IRBA} + K[L] - K[K_{IRBA}] + (d \times K_{IRBA} / \omega)\left(1 - e^{\omega(K_{IRBA} - L)/K_{IRBA}}\right) & \text{when } K_{IRBA} < L \end{cases}$$

The SFA for Securitization Exposures

The SFA is used to calculate the capital charges based on the functional relationship of several input parameters using a rather complex mathematical formula. More precisely, the capital charge depends on various bank-supplied inputs. Below we introduce the five most important ones,[4] of which three are pool parameters (K_{IRBA}, N, LGD) and two structural parameters (L, T).

The Supervisory Formula Approach Input Factors

The IRBA Capital Charge If the Underlying Exposures Had Not Been Securitized (K_{IRBA})[5]

K_{IRBA} is defined as the ratio of the IRB capital requirement (including the *EL* portion of the underlying exposures in the pool of underlying assets) to the exposure amount of the pool. The IRBA capital requirement is calculated according to the "standard" rules for IRBA banks, including accounting for potential credit risk protection.

The Pool's Effective Number of Exposures; and (N)

The Basel II framework refers to the effective number of underlying exposures (N) comprised in the securitized portfolio. Banks can choose between two alternative methods as to how to calculate N using one of the following formulas, of which the first is the inverse

$$\text{Herfindahl index}^6\text{: } N = \frac{\left(\sum_i EAD_i\right)^2}{\sum_i EAD_i^2} \text{ or } N = \frac{1}{C_1}$$

In the formulas above, EAD_i stands for the sum of all credit risk positions against the debtor i and C_1 for the share that the credit risk position with the highest *EAD* (more precisely, regulatory basis for assessment) has in the total of all individual *EAD* (or basis for assessment) in the securitized portfolio. In this context, debtors who are related to each other in a way which makes it likely that financial problems of one debtor lead to financial problems of the other debtor count as one single debtor and hence do not increase the input parameter *N*. Thus, both formulas for the calculation of the effective number of underlying exposures not only consider the total number of exposures within the pool but also the concentration within the pool.

*The Pool's Exposure Weighted Average Loss Given Default (*LGD*)*

The exposure-weighted average *LGD* is determined according to the following formula:

$$LGD = \frac{\sum_i LGD_i \times EAD_i}{\sum_i EAD_i}$$

In this context, LGD_i is determined as the weighted average *LGD* associated with all exposures to the i^{th} obligor. In the case of a resecuritization, an *LGD* of 100 percent must be assumed for the underlying securitized exposures.

The Tranche's Credit Enhancement Level (L)

The credit enhancement level is measured as the ratio of the amount of all securitization exposures subordinate to the tranche in question to the amount of exposures in the pool. For example, the 10 to 25 percent tranche of a securitization would represent a credit enhancement level of 10 percent. As a general rule, reserve accounts can increase the individual credit enhancement level. For this purpose, the reserve account must be junior to the tranche in question and funded by accumulated cash flows from the underlying assets. Unfunded reserve accounts cannot be considered when calculating *L*.

The Tranche Thickness (T)

The tranche thickness is defined as the ratio of the size of the tranche in question measured by the tranche's notional amount to the total notional amount of exposures in the pool that is securitized. For example, if a portfolio of loans with a nominal amount of $500 million is securitized and the tranche in question has a face value of $100 million, the tranche thickness *T* would equal 20 percent. Alternatively, using the same notation as in the example of a 10 to 25 percent tranche above, the tranche thickness *T* would equal 15 percent.

Individual Importance of the Supervisory Formula Approaches Input Factors

As the above input factors are interdependent to a certain extent, it is difficult to assess each of the factor's individual importance on the total capital charge resulting from a particular (securitization) exposure to which SFA is applied. However, in practice, the pool's effective number of exposures as well as its exposure weighted average loss given default effect on capital requirements is significantly weaker than the effect of the IRBA capital charge had the exposures not been securitized K_{IRBA}, the tranche's credit enhancement level L as well as its thickness T.

The charts in Figure 12.3 depict the sensitivity of the capital requirements calculated according to the SFA to the number of exposures. On the left-hand side, the risk weights for different values of the number of underlying exposures (N) are illustrated, assuming values for the average risk weight including expected losses K_{IRBA}*12.5, LGD, credit enhancement level L, and tranche thickness T of 62.5, 45, 5 and 95 percent, respectively. The graph clearly illustrates that the function shows the highest sensitivity for small values of N. Once the number of exposures reaches values of approximately 60, the curve begins to flatten rapidly. In other words, the impact of incremental increases of N on the risk weight decreases. As can be seen from the graph on the right-hand side, even if the input parameters are modified, the curve's shape remains similar and the general conclusion valid: Except for extraordinarily nongranular pools of underlying assets, the sensitivity of capital requirements to the number of underlying exposures N is low.

While the charts above depict the sensitivity of the capital requirements to the number of exposures N, Figure 12.4 shows on the z-axis by how much a tranche's risk weight increases if the tranche's average loss given default jumps from 45 to 100 percent. For this purpose, N is given as a fixed 125. The three other remaining parameters that are required for the

Figure 12.3 Sensitivity of Capital Requirements Calculated Using the Supervisory Formula Approach to *N*

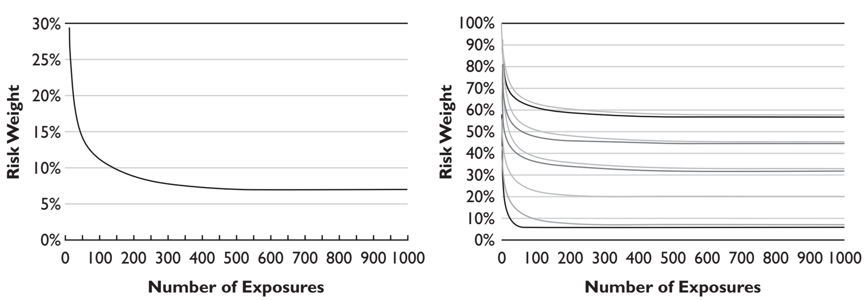
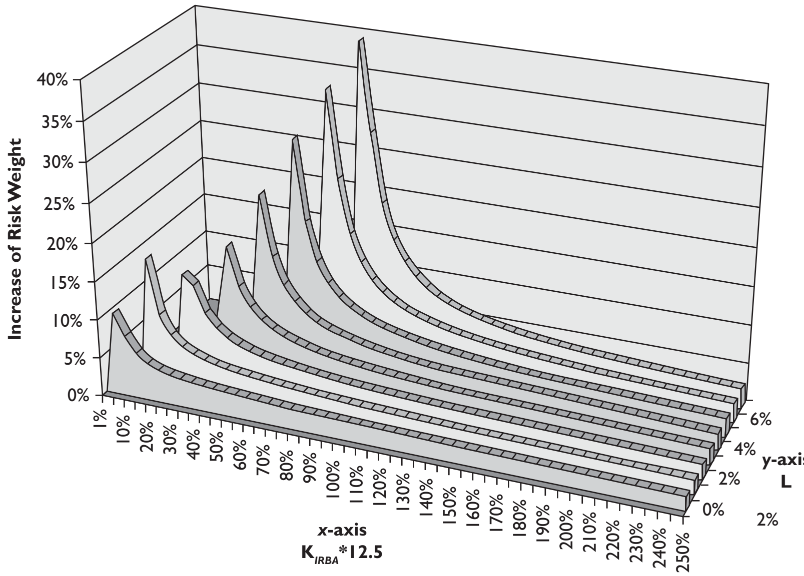


Figure 12.4 Sensitivity of Capital Requirements Calculated Using the Supervisory Formula Approach to *LGD*



calculation of a tranche's risk weight, i.e., K_{IRBA}, L, and T are given on the x-axis (K_{IRBA}) and y-axis (L and T, assuming T is given by $1-L$). In other words, the height of the peaks indicates for various combinations of K_{IRBA} and L (and thus T, as we assume a two-tranche securitization) the increase of a senior tranche's risk weight in percent given a jump of *LGD* from 45 to 100 percent.

While the maximum increase of the risk weights calculated according to the SFA as a result of an increase of *LGD* from 45 to 100 percent in the above example is approximately 35 percent, the graph also shows that for most combinations of K_{IRBA} and credit enhancement level L, the impact on the risk weights is significantly lower, i.e., in most cases below 5 percent. In the example above, in absolute terms, the maximum increase of the risk weight is 9.8 percentage points.

In Table 12.1, the respective risk weights for a senior tranche in a two-tranche securitization are given for different combinations of K_{IRBA}*12.5 and different credit enhancement levels. The number of underlying exposures and the average *LGD* are assumed to be constant at 125 and 45 percent, respectively.

As can be seen from this table, the risk weights of exposures can be reduced significantly under the SFA if the underlying exposures (had they not been treated according to the securitization rules of Basel II) carry a

Table 12.1 Risk Weights According to the Supervisory Formula Approach for a Most Senior Tranche (Example)

	Enhancement level									
K_{IRBA}* 12.5	**1%**	**2%**	**3%**	**4%**	**5%**	**6%**	**7%**	**8%**	**9%**	**10%**
20%	12.8%	**7.0%**	**7.0%**	**7.0%**	**7.0%**	**7.0%**	**7.0%**	**7.0%**	**7.0%**	**7.0%**
40%	35.4%	23.0%	10.4%	**7.0%**	**7.0%**	**7.0%**	**7.0%**	**7.0%**	**7.0%**	**7.0%**
60%	57.6%	45.4%	33.0%	20.3%	7.9%	**7.0%**	**7.0%**	**7.0%**	**7.0%**	**7.0%**
80%	79.4%	67.5%	55.3%	42.8%	30.1%	17.2%	**7.0%**	**7.0%**	**7.0%**	**7.0%**
100%	101.1%	89.4%	77.4%	65.2%	52.7%	40.0%	27.0%	13.7%	**7.0%**	**7.0%**
120%	122.6%	111.1%	99.4%	87.4%	75.1%	62.6%	49.9%	36.8%	23.5%	11.0%
140%	144.0%	132.7%	121.2%	109.5%	97.5%	85.2%	72.7%	59.9%	46.8%	33.4%
160%	165.3%	154.3%	143.0%	131.4%	119.7%	107.6%	95.4%	82.8%	70.0%	56.9%
180%	186.6%	175.7%	164.7%	153.4%	141.8%	130.0%	118.0%	105.7%	93.1%	80.2%
200%	207.8%	197.1%	186.3%	175.2%	163.9%	152.3%	140.5%	128.5%	116.1%	103.5%
220%	228.9%	218.5%	207.8%	197.0%	185.9%	174.6%	163.0%	151.2%	139.1%	126.8%
240%	249.9%	239.7%	229.3%	218.7%	207.8%	196.7%	185.4%	173.9%	162.0%	143.9%
260%	271.0%	261.0%	250.8%	240.4%	229.7%	218.9%	207.8%	196.5%	184.9%	173.0%
280%	291.9%	282.1%	272.2%	262.0%	251.6%	241.0%	230.1%	219.0%	207.7%	196.1%
300%	312.8%	303.3%	293.5%	283.5%	273.4%	263.0%	252.4%	241.5%	230.4%	219.1%
320%	333.7%	324.4%	314.8%	305.1%	295.1%	285.0%	274.6%	264.0%	253.2%	242.1%
340%	354.5%	345.4%	336.1%	326.6%	316.8%	306.9%	296.8%	286.4%	275.8%	265.0%
360%	375.3%	366.4%	357.3%	348.0%	338.5%	328.8%	318.9%	308.8%	298.4%	287.9%
380%	396.1%	387.4%	378.5%	369.4%	360.1%	350.7%	341.0%	331.1%	321.0%	310.7%
400%	416.8%	408.3%	399.6%	390.7%	381.7%	372.5%	363.0%	353.4%	343.5%	333.5%
420%	437.5%	429.2%	420.7%	412.1%	403.2%	394.2%	385.0%	375.6%	366.0%	356.2%
440%	458.1%	450.0%	441.8%	433.3%	424.7%	416.0%	407.0%	397.8%	388.5%	378.9%
460%	478.7%	470.8%	462.8%	454.6%	446.2%	437.7%	428.9%	420.0%	410.9%	401.5%
480%	499.2%	491.6%	483.8%	475.8%	467.6%	459.3%	450.8%	442.1%	433.2%	424.2%
500%	519.7%	512.3%	504.7%	496.9%	489.0%	480.9%	472.6%	464.2%	455.5%	446.7%

rather low-risk weight and benefit from a substantial level of credit enhancements. More precisely, assuming credit enhancement levels of up to 10 percent, the SFA's minimum risk weight of 7 percent can be achieved for exposures which would have received an average risk weight of up to 100 percent under the nonsecuritization rules.

In that context, it is important to understand that the risk weights would differ significantly if the assumption that the tranche thickness T equals 1 minus credit enhancement level L would not hold as this would mean that the relevant exposure is not considered a most senior position. All other factors unchanged, where the investor has an exposure to a mezzanine tranche with attachment point and detachment point equaling 5 and 20 percent, respectively, (i.e., a tranche thickness of 15 percent), the risk weights would increase substantially, as can be seen in the Table 12.2.

The examples in Table 12.2 clearly show that structuring securitization transactions can be a rather complex task given the various input parameters and covenants that have to be considered thoroughly in order to achieve the transaction's objectives. Due to this complexity, in practice securitizations are usually structured using models with different sophistication levels that aim at balancing the various constraints in an optimal way. However, as we will see below, the optimum is not necessarily achieved if a tranche's risk weight reaches the minimum value of 7 percent and thus cannot be further reduced.

Other Sources of Model Risk Resulting from the Application of the SFA

As we have seen previously, various input factors have to be considered when structuring a securitization deal. Finding the optimal balance between these input factors given a particular set of objectives that is to be achieved (e.g., tranching in a way that meets an individual customer's demands with respect to the tranche's risk profile, risk transfer that allows for reduction of capital requirements, achieving a certain cash flow profile, etc.) under the securitization transaction at a given point in time is, without a doubt, a complex task. However, to a certain extent, securitization transactions are somewhat like "living organisms" that are subject to permanent

Table 12.2 Risk Weights According to the Supervisory Formula Approach for a Mezzanine Tranche (Example)

	Enhancement level									
K_{IRBA}* 12.5	**1%**	**2%**	**3%**	**4%**	**5%**	**6%**	**7%**	**8%**	**9%**	**10%**
20%	84.3%	18.8%	7.0%	7.0%	7.0%	7.0%	7.0%	7.0%	7.0%	7.0%
40%	233.9%	150.6%	67.3%	20.7%	7.2%	7.0%	7.0%	7.0%	7.0%	7.0%
60%	380.0%	296.7%	213.3%	130.0%	50.0%	20.0%	7.8%	7.0%	7.0%	7.0%
80%	524.2%	440.8%	357.5%	274.2%	190.9%	107.5%	42.0%	18.6%	7.7%	7.0%
100%	667.1%	583.8%	500.5%	417.1%	333.8%	250.5%	167.1%	83.8%	36.3%	16.7%
120%	808.8%	725.7%	642.5%	559.2%	475.9%	392.5%	309.2%	225.9%	142.5%	66.0%
140%	948.1%	866.3%	783.5%	700.4%	617.2%	533.8%	450.5%	367.2%	283.9%	200.5%
160%	1080.1%	1002.9%	922.5%	840.4%	757.6%	674.5%	591.2%	507.9%	424.6%	341.3%
180%	1192.7%	1128.5%	1055.4%	977.2%	896.3%	814.1%	731.2%	648.1%	564.8%	481.5%
200%	1250.0%	1225.3%	1171.9%	1104.7%	1030.0%	951.0%	869.7%	787.3%	704.4%	621.2%

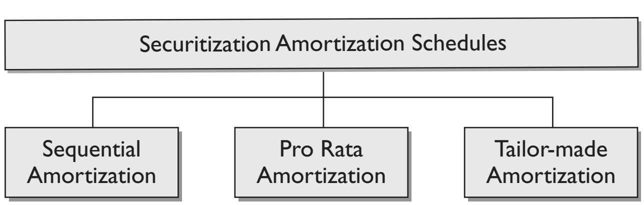
change over the life of a securitization transaction. Thus, structuring a deal "at the limit", i.e., by meeting the minimum requirements in order to achieve the desired result (e.g., risk weighting of a particular tranche) might turn out to be suboptimal from a broader (or longer-term) perspective.

SFA models allow one to "play" with the input parameters assuming either different scenarios with respect to the composition of the underlying pool of assets (thus assuming different levels of K_{IRBA}, LGD, or the effective number of exposures) or a different tranching of the transaction (thus assuming different values for the credit enhancement level L and tranche thickness T). Some models also use optimization routines that provide the most efficient tranching of a securitization transaction in terms of capital requirements given a set of predetermined constraints. However, the ineluctable dynamics of a portfolio of underlying assets and thus of the securitization itself usually remain out of the scope of such models. For a transaction to be successful, however, these very dynamics must be considered and one must attempt to anticipate their potential impact on future cash flows, capital requirements, as well as other relevant factors.

Two of the foremost factors in this context are rating migration (or rating transition)[7] as well as amortization of the transaction's pool of underlying assets. Rating migration, on the one hand, suggests that the average quality of the underlying pool of assets deteriorates over time and, all things being constant, increases K_{IRBA} and thus a securitization tranche's risk weight. Amortization, on the other hand, can (but not necessarily must) have an offsetting effect, depending on the transaction's individual amortization schedule as defined in the legal documentation. In general, for a nonrevolving transaction (i.e., where the pool of underlying assets is not replenished), there are three basic forms of amortization as shown in Figure 12.5.

First, a securitization can be structured as a sequential pay structure, i.e., the most senior tranche is the first to receive principal payments from the underlying pool of assets. Once the most senior tranche is repaid, the next most senior tranche receives payments, and so on. Another way of structuring is a pro rata pay structure where principal payments are used to pay each

Figure 12.5 Different Amortization Structures for Securitization Transactions



individual tranche of a securitization proportionally. The last option is to have a tailor-made amortization schedule that combines the sequential amortization and pro rata amortization based on agreed upon triggers. For example, the amortization schedule could be arranged so that the securitization amortizes sequentially first and changes to a pro rata schedule (or vice versa) once certain performance triggers have been breached.

As noted above, the particular amortization structure should be considered in the structuring process: While a pro rata amortization schedule does not impact the structuring results, sequential structures affect the supervisory formula's input factors credit level enhancement L and tranche thickness T. We will explain this feature of amortization in detail below.

Let us take an example assuming an underlying pool of assets having a total amount of $1,000,000 with an IRBA capital charge had the underlying exposures not been securitized (K_{IRBA}) of 5 percent (equaling an average pool's risk weight of 62.5 percent), an effective number of exposures N of 100 and an average LGD of 50 percent. Let us further assume that the deal structure comprises three tranches, i.e., 0 to 5 percent, 5 to 20 percent, and 20 to 100 percent, respectively. The values for the three individual tranche thickness T and credit enhancement level L as well as the resulting risk weights, risk-weighted assets, and capital charges for each individual tranche are given in Table 12.3.

Let us now assume that at two later points during the life of the securitization (denominated as t_1 and t_2) there have been amortization payments of $100,000 each according to a pro rata amortization schedule. As can be seen from Table 12.4, while the total amounts for the capital charge and risk-weighted asset amount are reduced proportionately, the tranche thickness and credit enhancement level and thus their individual risk weights remain unchanged.

However, where a sequential amortization schedule is used, the amortizations alter the parameters L and T. As shown in Table 12.5, the first amortization in t_1 lowers the credit enhancement level and tranche thickness of the tranches to 77.78 and 22.22 percent, 16.67 and 5.56 percent, as

Table 12.3 Example Portfolio

Exposure ($)	Capital Charge ($)	Risk-weighted Assets ($)	Risk Weight (%)	Thickness (T)	Credit Enhancement (L)
800,000.00	4,480.00	56,000.00	7.00%	80.00%	20.00%
150,000.00	8,603.84	107,547.97	71.70%	15.00%	5.00%
50,000.00	50,000.00	625,000.00	1,250.00%	5.00%	0.00%
1,000,000.00	**63,083.84**	**788,547.97**			

Table 12.4 Example Portfolio after Pro Rata Amortizations

	Exposure ($)	Capital Charge ($)	Risk-weighted Assets ($)	Risk Weight (%)	Thickness (T)	Credit Enhancement (L)
	720,000.00	4,032.00	50,400.00	7.00%	80.00%	20.00%
t_1	135,000.00	7,743.45	96,793.17	71.70%	15.00%	5.00%
	45,000.00	45,000.00	562,500.00	1,250.00%	5.00%	0.00%
	900,000.00	**56,775.45**	**709,693.17**			
	Exposure ($)	**Capital Charge ($)**	**Risk-weighted Assets ($)**	**Risk Weight (%)**	**Thickness (T)**	**Credit Enhancement (L)**
	640,000.00	3,584.00	44,80.00	7.00%	80.00%	20.00%
t_2	120,000.00	6,883.07	86,038.38	71.70%	15.00%	5.00%
	40,000.00	40,000.00	500,000.00	1,250.00%	5.00%	0.00%
	800,000.00	**50,467.07**	**630,838.38**			

Table 12.5 Example Portfolio after Sequential Amortizations

	Exposure ($)	Capital Charge ($)	Risk-weighted Assets ($)	Risk Weight (%)	Thickness (T)	Credit Enhancement (L)
	700,000.00	3,920.00	49,000.00	7.00%	77.78%	22.22%
t_1	150,000.00	4,665.31	58,316.32	38.88%	16.67%	5.56%
	50,000.00	48,078.15	600,976.86	1,201.95%	5.56%	0.00%
	900,000.00	**56,663.45**	**708,293.18**			
	Exposure ($)	**Capital Charge ($)**	**Risk-weighted Assets ($)**	**Risk Weight (%)**	**Thickness (T)**	**Credit Enhancement (L)**
	600,000.00	3,360.00	42,000.00	7.00%	75.00%	25.00%
t_2	150,000.00	2,468.56	30,857.05	20.57%	18.75%	6.25%
	50,000.00	44,414.51	555,181.33	1,110.36%	6.25%	0.00%
	800,000.00	**50,243.07**	**628,038.38**			

well as 5.56 and 0 percent, respectively. This in turn reduces the mezzanine and junior tranche's risk weights to around 39 percent (from 72 percent) and 1,202 percent (from 1,250 percent), respectively. After the second amortization payment of $100,000 in t_2, the risk weights are reduced even further to approximately 21 and 1,110 percent, respectively.

As stated previously, the amortization effect on risk weights can be offset by rating migration trends. Based on the example above, we now use a sequential amortization schedule: Let us assume that K_{IRBA} has increased to 5.5 percent (equals an average risk weight of 68.75 percent) in t_1 and 6 percent (equals an average risk weight of 75 percent) in t_2. Table 12.6 shows that this will partially offset the (positive) effect on risk weights from the sequential amortization. However, while the total effect remains positive, given a steeper increase in K_{IRBA}, the total effect of rating migration and amortization could easily turn out to be negative.

The examples in Table 12.6 clearly demonstrate that the level of complexity of structuring a securitization efficiently increases with the number of (to a certain extent) interdependent parameters. While many SFA models are likely to reach their limits given the structuring challenges discussed above, there are even more aspects that might influence a securitization transaction's success. One of which is the "efficient risk transfer criterion" that banks are required to meet in order to obtain any regulatory capital relief from a securitization according to Basel II. Currently, in most jurisdictions, the focus for identifying a significant risk transfer is on the mezzanine positions of a securitization. In this context, a mezzanine securitization position for which SFA is used is defined as an exposure (1) for which a risk weight of lower than 1,250 percent applies and (2) that is junior to the most senior position. In rating terms, for most cases this turns out to be everything in the range from BBB^- to A^+ (or where the most senior position is rated A^+ or below, everything in the range from BBB^- to the highest rated position which is still junior to the most senior position). A significant risk transfer is usually[8] considered to be achieved if the risk weighted exposure amounts that an originating bank retains from a securitization do not exceed 50 percent of the risk-weighted exposure amounts of all mezzanine positions of the securitization.

Table 12.6 Example Portfolio after Sequential Amortizations and Rating Migration

	Exposure ($)	Capital Charge ($)	Risk-weighted Assets ($)	Risk Weight (%)	Thickness (T)	Credit Enhancement (L)
	700,000.00	3,920.00	49,000.00	7.00%	77.78%	22.22%
t_1	150,000.0	7,685.93	96,074.08	64.05%	16.67%	5.56%
	50,000.00	49,971.45	624,643.12	1,249.29%	5.56%	0.00%
	900,000.00	**61,577.38**	**769,717.20**			
	Exposure ($)	**Capital Charge ($)**	**Risk-weighted Assets ($)**	**Risk Weight (%)**	**Thickness (T)**	**Credit Enhancement (L)**
	600,000.00	3,360.00	42,000.00	7.00%	75.00%	25.00%
t_2	150,000.00	6,000.03	75,000.35	50.00%	18.75%	6.25%
	50,000.00	49.603.01	620,037.66	1,240.08%	6.25%	0.00%
	800,000.00	**50,243.07**	**628,038.38**			

The rationale behind this is actually straightforward: The junior positions are considered irrelevant for the purpose of identifying an efficient risk transfer as these positions are risk weighted at 1,250 percent anyway and thus all the underlying risks are sufficiently accounted for from a regulatory capital perspective. The most senior pieces, despite in most cases being the largest piece with respect to its notional, usually bear only a very small fraction of the entire risks that have been securitized. Thus, the mezzanine positions represent the majority of the underlying risk and consequently are considered when determining whether a risk transfer has been achieved or not. However, this approach can become difficult to handle if securitizations are modeled using the SFA: Consider the securitization in Table 12.7 with three tranches we used as an example (see Table 12.3).

If the originator was to retain a 10 percent share in the 5 to 20 percent tranche (which we do not denominate as mezzanine tranche intentionally) and 70 percent in the 0 to 5 percent tranche,[9] the abovementioned 50 percent rule would be met, as the originator's individual risk-weighted exposure amount in what is considered the mezzanine tranche for regulatory purposes would be $10,754.80 or 10 percent of the regulatory mezzanine tranche's total risk-weighted exposure amount ($107,547.97). The originator's share in the junior tranche is not considered due to its 1,250 percent risk weight as discussed above. However, if the junior tranche's risk weight becomes less than 1,250 percent during the life of the securitization, the 50 percent limit might in fact be up for discussion. If, for example, amortization and rating migration effects reduce the 0 to 5 percent tranche's risk weight down to 1,249.29 percent (see example for t_1 in Table 12.6), it would subsequently have to be considered as a "regulatory mezzanine tranche" when assessing whether a significant risk transfer has been achieved or not. The total risk weighted exposure amount of the mezzanine tranche then would be $150,000 * 64.05% + $50,000 * 1,249.29% = $720,717.20. If the

Table 12.7 Example Portfolio after Sequential Amortizations

Tranche	Nominal ($)	Risk Weight (%)	Total Risk-weighted Exposure Amount ($)	Retained by Originator (%)	Retained Risk-weighted Exposure Amount by Originator ($)
20%–100%	800	7%	56	0%	—
5%–20%	150	71.70%	107,547.97	10%	10,754.80
0%–5%	50	1,250.00%	625,000.00	70%	437,500.00
	1,000,000		**788,547.97**		

originator's share in the 0 to 5 percent tranche and the 5 to 20 percent tranche were to remain unchanged, the originator's individual share in the regulatory mezzanine tranche would be $446,857.59, which exceeds the 50 percent limit (i.e. $360,358.60). Obviously, where achieving a risk transfer and thus regulatory capital relief is part of the originator's target setting (and thus considered when determining the transactions "financials"),[10] inadequate consideration of the problem area discussed above within the models would thwart the success of a securitization as the "financials" of a transaction necessarily depend on the transaction's regulatory treatment.

In addition to the above, the dynamic character of a securitization's underlying asset pool might cause SFA models to fail in another respect: Currently, IRB banks (only these are actually allowed to apply the SFA) have to meet minimum capital requirements, which for most assets are determined based on the rules for IRBA (which should be the standard case for IRBA banks). For some assets, however, IRBA banks are allowed to determine capital requirements based on the rules for the SA, either during a transitional period after the implementation of Basel II only, or for other assets on a permanent basis. However, in order for a securitization to be SFA-eligible, some supervisors require that the underlying pool predominantly consists of such assets that would be treated according to the IRBA rules had they not been securitized. Thus, for transactions that securitize a pool of assets which partially consist of assets that an originator or investor would not treat under the IRBA rules, it must be assured that the share of assets that are treated under the rules for the SA is not and cannot become larger than 50 percent of the total asset pool.[11]

CONCLUSION

There is little doubt that securitizations as such are and will continue to be an important instrument for achieving risk transfer and risk diversification as well as a means of funding for the financial system during as well as after the crisis. Given the fact that international supervisors have taken a rather negative stance in terms of risk weighting toward securitizations in general and resecuritizations, e.g., CDO^2 and similar products, in particular as well as the fact that credit rating agencies going forward will also take a rather conservative approach when assessing the credit quality of securitizations, the application of the SFA to securitizations might experience a boom. This is likely to be supported by the fact that the SFA reveals some interesting benefits compared with RBA as an alternative for calculating risk weights for securitization exposures. In practice, using the SFA is not only rather uncomplicated given its straightforward methodology, it also allows for securitization of a broader range of asset classes (for which rating agencies would require comprehensive information which goes far beyond the asset's individual risk weighting). In addition, the SFA also allows for higher transparency with respect to the structuring process as compared with a situation where rating agencies are involved. This could also mean cost advantages compared to RBA.

However, as discussed previously, the SFA for assessing the capital requirements of unrated securitization exposures bears some significant sources of model risk. First of all, models that are based on the SFA must allow for the different sensitivity of the risk weighting of securitization exposures to the formula's input factors. While some input factors develop a rather limited effect on the risk weighting of a securitization (*LGD*, *N*), others have significantly greater influence (K_{IRBA}, *L*, *T*).

However, it is not only the different input factors' general sensitivity as compared with each other that makes the application of the SFA models a complex task. In fact, the input parameters' sensitivity does not unfold in a linear manner. Instead, there are bands of critical values in terms of sensitivity. Outside these bands, modifications have only a negligible influence on risk weights. Obviously, the assessment of such sensitivities alone is a challenging task. Even though the SFA models are without a doubt important tools in the structuring process, this is particularly the case when looking at the securitization structure at one particular point in time. However, as a securitization's underlying pool is dynamic as it amortizes over time and is subject to a general rating migration trend for example, the "optimal" result that was achieved initially might not necessarily be optimal at any later point in time. As we have seen above, portfolio amortization, depending on the particular amortization schedule, and rating migration can have offsetting effects with respect to risk weights. On the one hand, this has clearly some impact on the "financials" of a transaction. Assuming that modifications to input parameters are incremental, one could suggest that the impact on risk weights and thus the transaction's financials would be incremental, too.

However, irrespective of the fact whether this makes sense from an economic perspective, regulatory rules define critical limits with respect to a transaction's underlying portfolio. We discussed this taking the example of the Basel II rules regarding an efficient risk transfer as well as the composition of the underlying asset pool with respect to IRBA and non-IRBA assets. Where these critical limits are breached, the originator and/or investor of a securitization may no longer be allowed to calculate the capital requirements, according to the SFA (which in turn might require a full capital deduction of a respective exposure), or to recognize any regulatory capital relief from the transaction, even though from an economic perspective a credit risk transfer has actually taken place.

Therefore, structuring a securitization "at the limit," i.e., by meeting the minimum requirements in order to achieve the predefined objectives might turn out to be suboptimal from a broader (or longer) perspective. In fact, it would be more appropriate to include a "risk buffer" that makes the transaction as a whole less sensitive for changes in its underlying parameters. However, how big should such a risk buffer to be included in the structuring process actually be? Unfortunately, there is no generally true answer to that question—and most likely none that SFA models might be able to provide—as this depends not only on the securitization's underlying pool of assets but also on its individual structural elements, i.e., the transaction's number of tranches, their attachment and detachment points, as well as several other (legal) constraints.

REFERENCES

Basel Committee on Banking Supervision. (2006) *International Convergence of Capital Measurements and Capital Standards—A Revised Framework*. Comprehensive version. Basel, Switzerland: BCOS.

Batchvarov, A., Picone, D., Hoogbruin, P.-P., et al. (2004) An Empirical Test of Basel Risk Weights Applied to Securitsation. In *Structured Credit Products—Pricing, Rating, Risk Management and Basel II*. London: Risk Books.

Deacon, J. (2004). *Global Securitisation and CDOs*. New York: John Wiley & Sons.

Gordy, M.B. (2004). Model Foundations for the Supervisory Formula Approach. In *Structured Credit Products—Pricing, Rating, Risk Management and Basel II*. London: Risk Books.

Hu, J. (2004). Rating Transitions in Global Structured Finance. In *Structured Credit Products—Pricing, Rating, Risk Management and Basel II*. London: Risk Books.

NOTES

1. Standard securitization exposures have to be distinguished from so-called first loss position (FLP). FLPs were already under Basel I, subject to full capital deduction.
2. Basel Committee on Banking Supervision (2006), p. 120.

3. Basel Committee on Banking Supervision (2006), p. 136. In this case, an unrated position is assigned the external rating of another securitization exposure which is subordinated as compared with the original exposure in every respect.
4. The SFA also provides for different treatment of retail and nonretail portfolios for which the calculation routine is slightly different. Thus, it is open to dispute whether the retail criterion is actually another input parameter or whether there are two different SFAs for retail and nonretail exposures, respectively.
5. Cf. Deacon (2004), p. 247.
6. Cf. Gordy (2004), p. 320.
7. For an overview on rating transitions, cf. Hu (2004), p. 87.
8. The 50 percent rule is used as a standardized test. However, banks are free to demonstrate that an efficient risk transfer has been achieved even if the 50 percent rule is not accomplished.
9. In order to avoid confusion with regard to the denomination of tranches as "mezzanine" from a regulatory perspective, we refer to the tranches simply by naming their respective attachment and detachment points.
10. Cf. Batchvarov, et al. (2004), p. 363.
11. For example, Germany's supervisory authority BaFin has issued a statement that beginning October 1, 2009, it will object to the application of the SFA to transactions where the share of non-IRBA assets within the underlying asset pool exceeds 50 percent.

This page intentionally left blank

CHAPTER 13

MODEL RISK IN CREDIT MANAGEMENT PROCESSES

Giacomo De Laurentis and Giampaolo Gabbi

ABSTRACT

The increasing usage of quantitative techniques in rating assignment and loan portfolio management is a great source of model risk and amplifies the tendency towards commoditization and short-termism of bank lending. Relationship banking is put in jeopardy. Roles and responsibilities of relationship managers, credit risk models structures, and statistical-based rating systems architectures are clear indicators of the magnitude and the nature of model risk in credit management processes. Results are relevant for banks' strategies and organization design, as well as for improving regulations on banks.

INTRODUCTION

We define model risk for ratings systems (MRRS) as the economic loss deriving from unexpected outcomes related to the use of rating models. The relation with unexpected outcomes is indirect when the unexpected outcomes are a consequence of (1) use of ratings as inputs of credit portfolio models, and/or (2) business and cultural modifications in the market. It is direct when the unexpected outcomes are a consequence of typical applications of rating systems, such as portfolio reporting, credit management, credit administration, and process auditing.

Given MRRS definition and Basel II definition for rating systems ("comprises all of the methods, processes, controls, and data collection," Basel Committee, 2004, §394), taxonomy of direct sources of MRRS is presented in Table 13.1. Direct sources of model risk are clearly addressed in Basel II, mainly in Part 2, III, H.

At the same time, "it is not the Committee's intention to dictate the form or operational detail of banks' risk management policies and practices" (Basel Committee, 2004, §389). As a consequence, individual banks directly and freely set up their models and should compare the model risk arising from a key choice: the chosen degree of mechanization of rating systems. However, national supervisory authorities are often using secondary regulation and/or moral suasion to orientate banks' choices toward statistical-based rating systems (SBRSs). In any case, the degree of mechanization of the assignment process is differentiated in different market and/or product segments. SBRSs can be appropriate for markets and/or product segments where model risks may be compensated by cost reduction of analysis, increased objectivity and consistency of ratings, faster underwriting, better separation of the risk-taking oriented loan officers (focused on commercial activity in loan departments), and risk-controlling oriented credit officers (focused on analysis and underwriting in credit departments).

Table 13.1 Direct Sources of Model Risk of Rating Systems and Treatment in Basel II

Direct Sources	Basel II References
Model design	Part 2, III, H, 3, i) Rating dimension
	Part 2, III, H, 3, ii) Rating structure
Dataset	Part 2, III, H, 4 iv) Data maintenance
	Part 2, III, H, 7 Risk quantification
Model building	Part 2, III, H, 3, iii) Rating criteria
	Part 2, III, H, 3, iv) Rating assignment horizon
Model calibration	Part 2, III, H, 7 Risk quantification
Model usage	Part 2, III, H, 4 Risk rating systems operations
	Part 2, III, H, 5 Corporate governance and oversight
	Part 2, III, H, 6 Use of internal ratings
	Part 2, III, H, 12 Disclosure requirements
Internal validation and compliance with Basel II requirements	Part 2, III, H, 8 Validation of internal estimates
	Part 2 (Pillar 1, in case of application for IRB approaches)
	Part 3 (Pillar 2)
	Part 4 (Pillar 3)

IRB, internal ratings-based.

The research question of this chapter is: Are SBRSs, and the implied model risk, also appropriate for segments of the credit market traditionally based on relationship banking?

METHODOLOGY AND POLICY IMPLICATIONS

The methodology is based on the distinction of direct and indirect relations between the use of rating models and unexpected outcomes. Our hypothesis is that transactional banking and relationship banking face model risk differently (Table 13.2).

The research hypothesis will be tested by

- Identifying which are the potential direct and indirect relations and implications
- Explaining why indirect relations are significant for relationship banking
- Clarifying why direct relations are stronger in relationship banking than in transactional banking, and why they are only partially controlled in the former case

Policy implication of positive testing of the research hypothesis is that it is necessary to reduce the degree of mechanization of rating systems used in relationship banking.

The issue we analyze is relevant because of the huge impacts on long-term profitability of banks, optimal allocation of financial resources, and economic growth.

The structure of the chapter is as follows: in the first and second section of this chapter are set definitions, research question, research hypothesis and methodology, relevance of the issue and policy implications; in the third section, the literature review is presented; in the fourth section, sources of model risk for SBRS are identified and the size of the model risk

Table 13.2 The Research Hypothesis

Use of SBRS and Unexpected outcomes	Relationship Banking	Transactional Banking
Direct relations	Strong and partially controlled	Medium and under control
Indirect relations	Strong and overlooked	Weak and ignored
Consequence	Model risks possibly higher than SBRS advantages	Model risks probably lower than SBRS advantages

SBRS, statistical-based rating system.

arising is outlined; in the fifth section, it is showed that the above identified indirect sources of model risk are real consequence of the use of SBRSs. The sixth section of this chapter shows why direct relations are stronger in relationship banking and why they are only partially controlled at the moment; in the final section conclusions are drawn.

LITERATURE REVIEW

There are many components in the rating process which could lead to model risk, such as (1) the economic cycle and the impact in pro-cyclical credit policies; (2) the size of companies demanding loans; and (3) the characteristic of industries that borrowers belong to.

Pro-cyclicality depends on three main factors: (1) how capital requirement has been designed by regulators; (2) how banks implement their capital management approaches; and (3) how rating are estimated through internal models. First, in order to make the capital accord less pro-cyclical, regulators corrected the relation between ratings and risk weights. Second, solutions such as introducing dynamic provisioning policies that allow the use of reserves accumulated in "sunny" periods (when returns are higher than the long-term average) help to reduce pro-cyclical effects (Gordy and Howells, 2004). Thirdly, rating systems can be designed *through-the-cycle* (TTC) or *point in time* (PIT). Rating agencies look for stable ratings TTC, whereas banks tend to use PIT ratings that change according to the stage of economic cycle. To reduce the risk of cyclicality and short-termism embedded in many credit risk models, some adjustments have been suggested (Pederzoli and Torricelli, 2005)

The second component is the firm size. A large part of the literature has focused on the special character of small business lending and the importance of relationship banking to face information asymmetries that affect small and medium enterprises (SMEs) in particular. Many studies discuss the role of soft information (Berger, Frame, and Miller, 2002; Allen, DeLong, and Saunders, 2004; Petersen, 2004; Degryse and Ongena, 2005). Some other risk factors for small business loans depend on monitoring costs, informative transparency, and recovery rates.

The third component is the industry impact. Portfolio diversification has been analyzed mainly in order to manage market risk. Applications of financial approaches to credit risk are more recent in financial literature; Morris (2001) compares the concentration issues for credit portfolios in different countries and sectors, demonstrating that: (1) most countries set limits on large exposures for banks, and (2) strong differences remain among countries in terms of limits for exposures to specific industries.

In the late 1990s two surveys froze the state of the art of internal rating systems in their early stages. The Basel Committee (2000) stated that alternative approaches can be viewed as points on a continuum with, at one extreme, systems focused on the judgment of expert personnel, and at the other, those based solely on statistical models. Treacy and Carey (1998) obtained similar findings on large U.S. banks. Since the late 1990s, there has been a clearly observable worldwide tendency of banks to develop rating systems that relay much more on statistical-based scoring models (De Laurentis, Saita, and Sironi, 2004). Many national supervisory authorities are pushing in this direction. This tendency is not driven by Basel II requirements—just the opposite, the Basel Committee has stated warnings on the use of mechanical approaches (Basel Committee, 2004, §417). In order to take into account all relevant and material information not considered by the model, banks are often combining model results with human judgment using the so-called "override process." A series of questions arise: Is override room large enough to take account of external-to-model information? Are SBRSs forward looking enough to enable relationship banking? In the mainstream literature relationship banking is associated with credit risk assessment processes based on the use of soft information, bottom-up credit analysis methodologies, customer proximity of those having lending authority (Diamond 1984; Berger and Udell, 2001; Petersen and Rajan, 2002; Degryse and Ongena, 2005). Brunner, Krahnen, and Weber (2000) have analyzed information production changes in credit relationships due to the increasing use of internal rating systems. Some large banks, traditionally less inclined to interrelate with small, informationally opaque and risky businesses with a relationship-oriented approach, saw divisionalization by customers segments as the way to attack the attractive markets of local banks (De Young, Hunter, and Udell, 2003; De Laurentis, 2005). Direct sources of model risk of SBRS derive from weaknesses of their discrimination and calibration properties. The assessment of these properties is part of typical validation of rating systems (Engerlmann and Rauhmeier, 2006; Committee of European Banking Supervisors, 2005; Basel Committee, 2005). Agencies' ratings performance can be considered benchmark measures for banks' rating systems, because of their long time-series and their publicly available results (Standard & Poor's, 2009).

INDIRECT SOURCES OF MODEL RISK: RATINGS AS INPUTS FOR PORTFOLIO CREDIT RISK MODELS

A rating system can be classified as cyclical, anti-cyclical, or neutral depending on its relation with the business cycle (Catarineu-Rabell, Jackson, and Tsomocos, 2005). TTC and PIT ratings are the extremes of a continuum. Rating agencies tend to assign stable ratings over the cycle (but associated default rates are volatile), whereas banks tend to use PIT methods focusing on short-term borrowers' credit quality. The choice of a limited time horizon is driven by a variety of factors, including data availability, internal budgeting cycle of banks, expected time needed to raise capital and implement risk mitigation actions. Both TTC and PIT ratings are dependent on the economic cycle, even if at different degrees. In any case, ratings measure the risk of individual instruments or borrowers; as such, they do not explicitly consider correlation and its changes over time that must be properly addressed by full portfolio credit risk models. These models have different structures, but most of them extrapolate from recent history, so that good current economic conditions signal good future prospects. When these models are not carefully designed to take account of ratings variability and ratings calibration over time, a first source of model risk can be observed.

Regulators have clearly identified this model risk: for instance, the Financial Stability Forum (2009) addressed recommendations to the Basle Committee: (1) to carry out regular assessments of the risk coverage of the capital framework in relation to financial developments and banks' evolving risk profiles; and (2) to make appropriate adjustments to dampen excessive cyclicality of the minimum capital requirements, in particular, to reassess mechanisms through which migrations in credit scores should be addressed. Nevertheless, the Basel Committee preliminary conclusion is to maintain the risk sensitivity of the inputs of capital requirements and instead focus on dampening the outputs. This approach does not eliminate a great part of such model risk.

When banks manage loans, they are expected to estimate not only the idiosyncratic risk, but also the concentration one, which could depend on many factors, such as the firms' size and industry. In the New Accord (Basel II), the highest asset correlation for corporate exposures (0.24) will apply to the lowest probability of default (PD) that are typical among large companies; the lowest asset correlation (0.12) applies to firms with the highest PD, typically small ones. SMEs exposures' risk weights depend on firm-size adjustments (firms' sales). For "other retail exposures," it is in the range of 0.03 to 0.16. The main reason for this differential treatment is that small business loans are generally found to be less sensitive to systematic risk, being more of idiosyncratic nature. Another reason is that maturities are generally shorter for loans to small firms (Dietsch and Petey, 2004, for France and Germany; Shen, 2005 for Taiwan). Gabbi and Vozzella, 2009a, analyzed Italian data (Table 13.3). The rigid treatment of firm size adopted by Basel II generate a model risk in terms of correlation errors (Gabbi and Vozzella, 2009b), as shown in Table 13.4.

Table 13.3 Asset Correlation and Risk Across Size Classes (Italy, turnover in million €)

Rating	<1M	1–5M	5–7.5M	7.5–10M	10–25M
A	0,132	0,062	0,070	0,091	0,093
BBB	0.144	0.065	0.096	0.106	0.103
BB−	0.144	0.086	0.099	0.079	0.074
B	0.159	0.112	0.119	0.121	0.137
B−	0.143	0.150	0.126	0.198	0.155
CCC	0.196	0.257	0.276	0.217	0.230

Table 13.4 Asset Correlation and Risk Across Industries (Italy)

Sectors	1	2	3	4	5	6	7	8	9	10	11
A	0.1054	0.1127	0.1136	0.1359	0.1016	0.1538	0.0533	0.1859	0.2131	0.1640	0.0913
BBB	0.0856	0.1209	0.1136	0.1051	0.1186	0.1279	0.0801	0.0981	0.0667	0.1321	0.1123
BB−	0.1280	0.1079	0.1022	0.0970	0.0798	0.1449	0.0832	0.0913	0.1610	0.0911	0.1126
B	0.1330	0.1292	0.1440	0.1287	0.1247	0.1341	0.1130	0.0728	0.1716	0.1183	0.1096
B−	0.1513	0.1425	0.2476	0.1715	0.1148	0.1468	0.1586	0.1230	0.2105	0.2096	0.1367
CCC	0.1802	0.2244	0.2298	0.2602	0.1914	0.2340	0.2456	0.1302	0.2934	0.3536	0.2334

Sectors: (1) real estate; (2) retail consumer; (3) car industry; (4) wholesale consumer; (5) building; (6) food; (7) mechanics; (8) oil and gas; (9) transport; (10) leather and shoes; (11) clothing industry.

DETERMINANTS OF INDIRECT SOURCES OF MODEL RISK FOR SBRS

In this section we prove that in relationship banking direct and indirect relations of model risk with SBRS are strong and only partially controlled or overlooked.

In relationship banking, banks are engaged in both assessing borrowers' credit-worthiness on medium-long term and feeding customers with the most appropriate products, advisory services, and assistance. For information-based theory of financial intermediation, banks exist because of information synergies they can exploit from credit risk assessment processes and commercial activities. The problem is that risk analysis based on SBRS does not produce information spillovers beneficial for commercial activities.

A different picture arises when relationship managers and credit analysts interact to elaborate information. There are different degrees of spillovers coming from more or less sophisticated judgmental approaches. A sophisticated approach, in which analysts are required to assign partial ratings to firm's "risk factors" such as business risks, financial risks, (borrower's own) credit risks, and operating risks, would require to integrate all available information sources in order to address the key risks a firm is facing. Doing so, the bank achieves a deep understanding of a firm's strength and weaknesses and opportunities and needs, also useful for relationship managers who can comprehensively servicing and advising customers. At the same time, they can provide to credit analysts valuable private information for a better assessment of risks on a longer time horizon.

A simpler judgmental approach, where analysts are required to determine the final borrower rating by assigning partial ratings to different "data sources" (typically, income statement, balance sheet, flow of funds statement, behavioral data, credit register data, business sector, strategic positioning) would result in a poorer understanding of firms, because the analysis misses the key risk factors triggering credit risk: information spillovers for commercial activities are much lower and credit risk assessment is less forward looking. Compared with the outlined simpler judgmental approach, SBRSs are even less informative.

The hypothesis of banks requiring the use of judgmental analysis of borrower credit-worthiness for purposes other than rating assignment and credit underwriting suffers from severe limitations: (1) bearable costs (and, consequently, effectiveness) of this analysis because of its limited scope, (2) low incentives to undertake deep firm's analysis, (3) relationship managers do not benefit from credit analysts" expertise and bear the entire burden of the analysis. This is the framework in banks using SBRS as the key tool to assess PDs.

In case the judgmental analysis is developed for credit underwriting, whereas SBRS are used for other risk management purposes, the drawbacks are: (1) ratings are not benefitting from judgmental analysis achievements, as they are separate processes; and (2) ratings do not reflect risk valuations developed for underwriting decisions, that is, provisions, capital requirements, and risk-adjusted performance measures based on rating are not directly linked with individual lending decisions. This is the peculiar framework envisaged by supervisors requiring that "those having credit underwriting authorities . . . must not have the power of the final assignment of ratings" (Bank of Italy, 2006).

To overcome some weaknesses of SBRS, banks are often combining model results with human judgment ("override process"). This seems appropriate also to take into account all relevant and material information not considered by the model, that is a Basel II–specific requirement. However, in the Basel II framework, for "model-based ratings," overrides are strictly regulated exceptions. Banks are actually limiting override room, above all when it leads to improvements of ratings. Rationale of limiting changes to model-based ratings through overrides is twofold: override proposals are usually set by relationship managers, who can be interested to increase loans extended and reaching their personal targets; qualitative considerations, considered for overrides, are simplified valuation of some aspects of the borrower, usually based on multiple-choices questionnaires.

We conclude that these approaches are also not compatible with relationship lending.

DIRECT RELATIONS ARE STRONGER IN RELATIONSHIP BANKING AND PARTIALLY CONTROLLED

In this section, we focus on model risk directly arising from weaknesses of SBRS to produce a fair and stable rank ordering of risk among borrowers. Basic validation tools for discriminatory power are transition (or migration) matrix, Lorenz curves, and Gini ratios (or their twins ROC curves and AuROC).

A transition matrix indicates satisfactory discriminatory power when: (1) transitions to default (default rates) are higher for worse ratings; (2) values on the diagonal are high, indicating that ratings are stable and forward looking (if the first condition holds, we can exclude that stability denotes poor sensitivity); and (3) transition rates to closer classes are higher than transition rates to less contiguous classes. Benchmark measures for banks' rating systems are agencies' ratings, due to their widespread availability and long time series. Table 13.5 shows that, on the long run (1981–2008), Standard & Poor's ratings satisfy all three conditions.

Table 13.5 Corporate Transition Matrices (1981–2008)—One-Year Transition Rates (%)

From/To	AAA	AA	A	BBB	BB	B	CCC/C	D	NR
AAA	**88.39**	7.63	0.53	0.06	0.08	0.03	0.06	0.00	3.23
AA	0.58	**87.02**	7.79	0.54	0.06	0.09	0.03	0.03	3.86
A	0.04	2.04	**87.19**	5.35	0.40	0.16	0.03	0.08	4.72
BBB	0.01	0.15	3.87	**84.28**	4.00	0.69	0.16	0.24	6.60
BB	0.02	0.05	0.19	5.30	**75.74**	7.22	0.80	0.99	9.68
B	0.00	0.05	0.15	0.26	5.68	**73.02**	4.34	4.51	12.00
CCC/C	0.00	0.00	0.23	0.34	0.97	11.84	**46.96**	25.67	14.00

Source: Standard & Poor's, 2009.

Table 13.6 denotes that, considering transitions in a five-year period, of course stability decreases but other conditions are still met.

Rating models discriminatory power can be quantified using Lorenz curves. Once observations are ordered from worse to better ratings/scores, the Lorenz curve is a graphical representation of the proportionality of a distribution of the cumulative share of issuers by rating (*x* axis) plotted against the cumulative share of defaulters (*y* axis). If the rating system is able to perfectly separate nondefaulting and defaulting borrowers, the curve would reach 100 percent of defaults on the *y* axis while having considered only the exact percentage of defaulted borrowers in the sample on the *x* axis. On the other hand, if the system assigns ratings randomly, the Lorenz curve falls along the diagonal. Thus, a good model has a curve quite vertical and close to the perfect model curve. The Gini coefficient is a summary statistic of the Lorenz curve, representing the area between the random model curve and an actual model curve, divided by the area between the perfect model curve and the random model curve: zero indicates that the actual model behaves randomly, one that mirrors the perfect model (Figure 13.1). Of course, discriminatory power decreases when a longer time horizon is considered, as it is more difficult to predict the issuer-quality five years ahead than only one year from the time the rating is assigned (Figure 13.2). But performances are still satisfactory.

Summary statistics of Gini coefficients indicate valuable discriminatory power, both at one-year and longer time horizons, above all for nonfinancial corporates (Table 13.7).

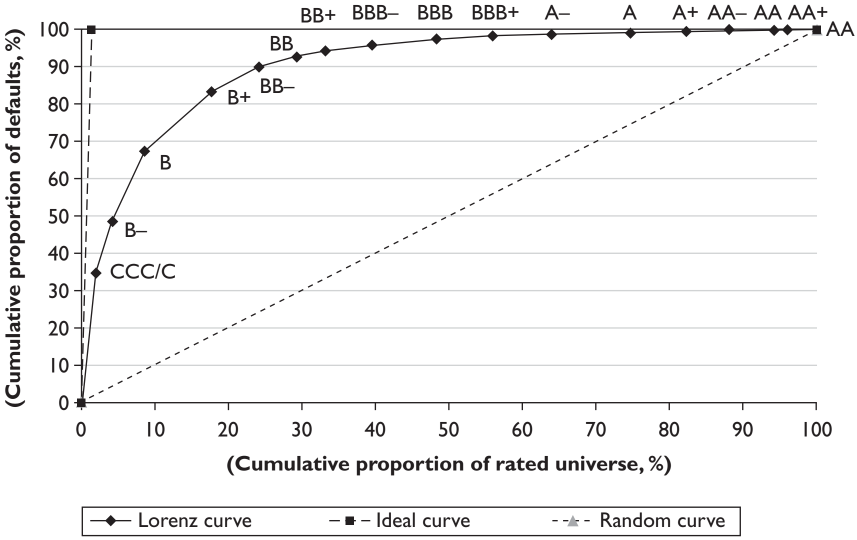
How do banks' SBRSs perform compared with agencies' ratings? Gini coefficients are sample-dependent measures. Moody's Investors Services (March 2000) is among a few studies that has benchmarked SBRSs on a

Table 13.6 Corporate Transition Matrices (1981–2008)—Five-Year Transition Rates (%)

From/To	AAA	AA	A	BBB	BB	B	CCC/C	D	NR
AAA	**54.23**	23.49	5.10	0.93	0.12	0.09	0.06	0.28	15.69
AA	1.75	**51.73**	23.52	4.08	0.60	0.36	0.04	0.30	17.62
A	0.12	5.92	**53.37**	15.23	2.44	0.95	0.17	0.68	21.11
BBB	0.05	0.78	10.84	**47.07**	8.28	2.91	0.52	2.57	26.99
BB	0.02	0.12	1.51	12.26	**28.12**	11.03	1.59	9.98	35.37
B	0.03	0.06	0.50	2.13	10.92	**20.83**	2.88	23.18	39.47
CCC/C	0.00	0.00	0.23	1.21	3.48	11.21	**3.33**	47.80	32.73

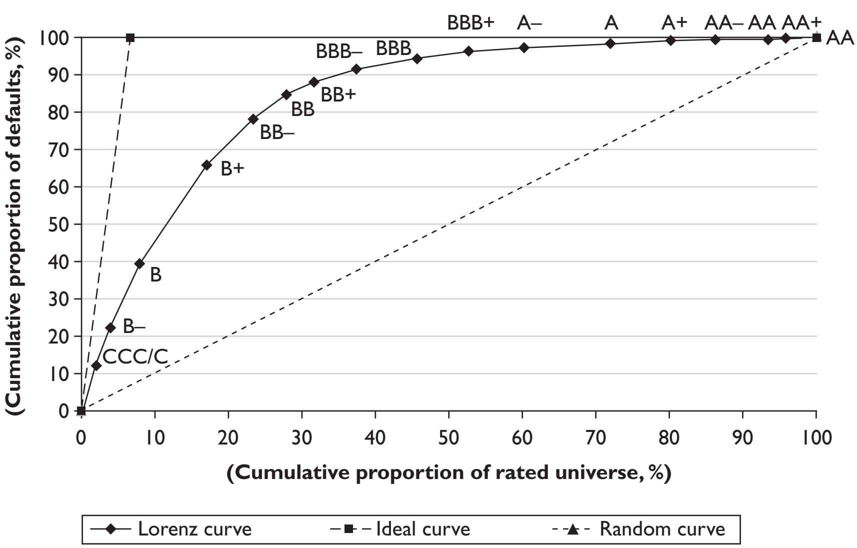
Source: Standard & Poor's, 2009.

Figure 13.1 Global One-Year S&P's Corporate Ratings Performance (1981–2008)



Source: Standard & Poor's, 2009.

Figure 13.2 Global Five-Year S&P's Corporate Ratings Performance (1981–2008)



Source: Standard & Poor's, 2009.

Table 13.7 Gini Coefficients for Global Corporates by Broad Sector (1981–2008)

	Time horizon (years)			
Sector	**1**	**3**	**5**	**7**
Average financial	78.53	72.43	66.08	61.81
Average nonfinancial	82.95	76.64	73.20	70.22

Source: Standard & Poor's, 2009.

common dataset (most of the tested models had one-year Gini ratio in the range of 50 to 75 percent for out-of-sample and out-of-time tests). The solution, envisaged by supervisory authorities, of creating a reference data set of borrowers to be used to benchmark rating systems performances is frustrated by the different structures of independent variables, in particular internal behavioral and qualitative data.

Qualitative data create the biggest problem for comparisons. If they are considered on a merely judgmental basis, it is not possible to replicate the rating on a large scale: this is the case for overrides of SBRS based on experts' judgments. In SBRSs qualitative data can be incorporated as nominal and ordinal variables, usually collected by closed-form questionnaires filled in by relationship managers. The issue of consistency in the treatment of qualitative information is greatly reduced when relationship managers are only required to fill in a questionnaire, but it still remains for questions with subjective answers and in presence of incentives schemes based on the amount of loan "sold" (as internal analysis conducted by banks show).

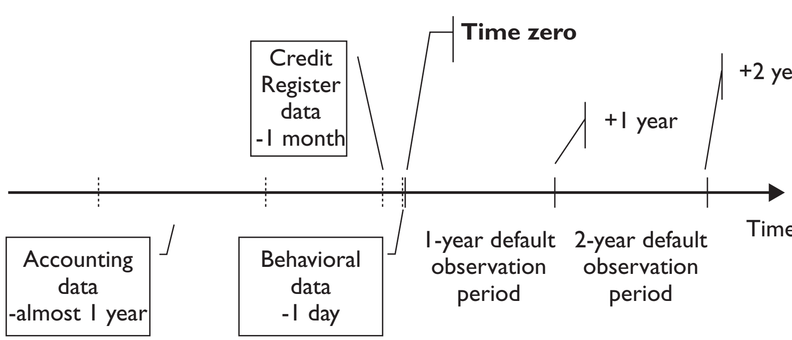
Qualitative data collected by questionnaires may participate to the final model in two different ways (De Lerma, Gabbi, and Matthias, 2007). The first approach is to use individual nominal and ordinal data as possible explanatory variables, together with quantitative data, in the estimation of the final algorithm representing the SBRS. This approach is rare because: (1) most qualitative data are either nominal or ordinal, so they are crowded out by variables selection procedures because of their low discriminatory power; (2) the collection of questionnaire-based qualitative data has been implemented only recently, so data sets are small; and (3) different types of quantitative data are not always available for all borrowers. This is why banks tend to use a second approach: a specific "qualitative module" is built; it produces a "qualitative rating" to be subsequently combined with "partial ratings" obtained by other "quantitative" modules (typically for financial statement, credit register, and internal behavioral data). Using this approach, more qualitative data indirectly enter into the SBRS. But the nature of qualitative data brings a low Gini ratio of the qualitative module. Thus, when combining partial ratings into the final algorithm, its relative relevance is low. The final result is that SBRSs do not leverage much on qualitative information.

Quantitative data entered into SBRSs may either derive from the past or represent economic forecasts. The use of forecast as explanatory variables in SBRS appears to be rare and, if present, it is more typical to include them in the qualitative module rather than to mix them up with historical, more objective, and certified data.

A key aspect is to get the time frame by which quantitative objective information are collected (Figure 13.3). "Time zero" is the point in time on which credit analysis is performed, so all information available at the moment can be incorporated into the model that will try to forecast if the borrower will go into default during the "observation period." Behavioral data produced by bank information systems is usually available the day after. Credit register data can be available according to the frequency of data distribution set up; in Italy, on average, it is available after one month. If annual financial statements are approved after a few months from year-end, they become available for credit analyses that take place, on average, about one year after year-end. Quarterly results shrink the lag.

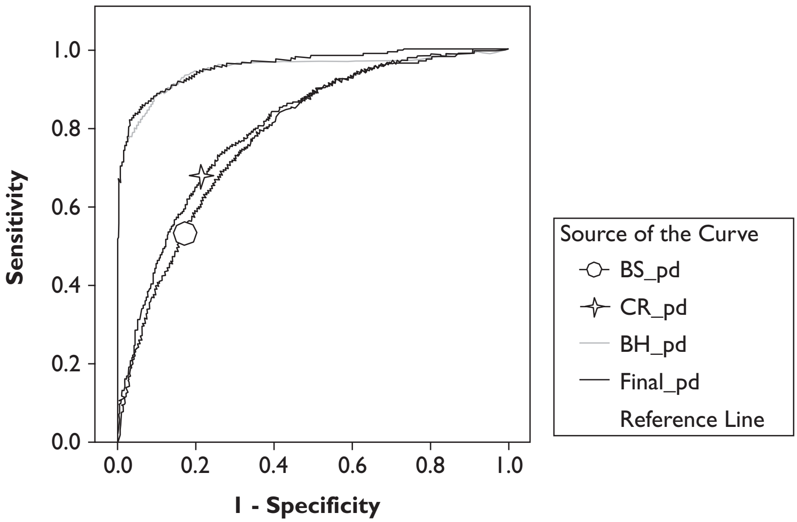
Different sources of quantitative objective information have a different freshness. Is it positively correlated with predictive power? Yes, if we consider a one-year observation period. In this case, the typical result you get when calculating ROC curves for accounting module, credit register module, and behavioral module, and then for the final model that combines them (using partial scorings as explanatory variables), is depicted in Figure 13.4. Behavioral module has the best performance among all partial modules and

Figure 13.3 Time Frame of a SBRS



SBRS, statistical-based rating system.

Figure 13.4 ROC Curves for Quantitative Modules of SBRS and the Final Model



Diagonal segments are produced by ties.

it is only slightly improved when combined with other modules in the final model. A final model that closely resembles results obtained from the behavioral module is a great source of model risk, in terms of short sightedness.

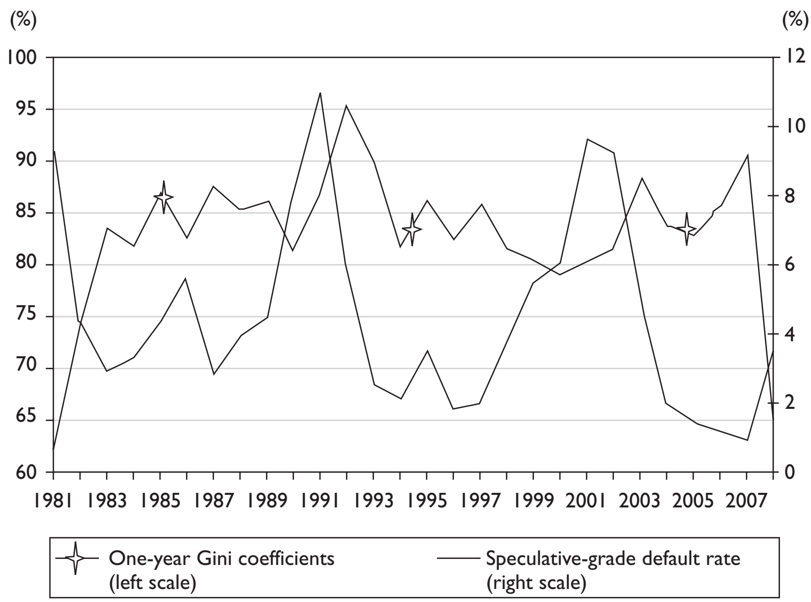
In fact, behavioral data suffers from two severe limitations: (1) they only reflect the current conditions of debts for the borrower (in other words, they reflect what is going on at the specific point in time, from the crude perspective of debt balances and cash flows, ignoring value creation and firm's potential); and (2) they are reflective information because they depend on a bank's own credit decisions: if a bank extends more credit to a given borrower, its debt behavior improves. At a lower level, also credit register data suffer for the same weaknesses. At the end, the final model is relying on point in time, short-term, and reflective information.

The majority of banks use a one-year observation period from time zero as the time frame for building SBRS to be used in daily operations, even if this is not what Basel II expects form them (§414 states that "although the time horizon used in PD estimation is one year, banks must use a longer time horizon in assigning ratings"). If a two-year or three-year period of observation from time zero is considered, behavioral and credit register data lose their apparent discriminatory power, and accounting data gain relevance in the final SBRS. On the other hand, it would result in much lower Gini ratios and/or AuROC.

The current bad state of the economy increases the demand for "banks to be bankers" and using a longer time horizon when assessing borrower creditworthiness. In fact, we know that the pattern of Gini coefficients for agency ratings appears to be broadly cyclical (Figure 13.5). In periods of economic stress, there is an increased likelihood of companies from across the rating spectrum suffering a more rapid deterioration of credit quality, which reduces the Gini ratio. In 2008, the one-year Gini ratio dropped to an all-time low of 65 percent, mainly attributable to extraordinary turbulence among global financials. Thus, strong discontinuities in the state of the economy worsen performance even in ratings that, being judgment based, can better accommodate a larger variety of factors into the rating process.

Do SBRSs perform better or worse than judgment-based ratings in periods of economic turnarounds? No robust analyses exist, because SBRS are very recent, dating back to only the late 1990s for a few banks and much later for other banks; and SBRS are continuously improved, often so much that is impossible to back-test them on old datasets. Theoretically, SBRSs are more point in time and based on a much smaller set of variables, so their performance probably lags behind judgment-based approaches in rapidly changing times. This is a further source of model risk.

Figure 13.5 Gini Ratio Cyclicality



Source: Standard & Poor's, 2009.

CONCLUSION

Model risk has many sources in credit processes, due to exogenous and endogenous factors to rating models. Among external ones, economic cycles, firm size, and industry reduce the aptitude of credit models to identify actual counterpart and concentration risk. Many empirical researches show that also regulatory approaches may fail to ease the model risk.

SBRSs are a great source of model risk, in particular in cases of relationship lending. When credit decisions concerning individual borrowers depend on SBRSs, their discriminatory capability and their calibration are required to hold at a satisfactory level, case by case, and on longer (than one-year) time horizons. Whereas, for transaction-based lending and for other applications, such as calculating bank provisions and capital adequacy, good Gini ratios on vast aggregates obtained by point in time rating systems can be sufficient.

REFERENCES

Allen, L., DeLong, G., and Saunders, A. (2004) Issues in the Credit Risk Modeling of Retail Markets. *Journal of Banking and Finance*, 28(4): 727–752.

Bank of Italy. (2006) Nuove disposizioni di vigilanza prudenziale per le banche no. 263, December 27.

Basel Committee on Banking Supervision. (2000) Range of Practice in Banks' Internal Ratings Systems. Discussion paper. Basel, Switzerland: BCOS.

Basel Committee on Banking Supervision. (2004) *International Convergence of Capital Measurement and Capital Standards—A Revised Framework*. Comprehensive version. Basel, Switzerland: BCOS.

Basel Committee on Banking Supervision. (2005) Studies on validation of Internal Rating Systems. Working paper 14, Basel, Switzerland.

Berger, A. N., Frame, W.S., and Miller, N.H. (2002) Credit Scoring and the Availability, Price and Risk of Small Business Credit. Federal Reserve System Working Papers, Federal Reserve Board.

Berger, A.N. and Udell L.F. (2001) Small Business Credit Availability and Relationship Lending: the Importance of Bank Organizational Structure. Federal Reserve System Working Papers, Federal Reserve Board.

Brunner, A., Krahnen, J.P., and Weber, M. (2000) Information Production in Credit Relationships. Working paper 10, Center for Financial Studies of University of Frankfurt, Germany.

Catarineu-Rabell, E., Jackson, P., and Tsomocos, D. (2005) Procyclicality and the New Basel Accord—Banks' Choice of Loan Rating System. *Economic Theory*, 26: 537–557.

Committee of European Banking Supervisors. (2005) Guidelines on the Implementation, Validation and Assessment of AMA and Internal Ratings Based Approaches. London: CEBS.

De Laurentis, G. (ed.). (2005) *Strategy and Organization of Corporate Banking*. Berlin: Springer Verlag.

De Laurentis, G., Saita, F., and Sironi, A. (eds.). (2004) *Rating Interni e Controllo del Rischio di Credito*. Rome: Bancaria Editrice.

De Lerma, M., Gabbi, G., and Matthias, M. (2007) CART Analysis of Qualitative Variables to Improve Credit Rating Processes. Mimeo, Limassol, Cyprus, Society for Computational Economics 12th International Conference on Computing in Economics and Finance.

Degryse, H. and Ongena, S. (2005) Distance, Lending Relationship and Competition. *Journal of Finance*, 9(1): 231–266.

DeYoung, R., Hunter, W.C., and Udell, G. F. (2003) The Past, Present and Probable Future for Community Banks. Working paper 14, Federal Reserve Bank of Chicago.

Diamond, D. (1984) Financial Intermediation and Delegated Monitoring. *Review of Economic Studies*, 51: 393–414.

Dietsch, M. and Petey, J. (2004) Should SME Exposures Be Treated as Retail or Corporate Exposures? A Comparative Analysis of Default Probabilities and Asset Correlations in French and German SMEs. *Journal of Banking and Finance*, 28(4): 773–788.

Engerlmann, B. and Rauhmeier R. (2006) *The Basel II Risk Parameters*. Berlin: Springer.

Financial Stability Forum. (2009) Report of the Financial Stability Forum on Addressing Procyclicality in the Financial System. Basel, Switerland: Financial Stability Board.

Gabbi, G. and Vozzella, P. (2009a) Default and Asset Correlation: An Empirical Study for Italian SMEs. University of Sienna Financial Economics Department Working Papers Series, n. 3/09.

Gabbi, G. and Vozzella, P. (2009b) Industry Factors to Explain Default and Asset Correlation. University of Sienna Financial Economics Department Working Papers Series, n. 5/09.

Gordy, M. and Howells, B. (2004) Procyclicality in Basel II: Can We Treat the Disease without Killing the Patient? Workshop "Accounting, Transparency and Bank Stability." Basel, Switzerland: BIS.

Moody's Investors Service. (2000) *Benchmarking Quantitative Default Risk Models: A Validation Methodology*. New York: Moody's.

Morris, J.A. (2001) Risk Diversification in the Credit Portfolio: An Overview of Country Practices. Working paper 01/200, International Monetary Fund.

Pederzoli, C. and Torricelli, C. (2005) Capital Requirements and Business Cycle Regimes. *Journal of Banking and Finance*, 29: 3121–3140.

Petersen, M.A. (2004) Information: Hard and Soft. Working paper, Kellogg School of Management, Northwestern University.

Petersen, M. and Rajan, R. (2002) Does Distance Still Matter? The Information Revolution in Small Business Lending. *Journal of Finance*, 57(6): 2533–2570.

Shen, T.P. (2005) *Empirical Study on the Effect of New Basel Accord on SMEs*. Taipei, Taiwan: Joint Credit Information Center.

Standard & Poor's. (2009) *Default, Transition, and Recovery: 2008 Annual Global Corporate Default Study and Rating Transitions*. New York: Standard & Poor's.

Treacy, W.F. and Carey, M.S. (1998) Credit Risk Rating at Large U.S. Banks. Federal Reserve Bulletin, Federal Reserve Board.

CHAPTER 14

NEGLECTING CASH FLOW INFORMATION IN STRUCTURAL CREDIT PORTFOLIO MODELS—A SHIPPING PORTFOLIO EXAMPLE

Torsten Seil and Florian Heitger[1]

ABSTRACT

Choosing an appropriate approach for taking into account loss dependencies in loan portfolios is a well-known, but also a challenging problem in modeling credit portfolio risk. In our chapter, we investigate model risk in terms of loss dependencies by using two different structural credit portfolio models, namely, a standard Merton-style model and an augmented approach which incorporates a cash flow default trigger in addition to the standard default trigger based on asset values. A rigorous simulation analysis is conducted in order to observe the impact of this source of model risk on the portfolio's expected shortfall and value at risk statistics. We base our study on a generic portfolio of shipping loans because asset value and cash flow data are comparatively well observable in the shipping market. Our results have implications on the widely used structural model approach in practice.

INTRODUCTION

Structural portfolio models are based on the so-called Merton approach which was pioneered by Robert C. Merton (1974) more than 35 years ago, who introduced the idea of using an asset value, or structural model, to evaluate credit risk. In his eminent contribution, Merton proposed to apply the Black-Scholes option model to the capital structure of a company and identifies its equity holders as the owner of a European call option on the company's assets with a strike price equal to the book value of the company's debt. This is why structural models are also called option-theoretic or contingent-claim models. If the asset value of the company falls below its debt level at option's maturity, the company is in default. In this framework, default is a function of both the asset value and the liability structure of a company and, thus, can be solely derived from the evolution of the company's structural variables. The Merton model provides a very useful framework for modeling credit risk and is widely accepted and used by financial market participants.

Without denying the outstanding relevance of Merton's contribution to the credit risk literature there is, however, also some basic criticism on his proposed model approach. It mainly refers to the rather simplified assumptions with respect to two aspects in the original publication, namely, (1) the assumed capital structure which implies a single default trigger and (2) the default timing which is solely restricted to the maturity of the company's debt. Concerning the latter aspect, one of the first generalizations of the Merton framework was proposed by Black and Cox (1976). In their model variant a company's default can occur in principle at any time during the company's lifetime. More precisely, the company's default time is defined as the point in model time at which the asset value falls below a certain ("default") threshold.[2] While in the meantime a lot of literature can be found that concentrates on the problem of how to determine the "correct" default threshold with respect to aspect (1), on the contrary, the de facto standard approach still assumes that defaults are exclusively triggered by the company's assets which have to be below such a threshold. However, one recent contribution to the question, whether several default triggers should be considered instead, is proposed by Davydenko (2007). He studies whether default is caused by low market asset values or by liquidity shortages which he denotes as economic distress and financial distress, respectively.

According to this terminology, on the one hand, it can be argued that a company's default is driven by economic rather than by financial distress, because typically shareholders will be willing to balance a temporary cash short fall by raising external financing as long as the asset value remains above the aforementioned boundaries. This argument implies that purely financial distress becomes more or less irrelevant. However, on the other hand, financial distress does appear to be a significant variable by all means, even though empirical analysis emphasizes that the asset value is the most important determinant of companies' default behavior.[3]

The well-known procedure to generate correlated asset price developments across a pool of obligors establishes a straightforward method to expand the standard structural credit model approach for single borrowers outlined above to a structural portfolio model in which correlated defaults of obligors are simulated. The relatively easy implementation as well as the well-founded underlying theory probably make structural portfolio models the most popular approach of modeling bonds and bank loans today. For instance, rating agencies use such a model framework to valuing synthetic collateralized debt obligations, so-called CDOs, which are securities linked to a pool of bonds or loans (see the rating tools named CDOEvaluator, CDOROM or the PCM Suite, formerly VECTOR model, provided by Standard & Poor's, Moody's, and Fitch, respectively). Regulators and bank portfolio managers also apply similar models for assessing the credit risk of portfolios.

However, as already indicated above for the case of credit risk models of single borrowers, these structural portfolio models used by the market participants typically concentrate on economic rather than financial distress in terms of the terminology proposed by Davydenko (2007). The main reason for neglecting the influence of cash shortages on the default characteristic of a portfolio within the "classical" structural model approach might be the inaccessibility of the relevant data, which, however, will lead to an overestimation of risk in general and, as a consequence, to a misallocation of capital.

In our contribution we will investigate the bias that arises from using a classical structural model approach versus considering cash shortages as an extra default trigger in the proposed model framework, which we understand as a sort of model risk in this context. We conduct our study by using a generic portfolio of shipping loans since asset value and cash flow data can be comparatively well observed in the shipping market. The reason for the relatively good availability of those data is the following. Shipping loans are typically granted to single-purpose companies (SPCs). Thus, the asset value of each company can be identified with the second-hand value of the financed vessel and charter income makes up the only cash flow to meet the debt service. Since we will exclusively deal with standardized ship types (e.g., container carriers, bulkers, and tankers), the development of the second-hand values and charter rates of a benchmark ship can be derived from historical time series which are available from leading ship brokers, who publish those worldwide valid market data.

Since we want to concentrate our discussion on portfolio modeling risk (with respect to cash shortage as an additional default trigger), we assume that individual probabilities of default (PDs) and losses given default (LGDs) for borrowers are exogenously given in our portfolio. This coincides with the usual treatment of credit portfolio risk in practice where organizational separation of rating development and portfolio modeling is typical, even though not optimal.[4]

MODEL OUTLINE

Asset Value and Cash Flow Representation

For our "classical structural model" (CSM), which will be our benchmark model in the following, we use a structural model approach where the default of a company is driven by a single latent variable, i.e., its asset return. Correlation in asset returns is modeled by systematic factors ("risk drivers"), which describe the current state of the shipping market. For this reason, we divide the shipping market into $n = 1, \ldots, 11$ shipping subsegments where each company of our generic portfolio can be assigned to exactly one of them.[5] The asset return $A_{i,t}$ of company i at time t can be written as

$$A_{i,t} = \sqrt{d}\left(\sqrt{\rho_{n,A}}\Psi_{n,t}^{(A)} + \sqrt{1-\rho_{n,A}}\varepsilon_{i,t}^{(A)}\right)/\sqrt{X}_t, \ \mathrm{Cov}\,(\varepsilon_{i,t}^{(A)}, \varepsilon_{j,t}^{(A)}) = 0 \text{ for all } t \text{ with } i \neq \mathrm{j},$$

$$\mathrm{Cov}\,(\Psi_{n,t}^{(A)}, \varepsilon_{i,t}^{(A)}) = 0 \text{ for all } n,i,t, \quad \Psi_{n,t}^{(A)} \sim N(0,1) \text{ for all } n,t, \quad \varepsilon_{i,t}^{(A)} \sim N(0,1) \text{ for all } i,t,$$

where $\Psi^{(A)}$ denotes the systematic risk associated to the shipping subsegment and $\varepsilon^{(A)}$ denotes the idiosyncratic risk. Moreover X is a χ^2-distributed random variable with d degrees of freedom (independent of the $\Psi^{(A)}$s and $\varepsilon^{(A)}$s), which leads to a Student-t distribution for the asset return $A_{i,t}$. The Student-t distribution of $A_{i,t}$ is motivated by an analysis of historic data of second-hand values provided by Clarksons,[6] which indicate for a fat-tail distribution of the log-differences of the time series. Note, that in the case of SPC financing in the shipping industry, it is straightforward to identify the asset value of each company with the second-hand value of the financed vessel, because the ship is the only asset on the balance sheet. Hence, we can assume that the asset value return is mainly determined by the log-differences of the second-hand values of the subsegment specific benchmark ship represented by the risk driver $\Psi^{(A)}$ in our model equation. Note further, that by scaling the standard normal distributed random variable $\left(\sqrt{\rho_{n,A}}\Psi_{n,t}^{(A)} + \sqrt{1-\rho_{n,A}}\varepsilon_{i,t}^{(A)}\right)$ by the factor $\sqrt{d/X_t}$, we use in fact a Student-t copula representation which shows tail dependence in contrast to the standard Gaussian copula.[7] Moreover it is important to remark that a further look to the Clarksons data shows a high correlation of log-returns of the second-hand values of vessels across different shipping subsegments. Thus, the $\Psi^{(A)}$ are assumed to be highly correlated across the 11 subsegments in our model framework. Small variations in the ship's equipment, its technical standard, and its overall condition within the respective subsegment lead to the idiosyncratic influence $\varepsilon^{(A)}$. Since an empirical evidence for the parameters $\rho_{n,A}$ is hard to obtain (and since our contribution does not aim to focus on the uncertainty of model parameter estimation), we use the average intrasegment correlation as a proxy for the factor loading parameter $\rho_{n,A}$.[8]

Our "augmented structural model" (ASM) is an extension of our benchmark model and also takes the incoming cash flows of each company into account. Here, the charter income of the vessels serves as a proxy for the financial situation of the company. Such information is also provided for the $n = 1,\ldots,11$ different shipping subsegments by Clarksons.[9] In the ASM we proceed in a similar way as for the asset return $A_{i,t}$ outlined above and assume an additional structural model equation for the evolution of the cash flow situation of each company. More precisely, we denote the log-differences in the companies' cash flows by $C_{i,t}$ which evolves according to $C_{i,t} = \sqrt{d}\,(\sqrt{\rho_{n,C}}\,\Psi^{(C)}_{n,t} + \sqrt{1-\rho_{n,C}}\,\varepsilon^{(C)}_{i,t})/\sqrt{X_t}$, $\mathrm{Cov}(\varepsilon^{(C)}_{i,t}, \varepsilon^{(C)}_{i,t}) = 0$ for all t with $i \neq j$, $\mathrm{Cov}(\Psi^{(C)}_{n,t}, \varepsilon^{(C)}_{n,t}) = 0$ for all n,i,t, $\Psi^{(C)}_{n,t} \sim N(0,1)$ for all n, t, $\varepsilon^{(C)}_{i,t} \sim N(0,1)$ for all i, t, where again the risk drivers $\Psi^{(C)}$ are correlated via intercorrelations across the shipping subsegments and the factor loading parameters $\rho_{n,C}$ are given by the average intrasegment correlation of the logarithmic changes of the charter rates. Moreover, the $\Psi^{(C)}$ are correlated with the former introduced $\Psi^{(A)}$. Thus, in summary, the ASM consists of the combination of two dependent structural model components, namely, on the one hand, our benchmark CSM describing the asset return of the shipping companies, and a second structural model part, for which the companies' cash flow changes are chosen as an additional latent variable.

So far, neither the number of degrees of freedom for the marginal Student-t distributions of asset returns $A_{i,t}$ and cash flow changes $C_{i,t}$ nor the correlation assumptions for the risk drivers $\Psi^{(A)}$ and $\Psi^{(C)}$ are specified. To become able to estimate the number of degrees of freedom for the distributions of the variables $A_{i,t}$ and $C_{i,t}$ from a relatively large data set (as well as to simplify matters), we assume the same number of degrees of freedom for all latent variables and obtain an estimate of approximately $d = 3$ by fitting all corresponding historic data to one single Student-t distribution. The dependence structure for the risk drivers $\Psi^{(A)}$ and $\Psi^{(C)}$ are determined by estimating the Spearman correlation coefficients for the log-returns of the second-hand values, the log-differences of charter rates as well as the dependence between these two sets of variables among the 11 shipping subsegments by using historic Clarksons time series. Table 14.1 shows the average segment correlations calculated from the original (22×22) matrix. The relatively high correlation across the risk drivers in shipping can clearly be detected.

Default and Loss Identification

Assuming a rating for each company and consulting a cumulative default table published by a rating agency,[10] the (conditional) default probabilities $PD_{i,t}$ of company i in year t, given that no default has occurred before, are derived. Afterward, the $PD_{i,t}$ are transferred into default thresholds TH which trigger a default in our ASM in time step t whenever $A_{i,t} < TH_{A,i,t} \wedge C_{i,t} < TH_{C,i,t}$ holds. The thresholds are quantiles of the (correlated) distribution of $A_{i,t}$ and $C_{i,t}$ so that $\Pr(A_{i,t} < TH_{A,i,t}, C_{i,t} < TH_{C,i,t}) = PD_{i,t}$. The pairs of $TH_{A,i,t}$ and $TH_{C,i,t}$ are not unique because a reduction of one threshold can be compensated by an appropriate increase of the other threshold in order to hit the same $PD_{i,t}$. Obviously this provides the possibility to allow for differentiating between individual loan to values and debt service coverage ratios of borrowers in the ASM framework.[11] However, we will abstract from such distinction in our analysis in the following and define $TH_{A,i,t} = TH_{C,i,t} := TH_{Condition\_AC,i,t}$, which leads to $\Pr(A_{i,t} < TH_{Condition\_AC,i,t}, C_{i,t} < TH_{Condition\_AC,i,t}) = PD_{i,t}$. Monte Carlo simulation techniques are employed to derive the thresholds for the analysis.

In our CSM approach the default thresholds, i.e., quantiles, are determined by $\Pr(A_{i,t} < TH_{Condition\_A,i,t}) = PD_{i,t}$. Since the $A_{i,t}$ are one dimensional Student-t distributed with $d = 3$ degrees of freedom, the thresholds can be calculated using the inverse of the respective cumulative distribution function.

Table 14.1 Segment Averages of the Correlations of Subsegment Specific Risk Drivers

		Log-differences of Short-term Charter Rates			Log-differences of Second-hand Values		
		Container	Bulker	Tanker	Container	Bulker	Tanker
Log-differences of Short-term Charter Rates	Container	87%					
	Bulker	52%	86%				
	Tanker	1%	24%	80%			
Log-differences of Second-hand Values	Container	68%	47%	19%	80%		
	Bulker	43%	72%	26%	49%	76%	
	Tanker	23%	29%	71%	41%	43%	84%

ANALYSIS

Our analysis is based on a generic portfolio consisting of 550 loans equally sized and evenly spread across the 11 shipping subsegments. Furthermore, the rating of each loan is assumed to be BBB-, where the LGD is set to 30% of the loans' initial outstanding and each loan matures in five years.[12] As already mentioned above, only SPCs are financed in our portfolio which are characterized by owning one single ship. Moreover, we assume that all vessels in the portfolio operate in the short-term charter market. Moreover, additional securities such as guarantees from third parties or any other risk mitigating factors do not exist for the loans in our portfolio.

Table 14.2 shows our base case simulation results. We note the following:

- Statistics on the left-hand side are simulated with the CSM, in which defaults are solely determined by the asset value triggers $TH_{Condition\_A,i,t}$. The results of the ASM, which uses asset value and cash flow triggers $TH_{Condition\_AC,i,t}$, are presented on the right-hand side of the table.
- The portfolio expected loss (EL) is the sum of the expected losses of the individual borrowers and is therefore independent of the chosen dependence structure. Consequently, the portfolio ELs are identical in both model approaches, because the different thresholds $TH_{Condition\_AC,i,t}$ and $TH_{Condition\_A,i,t}$ ensure that in both of the two models the same desired $PD_{i,t}$ are simulated.
- The correlation assumption of the asset and cash flow variables (along with the choice of the Student-t copula dependence structure) determines the portfolio risk, which we measure by means of expected shortfall (ES) and value at risk (VaR).
- Taking into account the additional cash flow triggers vs. triggering defaults only by the asset values, lowers the measured portfolio risk as

Table 14.2 Base Case Analysis

	$TH_{Condition\_A,i,t}$	$TH_{Condition\_AC,i,t}$
Expected Loss	1.09%	1.09%
ES(95%)	12.86%	12.21%
ES(99%)	21.23%	19.89%
VaR(95%)	7.09%	7.09%
VaR(99%)	16.91%	15.82%

ES, expected shortfall; VaR, value at risk.

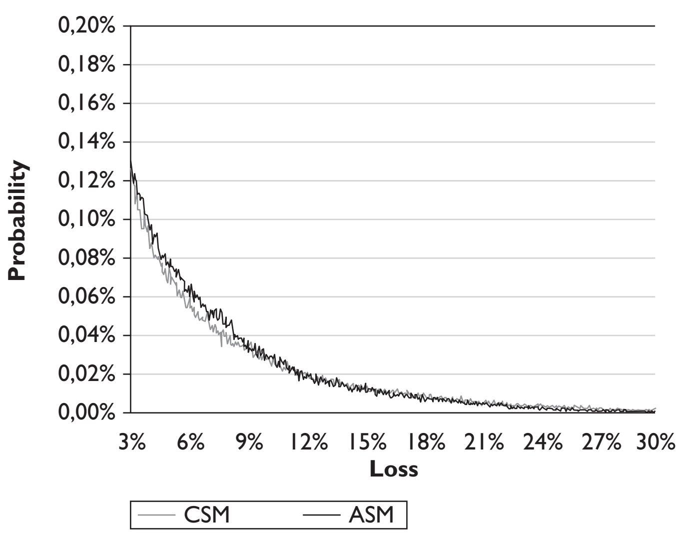
expected. In fact, small differences in the ES figures can be observed, e.g., 0.65 percent for ES(95%) and 1.34 percent for ES(99%).

- If cash flow information is included, the VaR calculated at the 99 percent quantile changes by 1.09 percent (from 16.91 to 15.82 percent). However, the simulation results for the VaR(95%) are identical for both model variants.
- In sum, the differences between both model approaches are relatively small. The reason for this finding is the high correlation of cash flows and asset values, which leads to the fact that incorporating a liquidity dimension for the default trigger of the borrowers adds only limited extra information.
- The similarity of the results can also be seen in Figure 14.1. Both graphs describe the loss distributions from which the portfolio risk statistics summarized in Table 14.2 have been calculated. More precisely, the graphs are histograms of the simulated relative portfolio losses (*x* axis), where the frequencies are divided by the total number of runs and hence can be interpreted as probabilities (*y* axis).

The high dependence of risk drivers is typical for the shipping industry. But the existence of long-term charter contracts reduces the correlation in shipping loan portfolios significantly.

In order to investigate this effect in further details, we lower the intra cash flow correlations and the correlation between cash flows and asset

Figure 14.1 Base Case Loss Distributions



values stepwise to 75, 50, and 25 percent of their initial values. Table 14.3 presents the results:

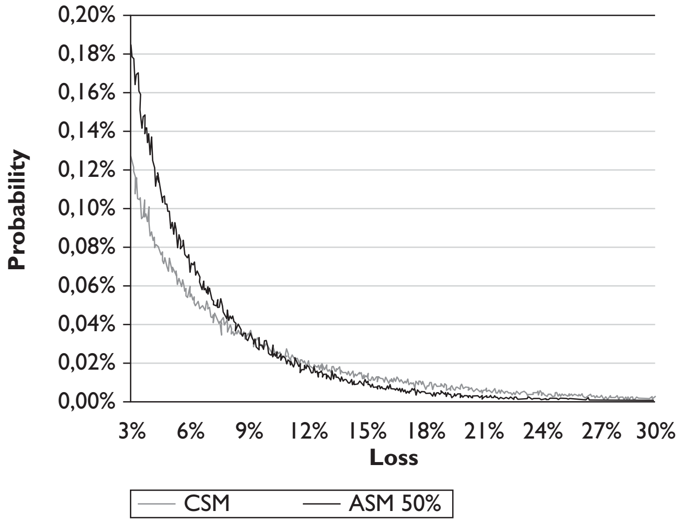
- The statistics are based on the ASM approach. Numbers in parentheses indicate the difference in comparison with CSM results from Table 14.2. Note that intra asset value correlations are not changed in our sensitivity analysis, which leave the CSM results unchanged.
- Since the left-hand side of Table 14.3 starts with the 100 percent correlation level, it shows the base case investigation already known from Table 14.2, while the other columns exhibit results for the lowered intra cash flow and inter cash flow/asset value correlations. We observe that the portfolio risk in terms of ES and VaR shrinks significantly. For instance, if the correlation level is only 25 percent of its initial value the ES(99%) will decrease from 19.89 to 14.10 percent in the ASM. This means, that the CSM results overestimate the ES(99%) by 7.13 percent for this case.
- Fifty percent is a realistic assumption for vessels in shipping loan portfolios to be long-term employed. This lowers the average intra cash flow correlation as well as the correlation between cash flows and asset values by about 50 percent.[13] Using the correlation sensitivity analysis as a proxy for the impact on the measured portfolio risk, we overestimate the ES(99%) by 4.99 percent in the CSM due to the fact that the cash flow default trigger is ignored.
- Figure 14.2 exhibits corresponding loss distributions derived from histograms of the simulation runs (see Figure 14.1 for further details). It is well observed that the probability to simulate portfolio losses below 10 percent (above 10 percent) is significantly larger (smaller) with the 50 percent reduced correlations used in the ASM 50% than in the CSM.

Table 14.3 First Correlation Sensitivity Analysis, Starting from Original Overall Correlation Level

Correlation	100%	75%	50%	25%
ES(95%)	12.21% (0.65%)	11.25% (1.61%)	10.18% (2.68%)	9.03% (3.83%)
ES(99%)	19.89% (1.34%)	18.22% (3.01%)	16.24% (4.99%)	14.10% (7.13%)
VaR(95%)	7.09% (0.00%)	6.65% (0.44%)	6.22% (0.87%)	5.78% (1.31%)
VaR(99%)	15.82% (1.09%)	14.40% (2.51%)	12.76% (4.15%)	11.18% (5.73%)

ES, expected shortfall; VaR, value at risk.

Figure 14.2 Loss Distributions of the First Sensitivity Analysis



In our first sensitivity analysis outlined above, the intra asset correlations were held constant and, thus, were remained at their relatively high original level. Furthermore, even the reduction of the remaining coefficients left the overall dependence in the portfolio comparatively strong.

However, other pools, for instance, a corporate debt portfolio, are characterized by less joint behavior. Therefore, our next investigation will provide an approximation of the effect of neglecting cash flow information in rather weakly correlated portfolios. For this purpose all correlations, including the intra asset value correlations, are lowered by 50 percent in the first step. In the second step, the intra cash flow and inter cash flow/asset value correlations are again decreased exactly in the same way as for the sensitivity analysis described before. Table 14.4 contains our findings:

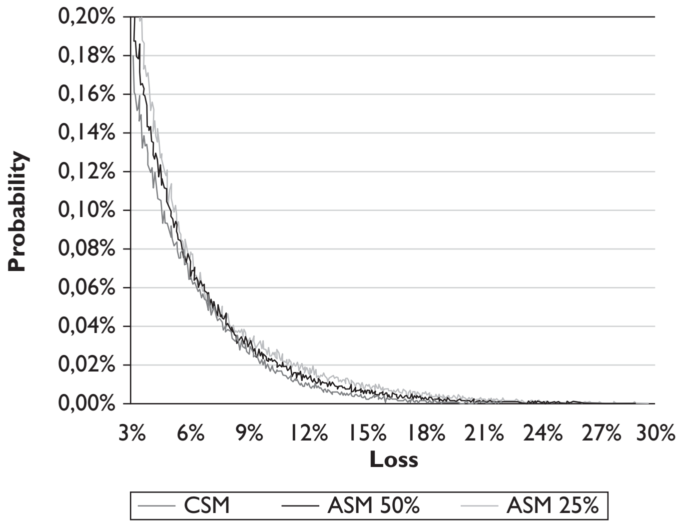
- All results shown in Table 14.4 are simulated with the ASM. Values in parentheses show the mistakes that are made when the CSM approach would be applied.
- Note that the results of the CSM have changed compared to Table 14.2 (and Table 14.3), since the intra asset value correlations have been manipulated.
- The first results on the left-hand side of the table are based on the initial reduction of all correlations to 50 percent of their original level. The decrease of all correlations makes the discrepancy between the CSM and ASM more significant than in our base case analysis

Table 14.4 Second Correlation Sensitivity Analysis, Starting from Reduced Overall Correlation Level

Correlation	50%	37.5%	25%	12.5%
ES(95%)	9.35% (1.29%)	8.87% (1.77%)	8.29% (2.35%)	7.75% (2.89%)
ES(99%)	14.85% (2.20%)	13.91% (3.14%)	12.82% (4.23%)	11.73% (5.32%)
VaR(95%)	5.84% (0.54%)	5.67% (0.71%)	5.40% (0.98%)	5.18% (1.20%)
VaR(99%)	11.67% (1.80%)	10.91% (2.56%)	10.15% (3.32%)	9.38% (4.09%)

ES, expected shortfall; VaR, value at risk.

Figure 14.3 Loss Distributions of the Second Sensitivity Analysis



shown in Table 14.2. For instance, we observe a deviation of 2.20 percent instead of 1.34 percent by neglecting the cash flow default trigger for ES(99%).

- Decreasing the intra cash flow and inter cash flow/asset value coefficients leads to an increasing bias as already observed in our investigation shown in Table 14.3. However, the absolute and relative biases get smaller in comparison to our first sensitivity analysis, because the incremental diversification effects by the cash flows get exhausted in a lower overall correlation environment.
- Figure 14.3 shows the loss distributions for the second sensitivity analysis. The graphs illustrate a gradual reduction of the probability to simulate extreme portfolio losses. High portfolio losses are

simulated more frequently by the CSM compared with the ASM 50% when starting the sensitivity analysis from the reduced overall correlation level. After the extra decrease of intra cash flow and inter cash flow/asset value correlations in the ASM framework (i.e., ASM 25%) the probability of high losses is further reduced.

FURTHER REMARKS

In contrast to the Student-t copula, the Gaussian copula is missing tail dependence. Lacking the property of tail dependence seems especially questionable with regard to the experience in the current financial crisis, in which the behavior of risk factors appears highly dependent. But also events like the September 11 terror act have provided important insight into the dependence mechanics of industry sectors when a shock induces sharply increasing correlations. To conclude, it can be stated that a realistic credit portfolio model should not exhibit asymptotic independence like the Gaussian copula approach does (Bluhm and Overbeck, 2007).

It is well known that applying a Student-t copula (with a low number of degrees of freedom) instead of a Gaussian copula has a large impact on the loss distribution. A final analysis will show these differences on the basis of the CSM. See Table 14.5 for the results:

- All results shown in Table 14.5 are simulated with the CSM.
- We observe striking differences, i.e., an almost doubled risk statistic for the Student-t copula compared to the simulation results obtained for the Gaussian copula.
- The significant differences are also reflected in the loss distributions depicted in Figure 14.4. Portfolio losses above 7 percent are triggered with significantly higher probability in case of the applied Student-t copula compared with the Gaussian copula.

CONCLUSION

Structural portfolio models are widely used in credit portfolio risk modeling today. Typically, the rating information of the borrowers is a model input in these types of portfolio models, because organizational separation of rating development and portfolio modeling is common in financial institutions, even though not optimal.

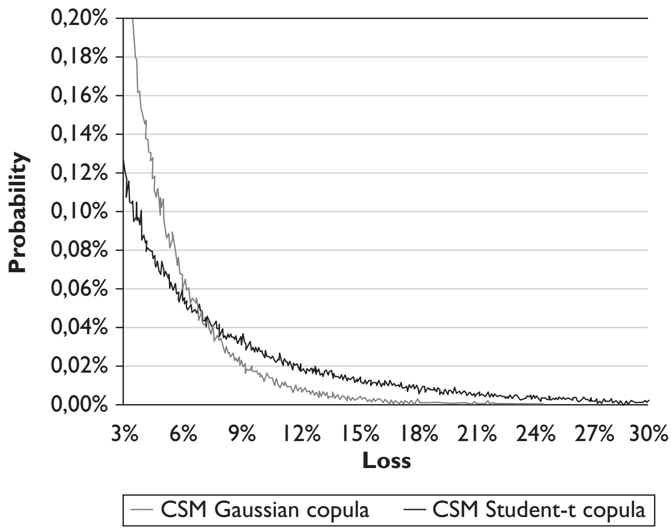
Structural models mainly concentrate on economic distress rather than on financial distress and, thus, neglect the correlation reducing effect of the borrowers' liquidity endowment in a loan portfolio. Ignoring such cash flow

Table 14.5 Different Copula Choices in the Classical Structural Model Approach

Correlation Reduction	Gaussian Copula $TH_{Condition\_A,i,t}$	Student-t Copula $TH_{Condition\_A,i,t}$
ES(95%)	7.67%	12.86%
ES(99%)	12.10%	21.23%
VaR(95%)	4.96%	7.09%
VaR(99%)	9.33%	16.91%

ES, expected shortfall; VaR, value at risk

Figure 14.4 Loss Distributions Resulting from Different Copula Choices



information leads to a significant overestimation of risk, if the overall level of correlation of cash flows and asset values remains moderate. But even in shipping loan portfolios, where cash flows are strongly correlated with asset values, and hence do not provide considerable extra information, a bias persists. Our example of a shipping loan portfolio suggests that the mistakes that are made by ignoring the liquidity situation of the borrowers become highly significant in the presence of long-term employment contracts because those contracts usually stabilize the income and hence reduce the intra cash flow as well as the inter cash flow/asset value dependence.

In accordance with our findings from historic shipping time series, we apply a Student-t copula in our ASM framework, which is characterized by tail dependence in contrast to a Gaussian copula. It is a broad consensus that modeling the correlation behavior of risk drivers for extremely rare events in structural portfolio models is a challenging but also an essential task which seems to us to become apparent, in particular, in the current financial crisis. But on the way to use more extreme dependence structures in future applications of credit portfolio models in practice, the necessity to allow for incorporating diversification effects in the model framework seems to be indispensable in order to ensure loss distributions in line with empirical evidence. Moreover, our investigation indicates that, on the one hand, cash flow information is an obviously missing diversifying parameter in structural portfolio models and, on the other hand, augmenting the standard approach with respect to this model feature is straight forward and relatively easy to implement.

REFERENCES

Black, F. and Cox, J.C. (1976) Valuing Corporate Securities: Some Effects of Bond Indenture Provisions. *Journal of Finance*, 31: 351–367.

Bluhm, C. and Overbeck, L. (2007) *Structured Credit Portfolio Analysis, Baskets & CDOs*. London: Chapman & Hall.

Davydenko, S.A. (2007) When Do Firms Default? A Study of the Default Boundary. AFA 2009 San Francisco Meetings Paper; EFA 2005 Moscow Meetings Paper; WFA 2006 Keystone Meetings Paper. Available at: http://ssrn.com/abstract=672343.

Elizalde, A. (2006) Credit Risk Models II: Structural Models. CEMFI Working paper 0606.

Merton, R. (1974) On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *Journal of Finance*, 29: 449–470.

NOTES

1. This article reflects the personal view of the authors and does not provide any information about the opinion of HSH Nordbank AG. The article has been written solely for academic purposes and should be read on this note.
2. See Elizalde (2006) for an overview, criticism and extensions of structural models in credit risk modeling.
3. See Davydenko (2007) for a detailed literature review on default triggers as well as for an empirical study of cash flow and asset value related default boundaries.
4. The typical organizational separation of loan origination and portfolio management activities in bank applications as well as the very

heterogeneous asset classes in the banks' overall portfolio are arguments for a clear organizational/process-related interface. But also the rating process of CDOs is a good example for this stepwise procedure: in the first step, a rating is determined for the individual assets of the underlying portfolio; in the second step, standard portfolio models are used to evaluate the portfolio risk of the CDO transaction. See Bluhm and Overbeck (2007) for a discussion on this topic.

5. We model the following subsegments: (1a) panmax, (1b) subpanamax, (1c) handymax, and (1d) handy container vessels; (2a) capsize, (2b) panamax, (2c) handymax, and (2d) handy bulker; (3a) VLCC, (3b) suezmax, and (3c) aframax tanker, which total 11 subsegments. See also the section "Analysis" for a description of the portfolio.
6. Clarksons is one of the leading brokers in the shipping industry and publishes historic time series of second-hand values, charter rates, etc. For our analysis we use data comprising a period from 1997 to 2008.
7. See Bluhm (2007) for a more detailed discussion on the construction of the standard structural model approach with different types of copula functions.
8. For example, the average second-hand ship value correlation between the container vessel subsegments.
9. Note that the used data set comprises the same time period as already mentioned above.
10. We use the cumulative default table for corporates provided by Fitch for demonstrative reasons.
11. The loan to value (LTV) and the debt service coverage ratio (DCSR) are commonly used risk statistics in object financing. The LTV is calculated by dividing the financed ship value by the outstanding debt amount of a single purpose company. Hence, a LTV below 100% indicates sufficient loan collateral as a rule of thumb. Dividing the income from operating the vessel by the sum of due interest and loan repayment in a certain time period gives the DSCR. Thus a DSCR below 100% indicates a problematic cash flow situation of a borrower.
12. Note that the rating assumption as well as the assumption on the LGD and the maturity of the loans is chosen arbitrary due to the fact that we consider a generic portfolio. Nevertheless, these characteristics of the loans seem to us to be in a range which are observable in real shipping portfolios.
13. The decrease in correlations has been estimated by assigning randomly chosen sequences of zeros to a simulation sample of the borrowers' cash flow changes in our portfolio and subsequently compare the new correlations with the original ones on average.

PART IV

MODEL RISK RELATED TO VALUATION MODELS

This page intentionally left blank

CHAPTER 15

CONCEPTS TO VALIDATE VALUATION MODELS

Peter Whitehead

ABSTRACT

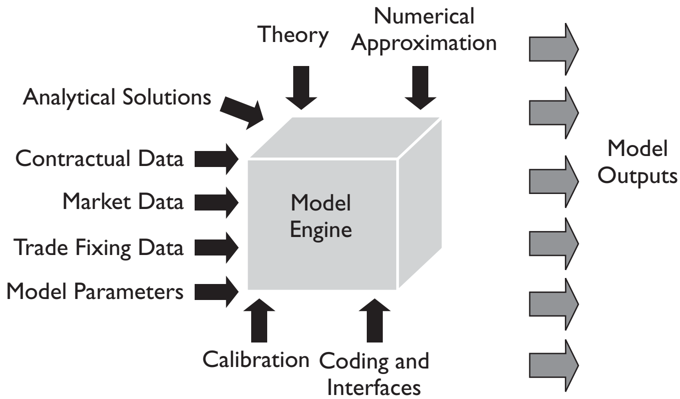
The independent review and validation of front office pricing models is a key component of any framework attempting to mitigate model risk. A number of complementary approaches can be used to validate valuation models and this chapter details these different concepts.

INTRODUCTION

A valuation (or pricing) model can be considered as a mathematical representation which is implemented within a trading and risk management system and is used to map a set of observable market prices for its liquid calibrating instruments to the price of an exotic product. At a basic level, a pricing model can be considered as having three components, namely, the input data, the model engine, and the output data as represented pictorially in the Figure 15.1. All three model components are possible sources of model risk which need to be addressed through the model validation process and the concepts explained in this chapter.

On one level, pricing models are theoretical constructs relying on a number of mathematical equations and assumptions. The first step in any attempt at validating valuation models should naturally start with a review of the theory underpinning the model and a re-derivation of all equations and theoretical results to ensure that no errors have been made in the theoretical specification of the model. A model cannot be reviewed in isolation

Figure 15.1 The Pictorial Representation of Models



from the product which it will be used to value, and the adequacy of the modeling framework for the product also needs to be considered as part of this step. Any contractual features of the product which are not captured by the model should be highlighted together with any relevant risk factors not being modeled. The use of incorrect dynamics and distributional assumptions for one or more of the underlying variables may also render the modeling approach inapplicable for the product under consideration. This review of the theoretical aspects of the model is arguably the most understood concept involved in validating pricing models. However, ensuring the correct transformation of a valuation model from its theoretical construct to its practical implementation within a trading and risk management system necessitates the consideration of a number of other validation concepts which are not all as familiar.

CODE REVIEW

Code review can be a contentious topic in the independent validation of valuation models with practitioners often divided over the usefulness of carrying out a line-by-line examination of the pricing model code in order to identify implementation errors. Detractors often assert that there is little value in such an exercise since the same results can be achieved through appropriate model testing and that, in any case, the model developers carry out such code review and testing as part of their developmental work. Setting aside the question of the independence of the validation, it is the author's experience that model developers are not necessarily natural programmers, are prone to favoring opaque coding techniques, and have little appetite for appropriately commenting to their code. Furthermore, the amount of actual testing and code review carried out by developers prior to independent validation is not always obvious. Although it is true that the existence of implementation errors can be detected (and brought to the attention of model developers) through the formulation of relevant model tests, a major concern with such an approach is that it will never be possible to second guess all implementation errors which may arise in practice and that the set of model tests carried out will never, by their very nature, be exhaustive. Carrying out a code review also provides the validator with a "feeling" for the model and some level of comfort around the developers' skills which all form part of the subjective picture being mentally built during the validation. Furthermore, there are some errors such as those leading to memory leaks and the unintentional "reading/writing" of memory locations resulting from overstepping array boundaries which can often only be caught through code analysis. Code reviews also highlight instances of hard-coded variables, the nature of any numerical approximations made, and allow the checking of all those minor calculation errors which may only be material on a portfolio basis.

The decision to perform a code review will invariably depend on the complexity of the model under consideration. Many models can be thought of as "framework + payoff" constructs in which the generation of sets of values for the model variables at different points in time is carried out separately from the payoff function which simply takes these sets of values as inputs and applies to them a set of deterministic rules reflecting the product payoff. Once the underlying framework for generating the sets of asset paths has been validated, then a new "model" is in reality just a new payoff function, and a review of the code which implements this functionality would definitely be recommended since it would not be time consuming (most payoff functions are a few hundred lines of code at most) and furthermore, this is the only way of determining with absolute confidence that the product implemented is exactly that described by the model developers. On the other hand, the framework engine itself may run to many tens of thousands of lines of permanently evolving code, and it could be reasonably argued that the time needed to check every such line of code would be better spent elsewhere. In some cases, carrying out a full code review is the necessary prerequisite to the independent validation due to the lack of appropriate model documentation; the code review then serves the dual purpose of model discovery and model validation. Such lack of appropriate model documentation should be addressed through the imposition of model documentary standards on the developers as part of the governance structure around models as detailed in Whitehead (2010).

Code review is one of those areas in which policy should not be prescriptive with regard to the requirement, or otherwise, of carrying out a full or partial independent code review; instead, flexibility should be given to the model validation group to exercise its judgment on a model-by-model case between the time required to carry out such a review and its perceived rewards. However, policy should make it a requirement for the model developers to appropriately comment their code, prepare detailed model documentation, and grant the model validation group full access to the pricing code so that this team can use the code to appropriately guide their validation efforts.

INDEPENDENT RECONSTRUCTION OF MODELS

The independent reconstruction of all or parts of the front office pricing models is considered best practice by regulators and auditors alike but, just like code review, is another divisive issue for practitioners since this requires the setting up of another team of model developers, almost identical in size, albeit acting this time independently from the front office business areas. The rationale for rebuilding models is that the independent model is extremely unlikely to recreate the exact same errors as the model being validated and therefore provides a useful validation on the implementation of the front office pricing model. However, emphasis should be placed here on "the exact same errors," because, just as front office model builders are prone to errors, the same can be said for independent model validators and the process of managing any valuation differences between the two models can be complicated. In any case, the rationale for rebuilding exactly the same model is arguable; any systematic reconstruction of front office models should focus instead on considering alternative quantitative techniques (for example, using lattice based techniques instead of Monte Carlo simulation) and on using a model with a greater number of risk factors or different dynamics for the underliers. This would have the advantage of permitting the testing of the model assumptions themselves as part of the independent validation and of investigating the impact of any perceived limitations in the front office pricing model. The implementation of the payoff formulations can still be carried out in such cases as all models can be degenerated to their deterministic cases, which allows for the comparison of outputs when both model set-ups are nonrandom. It should also be emphasized at this stage that model validators have a tremendous advantage over their front office counterparts in that their alternative model specifications will not be used under real trading conditions and, consequently, do not need to be fast.

A practical compromise to the joint issues of code review and model reconstruction would be to enable the model validation group to install a clone copy of the front office pricing source code on their own independent testing platform that would allow them to modify portions of the pricing code to build variants of the front office pricing model to test out various hypotheses and concerns which they may develop during their validation efforts. In such a set-up, the model validation group would be carrying out an implicit code review as part of their modification of the pricing model and the impact of any model changes carried out by the model validation group in constructing their model variants would be easy to explain since the differences with the original code base would be transparent.

BENCHMARK MODELS

A key motivation for building an alternative model as part of the validation process is to produce a benchmark against which the behavior of the front office model can be compared. The use of as many benchmark models as possible in model testing is primordial for the independent validation of valuation models. Most model validation groups will have access to a wide variety of front office production models from different front office teams working on different product areas. Each such model development team will tend to favor certain quantitative techniques and model dynamics and the model validation group should learn to leverage their access to such a wide set of comparative modeling tools. The difficulty consists in being able to collapse the product under consideration to a different product covered by an alternative front office pricing model (possibly employing different modeling assumptions) in such a way as to ensure consistency of model inputs and consistency of model payoffs. This may require the transformation of certain data inputs or the restriction of contractual parameters but if such equivalence can be achieved, then this provides ready-made alternative modeling and implementation benchmarks. Vendor models and independent valuation services could also be employed as benchmarks although their usefulness tends to be hindered by the lack of detailed information provided by such third parties on their model assumptions and implementation. Naturally, market prices would be the best yardsticks for the model if these were observable and, failing that, comparison against the market standard model would be essential but this is covered in detail elsewhere (Whitehead, 2010) and will not be considered any further here.

MODEL TESTING

Exhaustive testing of the valuation model, both in isolation and compared against other benchmarks, is a main component of any model validation process. This requires devising a number of scenarios, each one being a different combination of input parameters, which will enable the validator to determine if the pricing model, acting as a "black box" in which nothing is assumed known apart from the model inputs and its outputs, is behaving in an expected and consistent manner, and whether or not it is in agreement with the available benchmark models. The two main dimensions to such testing involve the varying of input contractual parameters and market data- related input parameters (the exact specification of which will depend on whether the model is internally or externally calibrated as detailed more fully later on in this chapter). Different types of model tests can also be recognized as detailed in the following paragraphs.

Vanilla repricing scenarios verify the accuracy of the model in valuing both the hedging instruments in terms of which the model was constructed and other basic financial instruments. Tests on limiting cases consider the natural boundaries of the product and aim to produce deterministic outcomes. Examples would include put options with zero strike; and barriers set to either very large or small values. Limiting cases can be considered as the extremes of monotonicity tests in which a single input parameter is varied from a lower to an upper bound with a view to determining whether the trend of values produced is correct. For example, a derivative in which the buyer can exercise an option on a set of dates should not decrease in value if the number of exercise dates is increased; call option values should decrease with increasing strikes; and a product in which the associated payoff is capped at a certain level should not decrease in value as the cap is increased. Such trend analysis should also be applied to all market data related input parameters such as current values for the underlying variables, volatilities, correlations, discounting rates, and so on. Other tests include exploiting put-call parity-type relationships to combine products to yield deterministic payoffs; setting values for the underlying variables so that they are deeply "in-the-money" and behave as forwards, so that the product should show little sensitivity to volatility inputs; and constructing scenarios which bound valuations between certain ranges of values.

Payoff implementation tests aim to verify that the actual product payoff correctly reflects that described by the model developers and requires valuing the product with different combinations of contractual input parameters in conjunction with input market data specified in such a way as to produce a deterministic evolution for the underlying variables. With a known deterministic evolution for the relevant quantities, even payoffs to path dependent options can be independently replicated on a spreadsheet and allow comparisons with the model produced outputs.

Convergence tests relate to varying nonmarket data-related model parameters and are invaluable for simulation (e.g., Monte Carlo) and grid (e.g., finite difference and tree) based models to ensure that a sufficient number of paths or grid density is being used to ensure accurate model outputs. Ensuring adequate convergence is particularly important when underlying variables are set close to barriers as the required number of paths or grid density required to attain an acceptable level of accuracy can dramatically increase as the underlying variables are moved toward the barrier. The same comment can be made when the valuation date is moved closer to a date on which a contractual feature of the product gets resolved.

Model stability tests highlight possible implementation problems and involve a slight perturbation in the model input parameters (mainly to market data-related parameters but also to contractual parameters), which should result in only a slight change in model outputs. Instability with regards to market data may also be indicative of problems in the calibration of the model.

Model stress testing involves constructing scenarios using market data which is significantly different in levels, shape, steepness and interrelationships from normal market conditions in order to ascertain how the model and its outputs would behave under such extreme conditions. The performance of the calibration of the model under these adverse conditions needs careful investigation since scenarios will always exist under which the model will simply not be able to calibrate to its data and, as a result, the model is not able to produce any outputs whatsoever. The results of such calibration failures under live trading conditions can be catastrophic. Even if the calibration process for a model does not fail under extreme conditions, the quality of the calibration may be significantly impaired. The aim of model stress testing is to raise awareness of the extreme market conditions under which the valuation and hedging of a product using a particular model and calibration targets break down, or are no longer recommended. The validation process should identify such conditions and raise them as possible model limitations, and the parties involved in controlling the model environment then need to consider the steps which would be required to manage such extreme situations should they occur.

The change in value of a product over a given period of time should be fully explainable in terms of the change in the underlying market data and the sensitivity of the product to this data. Trade profit-and-loss (PnL) explained reports form a standard part of any validation and control process and the existence of a large percentage of unexplained change in the valuation of a trade should be cause for concern. PnL explains are tests on the internal consistency of a model, ensuring that model prices and hedge sensitivities are compatible. Although the valuation of some products using specific models may result in closed-form formulae for option sensitivities, the majority of hedge sensitivities are produced through numerical schemes. Even in these cases, the numerical sensitivities output from the model itself (for example, using neighboring grid points in finite difference techniques for option deltas) will rarely be used in official trading and risk management systems since common practice is to run batch processes on all trading books in which the market data is perturbed externally to the model to obtain different prices from which the sensitivities are calculated using standard central difference formulae (the advantage of such external "bump and revalue" batch processes being that the systems do not need to "know" how a particular model produces its sensitivities). Hedge sensitivity tests should verify the accuracy of the option sensitivities output by the model by independently carrying out a "bump and revalue" approach. In addition, a similar analysis should be carried out on the batch system produced numbers on a sample trade basis.

Backtesting a model through an entire simulated lifecycle of a product is the ultimate test of its internal consistency but imposes such significant system and storage resource requirements that this is never more than just a theoretical concept. Such backtesting requires the specification and storage of plausible market data for every business day during the simulated life of the product, the calibration of the model and production of hedge sensitivities on a daily basis, in addition to formulating rules-based hedging strategies which would mimic the action of traders. This is equivalent to setting up a full test-trading environment as part of the model validation process and is the reason why carrying out simulated backtesting does not normally form part of the validation process.

A much more realistic goal is the production of a hedge simulation tool which would carry out simulated PnL explain tests for a large number of scenarios over a single time period only. This simply requires the specification of market data at the start of the period (which can be obtained for testing purposes from random deformations of real market data) from which trade valuations and hedge sensitivities can be obtained using the pricing model, together with the specification of the market data and corresponding trade valuations at the end of the period. Each set of values for the market data at the end of the period will lead to a single PnL explain test and construction of a routine for producing random sets of market data from an initial set would enable the automation of a large number of such scenarios.

TRADE SEASONING

The testing of pricing models can easily end up focusing uniquely on scenarios in which products are set up as "new" trades with the first contractual fixing date occurring after the chosen valuation date. However, the proper valuation of live trades is usually dependent on a history of realized past values for specific quantities; for example, the valuation of an Asian option in which a number of the averaging dates have already elapsed will be dependent on the actual, realized asset prices on those past dates and the pricing model must be able to take these values into account to obtain a correct trade valuation. The validation of such trade seasoning is easily overlooked and although errors in trade seasoning will not affect the initial price at which a trade is transacted, it will lead to valuation differences and hedging errors subsequently throughout the life of the trade. The logic for trade seasoning can either occur mainly within or outside of the pricing model, leading to different validation and control requirements. Using the above Asian option example, if trade seasoning occurs externally to the model, then the model would have to have an input parameter for the average asset price over elapsed fixing dates and this average would be calculated through some external process and then fed as an input into the model. This leads to higher operational risk since any failure to update that average in the upstream process would result in a stale past average being used by the model. In addition, the meaning of such trade seasoning input parameters must be clearly understood by all relevant parties. For an Asian option, the past fixings would actually be most easily captured through the past sum of realized fixings rather than through its elapsed average since use of the latter also requires an input parameter to reflect the number of past fixings. Confusion around the input requirements could easily lead to erroneously passing an elapsed average when the model expects a sum and leads to valuation errors. With an internal approach to trade seasoning, the history of past fixings would be the input parameter and the model itself would internally calculate the required sum over elapsed averaging dates. This approach is less prone to operational problems but places greater reliance on ensuring that this trade seasoning logic is validated through both code reviews and appropriate model testing (specifically, using scenarios in which a deterministic evolution of the underlying variables is guaranteed and replicating the seasoned payoff in a spreadsheet). Apart from the operational concerns, the decision to use internal or external trade seasoning for such simple options might be open to personal preferences. However, as soon as very path-dependent products are considered, for example, with a single payment at maturity but where the coupon paid accrues over a number of periods depending on the behavior of multiple underlying variables during those periods, then the use of fully internal trade seasoning logic in which only the past fixings for the relevant underlying variables are model inputs becomes essential. Once the validation of the trade seasoning logic has been effected, then the only control requirement is to verify the accuracy of the past fixings input data for the live trade population.

PRICE VERIFICATION

Incorrectly entered historical trade fixings is one source of model risk related to the use of input data in valuation models. Another relates to incorrectly entered contractual data which should be addressed as part of standard, deal review processes carried out by middle office functions. These processes do however need to consider the common practice of shifting contractual parameters such as barriers and strikes for hedging purposes; deal review processes should identify those shifted trades and monitor on an ongoing basis the size of the resulting embedded reserves. It should be noted that the validation of model input parameters is not normally considered part of the model validation process but is nevertheless crucial to ensuring accurate valuations. In most firms, traders set the values of the input market data which is used to price trades; the rationale for this practice is that these input values will impact not only valuations, but crucially for risk management purposes, hedge sensitivities as well and traders should have the freedom to risk manage their positions according to their own views. The compensating control for this practice is the price verification process which is typically carried out by the valuation control group and aims to source independent market data with which to compare the input values used by traders and the impact, or variances, on product valuations resulting from differences in the market data used by the desk and the valuation control group. The price verification process needs to contend not only with incorrectly specified input market data but also with the use of proxies and historical data for illiquid markets. This process is further complicated by the prevalent use of calibrated model input parameters for which the linkage to the original market data used by the traders during the calibration process is often not available.

CALIBRATION

The role of calibrations in the validation of valuation models is crucial and far reaching. The formulation of pricing models contains either an explicit or an implicit description of the stochastic evolution of the underlying variables from which all other necessary quantities are derived. These dynamics are usually postulated in terms of mathematical equations through a parsimonious set of model parameters. Different sets of model parameters will imply different evolutions over time for the financial quantities under consideration and consequently different model prices for exactly the same product. Calibration is the process of assigning values to model parameters in a manner consistent with the observed market values of the simpler financial instruments which are being used to hedge the risk on the more exotic product valued with the pricing model. In effect, the more exotic product is being priced relative to the vanilla hedging instruments in terms of which it can be replicated. These vanilla instruments are said to be the "calibration targets" for that product when valued using this model. Since the use of different vanilla instruments as targets will lead to different values for model parameters and hence different valuations for the exotic product, the choice and transparency around calibration targets is essential. This is a major theme of Whitehead (2010) where the need to enforce a strict product-model calibration scope when approving and controlling models is emphasized. Calibration is in itself a complex numerical routine which attempts through an iterative process to minimize (some specified function of) the differences between the market values of the target calibration set and their model prices. The process may allow the placing of greater emphasis on certain individual elements of the calibration set and usually requires the specification of an initial guess for the model parameters.

The calibration process can either be carried out externally to the pricing model or else it can be subsumed within the model itself. This choice will impact not only the representation of market data-related input parameters but also the validation and control processes required to ensure the integrity and transparency of this calibration process and subsequent product valuations. With external calibration, the model parameters are themselves input parameters to the pricing model, whereas with internal calibration, it is the values for the set of calibration targets which are the actual market data-related input parameters. The debate around internal or external calibrations revolves around the trade-off between speed and controls. Use of externally calibrated model parameters leads to a significant speed advantage when large portfolios of positions are considered since a number of trades are likely to be using the same calibration sets and consequently the actual number of calibrations which need to be carried out will be smaller with the external calibration approach. This speed advantage only increases for models with multiple assets. On the other hand, pricing models in which the calibration routine is internal will actually be recalibrated every time that the model is invoked to return a valuation. However, the control advantages of internal calibrations are significant. First of all, the calibrations will never be stale since they are implicitly updated for every valuation performed; external calibrations can suffer from infrequent recalibrations. The calibration routine and choice of targets explicitly forms part of the model validation process for internal calibrations since it is part of the model leading to greater transparency around the calibration process and choice of targets. With external calibrations, the actual calibration routine and choice of targets may not be at all visible to the control functions, rendering it difficult to verify that the trader is not internally manipulating the calibration process and parameters. The price verification process is also a lot more straightforward to carry out for internal calibrations since this only requires the replacement of trader supplied market data in the model function calls with independently sourced market data. The possible lack of visibility around the selection of targets with external calibrations significantly complicates the price verification process that may now have to reference the market data values implied by the desk calibrations and the desk sensitivities rather than enabling a full, independent revaluation of the positions using independent data.

The choice of calibration targets and the number of sets of targets associated with a particular model will be dictated by whether a local or global calibration approach is being employed. With local calibrations, each product may have its own specific set of calibration targets, and the appropriateness of each particular set will need to be considered. Local calibrations should result in a good calibration fit with very close repricing of the calibration instruments. A global calibration approach would specify a single, larger set of target instruments applicable for a wide range of products being valued using that model and results in a more generic calibration with reasonable overall fit but greater local errors in the repricing of specific calibration instruments within the wider calibration universe. The debate around local and global calibrations can be illustrated through a specific example. Consider an option with a three-year maturity that is being valued through a global calibration to the implied volatility surface for maturities up to ten years. Should the trader really be concerned about the quality of the calibration fit beyond three years for this option if this results in a worse calibration fit up to three years when compared with an equivalent local calibration up to three years only? Now consider this in conjunction with a second option of seven years' maturity and assume that two local calibration sets are being used for these options. Both options are being valued as accurately as possible in isolation but since they are using different sets of model parameters, the trader may not be convinced that these options are being consistently valued with regards to each other which may have an impact when exposures are netted across positions. A global calibration would ensure consistency of pricing but would result in locally worse calibration fits. Note that a product using local calibration will only show sensitivities to its target instruments, whereas a global calibration approach would result in a product showing exposure to all available market data points.

The validation of the calibration process and choice of targets is an integral part of the overall validation of valuation models. As already mentioned, the choice of calibration targets to be used in conjunction with a particular model and for a specific product must be explicitly specified during the model approval process and a key component of the validation process itself relates to the appropriateness of the calibration set for the product and model under consideration. It should be emphasized that differences between the postulated calibration set and the actual set of instruments used by traders to risk manage the product on a daily basis may occur in practice and would result in inconsistencies between the model prices and sensitivities and the real life hedging of the positions. However, the impact of any such differences is likely to be obscured by the netting of sensitivities across positions and the macro hedging of trading books.

The stability of the calibration process should be investigated by perturbing the values of the calibration targets (both in isolation and in combination), recalibrating the model and investigating the impact of the changes on calibrated model parameters (for external calibration) and product valuations (for both internal and external calibrations). A stable calibration would result in only small changes to model parameters and valuations. Instability of model parameters and valuations may indicate a problem with the calibration routine itself, an implementation error in the model or a misspecified model. For externally calibrated model parameters, the stability of model parameters over time should also be considered. A well-specified model will have model parameters which do not change too dramatically over time. Finally, the goodness of the calibration fit should be examined together with the sensitivity of the calibrated model parameters and valuations to the initial guess for the model parameters.

CONCLUSION

This chapter has considered a number of different concepts involved in the validation of valuation models which is an essential part of any framework attempting to address model risk.

REFERENCES

Whitehead, P. (2010) Techniques for Mitigating Model Risk. In Gregoriou, G., Hoppe C., and Wehn, C. (eds.). (2010) *The Risk Modeling Evaluation Handbook*. New York: McGraw-Hill.

CHAPTER 16

MODEL RISK IN THE CONTEXT OF EQUITY DERIVATIVES PRICING

Bernd Engelmann and Fiodar Kilin

ABSTRACT

The simplest approach to assess an equity derivative pricing model is to check how well market prices of vanilla options are fitted. However, when pricing exotic options, this approach is not sufficient for the rational choice of a model. Generally, each type of exotic option has its own set of most suitable models that take into account specific risks of the contract type. In this chapter, we provide an overview of studies that analyze the question of which model should be chosen to price and hedge barrier options. We check the results provided in the literature using a set of numerical experiments. In our test we compare prices of forward-start options in the local volatility, Heston, and Barndoff-Nielsen–Shephard models.

INTRODUCTION

The market of equity derivatives can be split into a market for vanilla options, i.e., call and put options, and a market for exotic options. The exotic options market covers all products which are not standard calls and puts. Examples are forward-start call and put options, path-dependent options like barrier or Asian options, and products on several underlyings such as basket options. To understand the relevance of model risk for equity derivatives we start with a short description of the vanilla and the exotics market.

Vanilla options on indexes or single stocks are traded on exchanges. On the EUREX, vanilla options on the DAX or the EuroStoxx 50 are traded for a set of strikes and maturities that is defined by the exchange. These options are of European type. Options on single stocks like EON, Deutsche Bank, or Siemens that are traded on the EUREX are of American type. For these products prices are quoted every day for the set of strikes and expiries where these instruments are traded. No model risk exists for these options because prices are delivered every day by the exchange. If the price of an option for a different strike or expiry is needed, typically an interpolation using the Black-Scholes model is applied. Prices given by the exchange are converted into implied volatilities using the Black-Scholes model. These implied volatilities are interpolated to find the implied volatility corresponding to the option's strike and expiry. The option's price is computed using the Black-Scholes model with this interpolated implied volatility. An alternative would be to calibrate a pricing model like the Heston model to the prices given by the exchange and price the option using the calibrated model. In this context model risk is rather low because the price of the vanilla option with the irregular strike and expiry has to be in line with the set of prices of similar options that is given in the market.

For exotic options hardly any market prices exist. They are not traded on exchanges but typically over-the-counter between banks or banks and institutional investors such as asset management firms or insurance companies. For these products prices are determined by pricing models. These models are typically calibrated to vanilla options, i.e., the model parameters are determined to replicate the given prices of vanilla options as close as possible. After calibrating a model, the price of an exotic product is computed in this pricing model. The basic idea behind this procedure is to price an exotic product in line with the market for vanilla options. In this context model risk can be substantial. The more the characteristics of an exotic product differ from a vanilla option, the less its price is determined by the vanilla options market.

This chapter is structured as follows. In the first section, we provide an overview of pricing models for equity derivatives. In the second section, we describe the problem of model risk in more details. A numerical example illustrates the problem for the case of forward-start vanilla options in the following section. In the final section we discuss the practical implications.

EQUITY DERIVATIVES PRICING MODELS

In this section we give a short overview of equity derivatives pricing models. The most prominent equity derivatives pricing model is the

Black-Scholes model (Black and Scholes, 1973). Its dynamics under the risk-neutral measure is given by $dS_t = (r - d) \cdot S_t \cdot dt + \sigma \cdot S_t \cdot dW_t$, where S is the price of a stock or stock index, r denotes the risk-free interest rate, d the dividend yield, σ the volatility, and W a Wiener process. In the Black-Scholes model r, d, and σ are at most time dependent.[1] It is well known that prices in the market for vanilla options cannot be explained by the Black-Scholes model. Computing implied volatilities from observed market prices leads to a volatility smile for short maturities and to a volatility skew for longer expiries. Therefore, assuming a time-dependent volatility is inconsistent with the vanilla options market.

A natural extension of the Black-Scholes model is the local volatility model. Its dynamics is identical to the Black-Scholes model but the volatility σ is a function of spot and time $\sigma(S_t, t)$. It was shown by Dupire (1994) that for every surface of arbitrage-free prices of European call options $C(K, T)$, where K is the strike and T the option's expiry, exists a unique local volatility function $\sigma(S_t, t)$ that is consistent with the given price surface. Therefore, the local volatility model can be calibrated perfectly to every arbitrage-free surface of European call option prices. However, as shown in Hagan et al. (2002), the predictions of future volatility smiles and the smile dynamics under spot shifts implied by the model are unrealistic. For this reason, new models which are both able to explain the current implied volatility smile and give realistic predictions of future smiles had been developed.

It is observed in the market that stock volatility goes up when prices go down and vice versa. Further, it is observed that the level of volatility is fluctuating in time. Therefore, a natural extension to the Black-Scholes model is a stochastic volatility model. The most prominent example is the Heston (1993) model. Its dynamics is given by

$$\begin{aligned} dS_t &= (r - d) \cdot s_t \cdot dt + \sigma_t \cdot S_t \cdot dW_t, \\ d\sigma_t^2 &= \kappa \cdot (\eta - \sigma_t^2) \cdot dt + \theta \cdot \sigma_t \cdot d\overline{W}_t, \\ Cov[dW_t, d\overline{W}_t], &= \rho \cdot dt, \end{aligned}$$

where θ is the volatility of volatility, η the long-term variance, and κ is mean-reversion speed. The correlation between the driving Brownian motions is denoted with ρ. This correlation has to be negative to explain the observed co-movement of spot and volatility.

The Heston model still implies continuous spot paths. In reality sometimes huge jumps in the spot are observed. This can be included in the model by adding a jump component to the spot dynamics of the Heston model (Bates, 1996):

$$dS_t = \left(r - d - \lambda \cdot \mu_{\mathcal{J}}\right) \cdot S_t \cdot dt + \sigma_t \cdot S_t \cdot dW_t + \mathcal{J} \cdot S_t \cdot dN_t,$$
$$\log\left(1 + \mathcal{J}\right) \approx N\left(\log\left(1 + \mu_{\mathcal{J}}\right) - \frac{\sigma_{\mathcal{J}}^2}{2}, \sigma_{\mathcal{J}}^2\right),$$

where N_t is a Poisson process with intensity $\lambda > 0$. The Poisson process is independent of both Wiener processes and $\mathcal{J}$ is the percentage jump size which is log-normally distributed. This distribution is determined by the parameters $\mu_{\mathcal{J}}$ and $\sigma_{\mathcal{J}}$.

An alternative model class are Levy models with stochastic time (Carr et al., 2003). The underlying process is modeled under the risk-neutral measure as

$$S_t = S_0 \frac{\exp\left(\left(r - d\right) \cdot t\right)}{E\left[\exp\left(X_{Y_t}\right) \mid y_0\right]} \exp\left(X_{Y_t}\right),$$
$$Y_t = \int_0^t y_s \, ds,$$

where X_t is a Levy process, Y_t is a business time, and y_t the rate of time change. Typical choices for X_t are the variance gamma process or the normal inverse Gaussian process, for y_t the Cox-Ingersoll-Ross process or the Gamma-Ornstein-Uhlenbeck process. This class of processes is able to represent current implied volatility surfaces with reasonable accuracy and gives realistic predictions on the future shape of implied volatilities. However, compared to Heston or Bates, the parameters in this model class are less intuitive. For instance, a trader might be more comfortable with a volatility of volatility than with a mean-reversion speed of a stochastic clock.

A further alternative is the Barndorff-Nielsen–Shephard (2001) model. Its dynamics is given by

$$d\left(\ln\left(S_t\right)\right) = \left(r - d - \lambda k\left(-\rho\right) - \frac{\sigma_t^2}{2}\right) \cdot dt + \sigma_t dW_t + \rho \cdot d\mathcal{J}_{\lambda t},$$
$$d\sigma_t^2 = -\lambda \cdot \sigma_t^2 \cdot dt + d\mathcal{J}_{\lambda t},$$

with $j_t = \sum_{n=1}^{N_t} x_n$, where N_t is a Poisson process with intensity a, x_n is independent and identically distributed following an exponential distribution with mean $1/b$. The function k is given by $k(u) = \ln(E[\exp(-u \cdot \mathcal{J}_1)]) = -a \cdot u \cdot (b + u)$. In this model the volatility process is entirely a jump process and only the spot process has a diffusive component.

All models presented so far model the spot and at most in addition the instantaneous volatility. For all these models the characteristic function of the spot distribution is known in closed form at any future point in time which allows the calculation of expectations of the spot and the calculation of European vanilla options prices.[2] The calibration of these models is done by solving the optimization problem

$$\min_{\text{Parameters}} \sum_{\text{Call Options}} \left(\text{Market Price} - \text{Model Price}\left(\text{Parameters}\right)\right).$$

Since the model price of a European call option can be computed almost analytically in these models, the calibration can be carried out in a very efficient way.

MODEL RISK FOR EQUITY DERIVATIVES

In the last section we have introduced the most prominent equity derivatives pricing models. After calibrating these models they all give reasonable prices for European vanilla options. When pricing an exotic option, like a down-and-out put option or a cliquet option, the question arises as to which of the models presented so far is best suited for the specific product.

As a first step, one could ask the question if it makes a difference which model is used. This question was answered by Schoutens, Simons, and Tistaert (2004). In Schoutens et al. (2004), several exotic equity options are prices in the Heston model, the Bates model, four Levy models, and the Barnsdorff-Nielsen–Shephard model. They find that all models are able to replicate the prices of European call options with reasonable accuracy but that differences in prices of exotic options can be substantial. Especially prices of reverse barrier options are extremely model sensitive.

The reason for this behavior is that these models make very different predictions of the distribution of forward volatility. Especially for cliquet options or forward-start options, it is clear that forward volatility plays a crucial role. However, in these models the distribution of forward volatility is not transparent but hidden inside the model assumptions. For this reason, some modern modeling approaches attempt to model forward volatility explicitly. An example is Bergomi (2005). These modeling approaches have the advantage that the modeling of forward volatility is no longer opaque but the disadvantage that not enough market instruments are available to calibrate the model and hedge exotic options in this model. Therefore, the Bergomi model is not yet suited for practical applications and one has to rely on one of the models presented in the last section.

When choosing a model for an exotic product several aspects have to be taken into account. First of all one has to decide about the relevance of model risk for the specific product. We will present an illustrative example in the next section. If the result is that model risk is substantial then it is rather difficult to make a decision for a specific model. One way is to carry out a backtest as in Engelmann et al. (2006) to make a model decision. In this backtest historical market data is needed for several years. On the historical data the issuing of exotic options and hedging their risks can be simulated on real market data and the hedging error over the lifetime of the product can be measured. The model that delivers on average the least hedging error should be the most suitable model for the exotic product. Further, an indication for the most suitable hedging strategy is also found in this way.

If a long history of market data is not available, there is an alternative to a backtest. One could use an advanced model like the Bergomi (2005) model and assume that this model gives a realistic description of markets. Although it cannot be calibrated to the market it can be parameterized and it can be assumed that after parameterization it represents a realistic market. After that, prices of vanilla options are calculated and the models of the last section are calibrated. Then the exotic product is prices both in the Bergomi model and in all the other models. This is done for several parameterizations of the Bergomi model. The model that delivers prices that are closest to the prices in the Bergomi model can be considered the most realistic model for the specific product. A reference for this procedure is Kilin, Nalholm, and Wystup, (2008).

ILLUSTRATIVE EXAMPLE

In this section we describe a numerical example where we restrict the analysis to comparison of option prices in three models only. We calculate prices of forward-start vanilla options in the local volatility, Heston and Barndorff-Nielsen–Shephard models using different implied volatility surfaces as an input. We then analyze which patterns of the implied volatility surface lead to highest model risk.

We generate 24 different implied volatility surfaces from 24 different sets of parameters of the Heston model. Values of these parameters are reported in Table 16.1. When generating the implied volatility surfaces we use European call option with maturities of one, three, and six months and one, three, and five years. For the first maturity (one month) we use only one at-the-money option. For the second maturity we use three options

Table 16.1 Parameters of the Heston Model Used for Generating Implied Volatility Surfaces

Scenario	Long-run Variance	Mean-reversion Rate	Volatility of Variance	Short Volume	Correlation
1	0.04	0.25	0.5	0.25	−0.5
2	0.04	0.25	0.5	0.25	−0.6
3	0.04	0.25	0.5	0.25	−0.7
4	0.04	1.15	0.39	0.2	−0.64
5	0.04	2	1.5	0.3	−0.5
6	0.04	2	1.5	0.3	−0.6
7	0.04	2	1.5	0.3	−0.7
8	0.04	2	1.2	0.3	−0.7
9	0.049	2	1	0.8	−0.7
10	0.01	2	0.2	0.16	−0.7
11	0.04	2	0.5	0.3	0
12	0.04	2	0.5	0.9	−0.7
13	0.04	2	0.5	0.7	−0.7
14	0.04	2	0.5	0.3	−0.7
15	0.04	2	0.5	0.3	−0.5
16	0.04	2	0.5	0.3	−0.6
17	0.04	2	0.5	0.3	−0.8
18	0.04	2	0.5	0.3	−0.9
19	0.04	2	0.5	0.24	−0.7
20	0.04	2	0.5	0.26	−0.7
21	0.04	2	0.5	0.34	−0.7
22	0.04	2	0.5	0.38	−0.7
23	0.04	2	0.4	0.3	−0.7
24	0.04	2	0.6	0.3	−0.7

with 90, 100, and 110 percent strike. For all further maturities we calculate prices of 15 vanilla options with the strikes equidistantly distributed between 65 and 135 percent. After calculating the prices of these options in the Heston model, we calculate implied volatilities from these prices using the inverse of the Black-Scholes formula. These implied volatility surfaces are then used for calibration of the local volatility and Barndorff-Nielsen–Shephard models. The local volatility model is calibrated using the algorithm described in Andersen and Brotherton-Ratcliffe (1998). The Barndorff-Nielsen–Shephard model is calibrated using the direct integration method and caching technique described in Kilin (2007). After these calibrations we have parameters of three models for each of the 24 scenarios. Using these parameters we calculate prices of ten forward-start options specified in Table 16.2. The payoff of a forward-start call is

Table 16.2 Specification of Forward-Start Options That Are Used in the Pricing Experiment

Instrument	Forward-start Time	Maturity	Relative Strike	Option Type
1	0.25	0.5	0.8	Put
2	0.25	0.5	1	Call
3	0.5	1	0.9	Put
4	0.5	1	1	Call
5	0.75	1	1	Call
6	1	3	0.8	Put
7	1	3	1	Call
8	2.5	3	1	Call
9	1	5	1	Call
10	4	5	1	Call

$\max(S(T) - kS(t_0), 0)$. The payoff of a forward-start put is $\max(kS(t_0) - S(T), 0)$, where T is the maturity of the option, t_0 is the forward-start time, k is the relative strike.

In all our experiments we make a simplifying assumption of zero interest rates and absence of dividends. The results of the experiment are reported in Tables 16.3 to 16.8 and analyzed below. In the rest of this section we describe special features of the implied volatility surfaces corresponding to scenarios used in our experiment. Scenarios 1 to 3 describe situations when at-the-money implied volatilities decrease very slowly as the expiries of options increases. Scenario 4 is based on parameters reported in Bakshi, Cao, and Chen (1997). This is an example of a Heston model parameterization commonly used in academic literature. Scenarios 5 to 8 deal with strong convexity of the implied volatility curves. This convexity is especially strong for scenarios 5 to 7. Parameters of scenario 9 produce an example of volatile markets. A typical example of calm markets is described by parameters of scenario 10. Scenario 11 describes implied volatility surface with almost symmetric smiles. Scenarios 12 and 13 correspond to the case of very high short-term at-the-money implied volatility. Scenario 14 illustrates a standard case for equity derivatives markets. Scenarios 15 to 18 are derived from scenario 14 by changing the skew of the implied volatility. The skew is modified by changing the correlation between the underlying process and variance process in the Heston model. Scenarios 19 to 22 are derived from scenario 14 by changing the short-term, at-the-money volatility. These changes are obtained by modifying the initial state of the variance process in the Heston model. Scenarios 23 and 24 are derived from scenario 14 by varying the convexity of the implied volatility curves. Specifically, the volatility of variance parameter is changed.

Table 16.3 Prices of Forward-Start Options in Different Models. Scenarios 1–4

Scenario	Instrument	Heston	Local Volatility	Barndorff-Nielsen-Shephard
1	1	0.0048	0.0041	0.0041
1	2	0.0326	0.0380	0.0404
1	3	0.0188	0.0224	0.0181
1	4	0.0589	0.0669	0.0494
1	5	0.0367	0.0454	0.0318
1	6	0.0231	0.0331	0.0339
1	7	0.1135	0.1131	0.0772
1	8	0.0475	0.0564	0.0267
1	9	0.2059	0.1853	0.1152
1	10	0.0607	0.0770	0.0395
2	1	0.0049	0.0043	0.0046
2	2	0.0321	0.0369	0.0383
2	3	0.0188	0.0222	0.0172
2	4	0.0584	0.0654	0.0456
2	5	0.0361	0.0439	0.0285
2	6	0.0228	0.0330	0.0347
2	7	0.1122	0.1097	0.0735
2	8	0.0466	0.0542	0.0230
2	9	0.2052	0.1855	0.1150
2	10	0.0597	0.0785	0.0383
3	1	0.0051	0.0047	0.0050
3	2	0.0316	0.0356	0.0358
3	3	0.0187	0.0219	0.0160
3	4	0.0578	0.0635	0.0407
3	5	0.0357	0.0422	0.0245
3	6	0.0226	0.0326	0.0357
3	7	0.1106	0.1053	0.0700
3	8	0.0456	0.0518	0.0209
3	9	0.2044	0.1880	0.1161
3	10	0.0590	0.0811	0.0378
4	1	0.0023	0.0020	0.0026
4	2	0.0281	0.0296	0.0336
4	3	0.0147	0.0164	0.0183
4	4	0.0592	0.0588	0.0461
4	5	0.0388	0.0399	0.0309
4	6	0.0225	0.0278	0.0333
4	7	0.1287	0.1139	0.0981
4	8	0.0600	0.0579	0.0445
4	9	0.2237	0.1835	0.1428
4	10	0.0833	0.0811	0.0665

Table 16.4 Prices of Forward-Start Options in Different Models. Scenarios 5–8

Scenario	Instrument	Heston	Local Volatility	Barndorff-Nielsen-Shephard
5	1	0.0061	0.0056	0.0053
5	2	0.0224	0.0364	0.0345
5	3	0.0133	0.0197	0.0174
5	4	0.0489	0.0610	0.0361
5	5	0.0299	0.0415	0.0202
5	6	0.0199	0.0312	0.0351
5	7	0.1135	0.1125	0.0814
5	8	0.0494	0.0571	0.0274
5	9	0.2121	0.1851	0.1265
5	10	0.0700	0.0803	0.0485
6	1	0.0062	0.0061	0.0054
6	2	0.0215	0.0341	0.0312
6	3	0.0133	0.0192	0.0170
6	4	0.0484	0.0590	0.0320
6	5	0.0295	0.0407	0.0173
6	6	0.0202	0.0315	0.0366
6	7	0.1128	0.1103	0.0805
6	8	0.0490	0.0557	0.0263
6	9	0.2119	0.1844	0.1271
6	10	0.0695	0.0800	0.0472
7	1	0.0062	0.0055	0.0055
7	2	0.0206	0.0307	0.0269
7	3	0.0132	0.0196	0.0167
7	4	0.0477	0.0575	0.0281
7	5	0.0290	0.0388	0.0146
7	6	0.0203	0.0322	0.0379
7	7	0.1118	0.1078	0.0795
7	8	0.0486	0.0540	0.0252
7	9	0.2113	0.1839	0.1276
7	10	0.0690	0.0799	0.0461
8	1	0.0063	0.0050	0.0059
8	2	0.0242	0.0324	0.0296
8	3	0.0149	0.0187	0.0187
8	4	0.0515	0.0583	0.0321
8	5	0.0317	0.0392	0.0173
8	6	0.0218	0.0298	0.0394
8	7	0.1169	0.1088	0.0870
8	8	0.0515	0.0548	0.0293
8	9	0.2159	0.1828	0.1350
8	10	0.0728	0.0808	0.0520

Table 16.5 Prices of Forward-Start Options in Different Models. Scenarios 9–12

Scenario	Instrument	Heston	Local Volatility	Barndorff-Nielsen-Shephard
9	1	0.0325	0.0289	0.0236
9	2	0.0860	0.0891	0.0945
9	3	0.0447	0.0470	0.0437
9	4	0.0939	0.0988	0.0828
9	5	0.0536	0.0609	0.0463
9	6	0.0322	0.0319	0.0552
9	7	0.1374	0.1245	0.0837
9	8	0.0586	0.0625	0.0195
9	9	0.2351	0.1950	0.1404
9	10	0.0825	0.0895	0.0380
10	1	0.0003	0.0002	0.0004
10	2	0.0182	0.0182	0.0225
10	3	0.0040	0.0043	0.0064
10	4	0.0422	0.0390	0.0282
10	5	0.0266	0.0252	0.0182
10	6	0.0035	0.0040	0.0081
10	7	0.0909	0.0699	0.0564
10	8	0.0423	0.0350	0.0256
10	9	0.1762	0.1243	0.0801
10	10	0.0574	0.0461	0.0383
11	1	0.0036	0.0031	0.0025
11	2	0.0399	0.0446	0.0474
11	3	0.0171	0.0208	0.0203
11	4	0.0676	0.0703	0.0568
11	5	0.0441	0.0472	0.0377
11	6	0.0188	0.0269	0.0315
11	7	0.1340	0.1233	0.1019
11	8	0.0628	0.0626	0.0450
11	9	0.2259	0.1870	0.1514
11	10	0.0872	0.0807	0.0662
12	1	0.0422	0.0419	0.0361
12	2	0.1113	0.1129	0.1186
12	3	0.0630	0.0673	0.0683
12	4	0.1237	0.1250	0.1130
12	5	0.0739	0.0775	0.0677
12	6	0.0339	0.0364	0.0479
12	7	0.1489	0.1345	0.1011
12	8	0.0622	0.0613	0.0194
12	9	0.2382	0.1951	0.1357
12	10	0.0860	0.0817	0.0302

Table 16.6 Prices of Forward-Start Options in Different Models. Scenarios 13–16

Scenario	Instrument	Heston	Local Volatility	BNS
13	1	0.0261	0.0255	0.0216
13	2	0.0857	0.0874	0.0915
13	3	0.0455	0.0481	0.0480
13	4	0.1018	0.1022	0.0895
13	5	0.0614	0.0639	0.0531
13	6	0.0295	0.0309	0.0504
13	7	0.1418	0.1256	0.0932
13	8	0.0620	0.0604	0.0215
13	9	0.2332	0.1892	0.1383
13	10	0.0859	0.0817	0.0382
14	1	0.0050	0.0043	0.0058
14	2	0.0370	0.0390	0.0407
14	3	0.0191	0.0201	0.0242
14	4	0.0654	0.0646	0.0502
14	5	0.0420	0.0429	0.0307
14	6	0.0238	0.0274	0.0386
14	7	0.1322	0.1160	0.1050
14	8	0.0617	0.0591	0.0457
14	9	0.2268	0.1843	0.1497
14	10	0.0857	0.0818	0.0707
15	1	0.0046	0.0040	0.0047
15	2	0.0378	0.0409	0.0445
15	3	0.0186	0.0204	0.0238
15	4	0.0661	0.0668	0.0545
15	5	0.0426	0.0443	0.0351
15	6	0.0226	0.0273	0.0373
15	7	0.1330	0.1183	0.1059
15	8	0.0620	0.0602	0.0473
15	9	0.2268	0.1851	0.1492
15	10	0.0863	0.0814	0.0713
16	1	0.0048	0.0042	0.0054
16	2	0.0374	0.0400	0.0429
16	3	0.0188	0.0202	0.0243
16	4	0.0657	0.0656	0.0527
16	5	0.0423	0.0436	0.0331
16	6	0.0233	0.0272	0.0377
16	7	0.1327	0.1171	0.1046
16	8	0.0619	0.0597	0.0466
16	9	0.2268	0.1846	0.1490
16	10	0.0860	0.0815	0.0706

BNS, Barndorff-Nielsen-Shephard

Table 16.7 Prices of Forward-Start Options in Different Models. Scenarios 17–20

Scenario	Instrument	Heston	Local Volatility	BNS
17	1	0.0052	0.0045	0.0065
17	2	0.0365	0.0378	0.0383
17	3	0.0193	0.0200	0.0242
17	4	0.0651	0.0635	0.0481
17	5	0.0418	0.0420	0.0286
17	6	0.0243	0.0272	0.0403
17	7	0.1318	0.1146	0.1045
17	8	0.0615	0.0585	0.0447
17	9	0.2267	0.1839	0.1503
17	10	0.0854	0.0819	0.0704
18	1	0.0053	0.0047	0.0064
18	2	0.0361	0.0365	0.0352
18	3	0.0195	0.0197	0.0244
18	4	0.0647	0.0621	0.0459
18	5	0.0415	0.0409	0.0269
18	6	0.0249	0.0272	0.0415
18	7	0.1312	0.1133	0.1065
18	8	0.0613	0.0577	0.0444
18	9	0.2267	0.1833	0.1530
18	10	0.0851	0.0821	0.0703
19	1	0.0034	0.0030	0.0039
19	2	0.0308	0.0326	0.0355
19	3	0.0166	0.0177	0.0212
19	4	0.0617	0.0608	0.0481
19	5	0.0402	0.0409	0.0311
19	6	0.0234	0.0276	0.0360
19	7	0.1314	0.1155	0.1034
19	8	0.0617	0.0589	0.0462
19	9	0.2263	0.1844	0.1470
19	10	0.0857	0.0818	0.0700
20	1	0.0039	0.0034	0.0046
20	2	0.0328	0.0346	0.0372
20	3	0.0174	0.0185	0.0225
20	4	0.0629	0.0621	0.0489
20	5	0.0408	0.0416	0.0311
20	6	0.0235	0.0276	0.0375
20	7	0.1317	0.1156	0.1032
20	8	0.0617	0.0590	0.0463
20	9	0.2264	0.1843	0.1473
20	10	0.0857	0.0818	0.0704

BNS, Barndorff-Nielsen-Shephard

Table 16.8 Prices of Forward-Start Options in Different Models. Scenarios 21–24

Scenario	Instrument	Heston	Local Volatility	BNS
21	1	0.0062	0.0055	0.0067
21	2	0.0414	0.0435	0.0446
21	3	0.0210	0.0221	0.0254
21	4	0.0682	0.0676	0.0519
21	5	0.0434	0.0444	0.0310
21	6	0.0242	0.0271	0.0420
21	7	0.1329	0.1163	0.1051
21	8	0.0617	0.0592	0.0437
21	9	0.2272	0.1843	0.1519
21	10	0.0857	0.0818	0.0691
22	1	0.0077	0.0069	0.0075
22	2	0.0460	0.0481	0.0502
22	3	0.0231	0.0242	0.0267
22	4	0.0713	0.0708	0.0553
22	5	0.0450	0.0460	0.0328
22	6	0.0246	0.0271	0.0446
22	7	0.1336	0.1168	0.1054
22	8	0.0617	0.0593	0.0407
22	9	0.2277	0.1843	0.1528
22	10	0.0858	0.0817	0.0673
23	1	0.0045	0.0041	0.0053
23	2	0.0390	0.0402	0.0432
23	3	0.0194	0.0205	0.0245
23	4	0.0677	0.0662	0.0543
23	5	0.0439	0.0439	0.0343
23	6	0.0236	0.0273	0.0386
23	7	0.1345	0.1177	0.1075
23	8	0.0634	0.0599	0.0483
23	9	0.2281	0.1850	0.1519
23	10	0.0878	0.0817	0.0728
24	1	0.0054	0.0046	0.0061
24	2	0.0349	0.0378	0.0381
24	3	0.0186	0.0197	0.0236
24	4	0.0631	0.0634	0.0466
24	5	0.0402	0.0421	0.0275
24	6	0.0239	0.0274	0.0400
24	7	0.1299	0.1144	0.1022
24	8	0.0600	0.0582	0.0426
24	9	0.2253	0.1837	0.1479
24	10	0.0836	0.0818	0.0675

BNS, Barndorff-Nielsen-Shephard

CONCLUSION

In this experiment we measure model risk as price differences of products between different models. The highest model risk is observed for the options with long maturities. In most of the cases the prices of the forward-start options in the Heston model are higher than in the local volatility model. The forward-start options in the Barndorff-Nielsen–Shephard model are typically cheaper than in the local volatility and Heston models. If we compare different patterns of the implied volatility surfaces, the highest model risk is observed for scenarios 9, 12, and 13, i.e., for scenarios with the highest values of the short-term at-the-money implied volatility. The lowest model risk is observed for scenarios 4 and 14 to 24. These are the cases that correspond to standard implied volatility surfaces in the equity derivatives market. We can conclude from these observations that the model risk becomes an extremely important issue especially in nonstandard market situations.

The results of our experiment imply possible practical recommendation for financial institutions dealing with forward-start and cliquet options. These institutions should be aware of the model risk, especially in cases of high short-term implied volatility and options with long maturities or in cases when an implied volatility surface has some uncommon features that have not been observed in the past.

Usually a trader has to decide which model he wants to use to hedge his options book and he cannot switch between models frequently. In this situation one has to live with the risk that a pricing model is inadequate for hedging purposes in some market situations. When buying or selling a product, the prices for the product should be calculated in several models. If one of the models gives a substantially different price than the model used by traders and sales the alarm bells should ring and one should think very carefully about the price where one is willing to make the trade or if the trade should be made at all.

REFERENCES

Andersen, L. and Brotherton-Ratcliffe, R. (1998) The Equity Option Volatility Smile: An Implicit Finite Difference Approach. *The Journal of Computational Finance*, 1(2): 5–38.

Bakshi, G., Cao, C., and Chen, Z. (1997) Empirical Performance of Alternative Option Pricing Models. *Journal of Finance*, 52: 2003–2049.

Barndorff-Nielsen, O. and Shephard, N. (2001) Non-Gaussian Ornstein-Uhlenbeck-based Models and Some of Their Uses in Financial Economics. *Journal of the Royal Statistical Society*, B, 63: 167–241.

Bates, D. (1996) Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options. *Review of Financial Studies*, 9: 69–107.

Black, F. and Scholes, M. (1973) The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81: 637–659.

Bergomi, L. (2005) Smile Dynamics II. *RISK*, 18: 67–73.

Carr, P, Geman, H., Madan, D., and Yor, M. (2003) Stochastic Volatility for Levy Processes. *Mathematical Finance*, 13: 345–382.

Dupire, B. (1994) Pricing with a Smile. *RISK*, 7: 18–20.

Engelmann, B., Fengler, M., Nalholm, M., and Schwendner, P. (2006) Static versus Dynamic Hedges: An Empirical Comparison for Barrier Options. *Review of Derivatives Research*, 9: 239–264.

Hagan, P., Kumar, D., Lesniewski, A., and Woodward, D. (2002) Managing Smile Risk. *Wilmott Magazine*, 1: 84–108.

Heston, S. (1993) A Closed-form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *Review of Financial Studies*, 6: 327–344.

Kilin, F. (2007) Accelerating the Calibration of Stochastic Volatility Models. Available at: http://ssrn.com/abstract=965248.

Kilin, F., Nalholm, M., and Wystup, U. (2008) On the Cost of Poor Volatility Modelling—The Case of Cliquets. Working paper.

Schoutens, W., Simons, E., and Tistaert, J. (2004) A Perfect Calibration! Now What? *Wilmott Magazine*, 3:66–78.

NOTES

1. We remark that for equity derivatives the modelling of discrete dividends is crucial especially for options on single stocks. In this article we neglect this important issue and assume that modelling dividends by a dividend yield is sufficient which can at least be justified for stock indexes.
2. An exception is the local volatility model where the local volatility can be computed analytically from the surface of given European call options prices.

CHAPTER 17

TECHNIQUES FOR MITIGATING MODEL RISK

Peter Whitehead

ABSTRACT

Model risk can be defined as the loss arising from a misspecified, misapplied, or incorrectly implemented model, or resulting from incorrect decisions taken on the basis of model outputs affected by model risk. Model risk affects all types of models used in financial institutions, whether pricing models used for official valuations and hedging purposes, market and credit risk exposure measurement models used to estimate the risks faced by those institutions and the amount of capital set aside to cover those risks, or models used to define the strategic directions taken by institutions. All models are wrong by design because they are simplifications of reality. Model risk is therefore inherent in the use of models and can never be fully eliminated. Model risk can only ever be controlled and managed but this first requires a clear understanding of the origins and evolution of model risk as reflected in the structure of this chapter. Emphasis is placed on treating model risk as a multidisciplinary subject with close cooperation between all parties involved in the development, implementation, and use of models; on the need for the creation of a new position of "Chief Model Risk Officer"; and on the need to apply a strict model-product scope.

INTRODUCTION

The widespread reliance on models in the banking and financial services industry was a little known fact outside of the arcane world of quantitative modeling, derivative pricing, and structured finance. The credit and banking crisis of 2007 to 2009 has placed the prevalence and role of models across banking and finance firmly in the public spotlight. Complex products, the models used for securitization and by credit rating agencies to assess the relative riskiness of such products, together with the general underpricing of risk and mark-to-market accounting have all been blamed for creating and exacerbating this crisis. If it was not beforehand, model risk is now a key concern for banking supervisors, auditors, and market participants alike. However, even before this crisis, the possible consequences of model risk related events should have been sufficient to make controlling and minimizing such risks a priority for senior management in financial institutions. The most obvious outcome of a model risk event, namely mark-to-market losses from revised lower valuations and those resulting from incorrect hedging, although having the potential to lead in extreme cases to earning restatements, are usually the easiest for firms to absorb and, beyond the pure monetary loss, are likely to have little lasting impact. It is the indirect results of model risk which can be far more damaging. Senior management routinely make strategic decisions about the allocation of capital between their different businesses and trading desks on the basis of the perceived relative risk and return profiles of the different activities in which their institution engages. Any errors in the models used to reach such decisions may have dramatic consequences if the outcome is erroneous overexposure to a particular business or product. A clear example of this is the massive build up of subprime securities in certain firms which resulted from the incorrect low risk assigned to such products by internal models. Finally, in an era where the branding and image of institutions are all important, the reputational risk following a publicized model-related loss can be immeasurable, especially if resulting from a lack of, or breakdown in, controls, and if followed by regulatory redress.

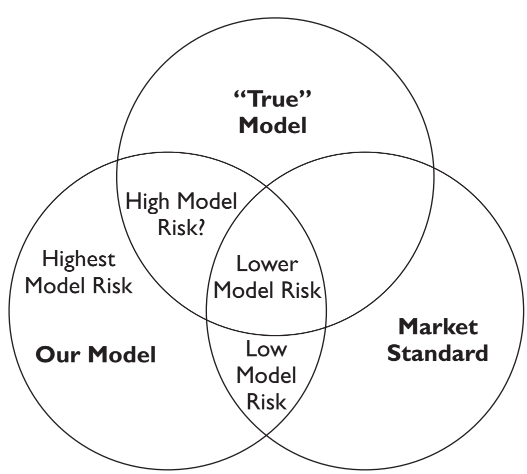
Given the possible impact of model risk, the lack of publications dealing with model risk as a whole is somewhat surprising and we can only make reference here to the excellent articles by Derman (2001) and Rebonato (2002).

THE MARKET STANDARD MODEL

Pricing models are used for the valuation of mark-to-model positions and to obtain the sensitivities to market risk factors of those trades which are mark-to-market through exchanges. In general, model risk can be defined as the loss which results from a misspecified, misapplied, or incorrectly implemented model. For pricing models, the implications of market practices and fair value accounting require us to consider a further definition model risk. Trading books must be mark-to-market and the valuations produced by pricing models must be in line with either market prices or else with those produced by the market standard models once these become visible to all market participants. Pricing model risk can therefore also be defined as the risk that the valuations produced by the model will eventually turn out to be different from those observed in the market (once these become visible) and the risk that a pricing model is revealed to be different from the market accepted model standard. This is also the definition of model risk used in Rebonato (2002). The challenge resides in the fact that the "true" model for valuing a product will never be known in reality. The market standard model itself will evolve over time and the process by which a model becomes accepted as the "standard" is inherently complex and opaque. The Venn diagram in Figure17.1 illustrates the interaction between the model being used to value a trade, the "true" model and the "market standard" model and their associated model risk. This diagram is static, whereas the realities of model risk would be better reflected through a diagram in which each component was moving dynamically and in which all boundaries were blurred.

The market standard model is simply the model which the majority of market players believe to be the closest to the "true" model. Other participants may disagree with the choice of market standard model. Regardless, they must value their positions in line with the appropriate market standard

Figure 17.1 The Interaction Between Models and Model Risk



model until such a time as the majority of market participants actively believe that this standard model is wrong. The reason for this is that, to exit a trading position, participants must either unwind their position with the existing counterparty, or novate it to a different counterparty, or else enter into an opposing trade so as to neutralize the market risk exposure of that position. If all other counterparties believe in a particular market standard model, then the price at which these counterparties will be willing to trade will be dictated by that model. In any case, the widespread use of credit agreements and the posting of collateral between counterparties further binds all participants into valuing their positions consistently with the market standard model. Any significant differences in valuations from the market consensus would lead to difficult discussions with auditors and regulators. Finally, in the presence of a number of knowledgeable and competitive market players, all dictated by the same market standard model, trading desks should not be winning or losing all trades. This would be a clear indication that the model being employed is under- or overpricing the trades.

The market standard model does not have to be invariably correct or the most appropriate for hedging purposes. Trading desks should have the freedom to adopt any model for their hedging purposes as long as all official valuations and risk sensitivities are in line with the market consensus. If a trading desk believes that the market standard model is wrong and that prices in the market are too expensive or cheap, then the desk would trade and risk manage according to their proprietary model with all valuation differences with the market model being isolated and withheld until such a time as the market adopts the new model as the correct standard. The trading desk can only realize the benefits of their superior model once it becomes the market standard and until that time, the desk will need to be able to withstand the losses from valuing their positions at the perceived incorrect market standard model. Model risk losses are thus realized when the market becomes visible and model prices are off-market or else through the implementation of a new model to better reflect the market standard model. Such adjustments tend to occur in sizeable amounts. On the other hand, the incorrect hedging resulting from using a wrong model will tend to result in small but steady losses which will often be both obscured and minimized by the macro hedging of trading books and the netting of long and short risk positions.

The market standard model may not always be apparent and price discovery mechanisms may need to be employed to build up a picture of this model. Participation in market consensus services offers some insights into where other participants are valuing similar products, although care must be taken when interpreting results of such surveys since some contributors may be deliberately biasing their submissions for their own reasons. Such services also tend to focus on the vanilla calibrating instruments rather than the exotic products themselves. Two-way trading of products and sourcing reliable broker quotes provide important data points as will internal information on collateral calls and margin disputes. Working papers and publications by market participants and academics, together with conferences and discussions posted on various internet forums, are all components of the market's perceptions around specific models and their possible replacements. This should all be complemented by market intelligence from the trading desks and sales teams.

Liquidity, the existence of observable market prices, and that of an accepted market standard model are intricately linked. Model risk is directly related to the observability of market prices with very liquid products having very low model risk. It should come as no surprise that model risk will be at its greatest in illiquid markets and for new, complex products (albeit mitigated by the high margins available to such products). However, price observability must always be considered in conjunction with its domain of applicability; model risk can be significant when models are used to extrapolate prices beyond the limits of their domain of observability. Model risk is prevalent when trading is primarily in one direction and consequently for trading books whose exposure to a product is predominantly long or short. In a balanced book, netting of positions would significantly minimize the impact of using a wrong model on overall valuation and market risk, although not on counterparty credit risk. Model risk may also result when the output from one model is used as an input to another model and the assumptions and limitations of the first model are not known to the developer of the second model. Prepayment and historical default models are topical examples given the recent subprime-related losses.

THE EVOLUTION OF MODELS AND MODEL RISK

The evolution of models and model risk can be described through three phases in the life cycle of a product, its valuation models, and their associated model risk. In phase I, the market for the product is completely new and illiquid. Indeed, there are not even any liquid hedging instruments. The models used in this phase will be very simple to allow participants to enter the market quickly and to benefit from the high margins associated with such a new product. There is no price consensus in the market. The lack of liquid hedging instruments results in infrequent recalibration of the model used to price the product in this first phase, which in turn leads to little valuation volatility being observed. One might argue that this phase exhibits the highest model risk, although this should be mitigated through appropriate reserving of the available high margins. The market in phase II has become more established with some price observability but is still illiquid. Margins are tightening and the pricing models are at their most sophisticated in this stage. The market has matured enough to enable the development and trading of liquid hedging instruments leading to frequent recalibration of model parameters and thereby to greater volatility in the valuation of trades. Phase II is characterized by high model risk. Once the market has become well established and liquid, the product enters the third phase in its life cycle which is characterized by small margins and full price observability. The models in this phase act as pure price interpolators and there is low model risk. Phase III does, however, display the greatest valuation volatility as prices continuously readjust to market information. This highlights a paradox in the evolution of model risk, namely, that those products which have the greatest uncertainty in their valuation exhibit the smallest amount of valuation volatility.

MODEL OBSOLESCENCE

The evolution of models just depicted hints at another source of risk, namely, model obsolescence. Models should evolve and improve over time. The concern here is that the market has become more sophisticated and this does not get reflected in a timely manner in the production pricing model. This can be linked to the existence of second-order effects and risk factors which did not need to be taken into account at inception of a new product but which get priced into the market over time as the market matures and margins narrow. Obsolete models often reside in model libraries which, due to changes in personnel or trading desk structures, are no longer being maintained by research and analytic groups but have instead become deeply embedded within technology systems and essentially forgotten. Ensuring adequate ownership of such legacy model libraries and the aged trades being valued on them is a necessary task. The production of model inventory and usage reports, and the decommissioning of obsolete models are key controls to avoid this source of model risk. The decommissioning of outdated models, in particular, can be a controversial subject since traders, research analysts, and software developers are usually reluctant to remove existing functionality from the trading and risk management systems.

The environment in which a model operates must be frequently monitored to ensure that the market has not evolved without corresponding changes in the model. Although the regular re-review and reapproval of models is usually considered best practice by auditors and regulators, this is not necessarily an optimal use of market participants' control resources. If a model has not been materially changed, then there would be little point in carrying out a full revalidation and re-review of the model as its payoff and implementation will not have changed; the same set of inputs will still produce the same set of outputs. In any case, research teams tend to be prolific in their model development and it is unlikely that the model used to value a particular product will not change over the lifecycle of a trade. Instead, the focus should be on the ongoing monitoring of model performance and the regular reappraisal of the suitability of a model for a particular product (and calibration set) given changes in market conditions and in the perception of market participants.

MODEL APPLICABILITY

The use of an obsolete model to value a trade is but one example of model inapplicability: a situation which arises when the model is internally reasonable and well implemented, but is being improperly used. Other examples of model inappropriateness include the booking of trades on the wrong model, the simplified representation of trades to enable the model to be used (i.e., approximate bookings), models being used for an incorrect product subclass (for example, using a high-yield bond option model to value high-grade bond options), and the application of existing models developed for a particular product area to a new and completely different product class (for example, applying interest rate models to commodity and energy products). Model inapplicability may also be related to particular implementation aspects; for example, a model relies on an upper (respectively lower) bound approximation for its solution and is therefore only applicable in the valuation of short (respectively long) positions, or the model solution is only valid for a certain domain of a risk factor and is employed outside of its range of applicability. Furthermore, the use of a model might only be inappropriate under specific market conditions; for example, in high volatility regimes or for steeply declining volatility surfaces. Finally, a very subtle example of model inapplicability relates to the possible lack of convergence of a model. This is often performance related in the case of Monte Carlo simulation models but can also be trade and market data specific; for example, if an asset is near a barrier, then a more densely populated grid is required, or if a coupon payment date is nearing, then a greater number of simulation paths may be required to attain the required level of accuracy.

A STRICT MODEL-PRODUCT SCOPE APPROACH

Enforcing a strict model-product scope when developing, using, and controlling models is crucial to limiting model appropriateness issues. A model cannot be considered in isolation without any reference to the products being valued with that model, or the restrictions under which it is applicable for those products. A model should always be associated with a product and, furthermore, with a set of calibration targets; and, it is that triplet of model-product-calibration instruments which is relevant. Indeed, the model approval process should in reality be product focused: permission is given to value a well-specified product using a particular model employing a specific calibration methodology on a precise set of target, vanilla hedging instruments for that product. This strict product-model-calibration approach must be enforced because the majority of models developed are in reality frameworks which can be applied quite generally to a variety of products using numerous different calibration sets. The temptation with such flexibility is simply to approve the modeling framework with either no, or at best, very general, product descriptions. However, the existence of different modeling frameworks within the same trading and risk management system, each being able and allowed to value the same product using different sets of calibration instruments, makes it very difficult to ensure that the trading desk is not internally arbitraging the models and/or calibration sets. The only solution is a strict product-based approval which strongly binds the product to the model and calibration set. Furthermore, being fully explicit about the product-model-calibration requirements assists the deal review processes, which aim precisely to ensure that each trade is booked correctly, and highlights any model inapplicability concerns.

The implementation of a strict model-product-calibration approach is not without its challenges. This requires the adoption and maintenance of a granular product classification system by both model developers and traders, and places constraints on the allowed representation of products in the trading and risk management systems. These concepts are anathema to model developers and traders and may necessitate efforts to persuade the front office that a change of mindset is required. It should be emphasized that such constraints on the flexibility of trade representation is perfectly compatible with having a scaleable system architecture since the required restrictions on the representation of products only need to be applied at the topmost user booking interfaces where trades would be captured within clearly defined and unique product templates. The issue of model appropriateness is then reflected in that of the product template. In overly flexible systems, the same trade can be correctly represented payoff-wise using several different templates or even using the same general template. This proliferation of possible representations leads to a lack of visibility and a non uniqueness of model use with regards to identical trades from which model applicability problems can easily arise. The extreme case is the common use of scripts which allow payoffs to be expressed in terms of characters, symbols, and even computer code and which all form part of the trade booking itself rather than the system software. Use of such scripts allows the traders to book almost any product using a multitude of models and calibration sets and necessitates detailed technical trade reviews simply to ensure that the payoff has been correctly represented, let alone to ascertain that it is using the appropriate model and calibration set.

A ROBUST MODEL CONTROL FRAMEWORK

The existence of a robust model control framework is essential to mitigate model risk. Controlling the deployment and use of models tends to be overlooked as the less glamorous side of dealing with model risk (compared with the development of ever more sophisticated models to address current model limitations), even though lack of model control probably accounts for the majority of model-related losses across the industry. The cornerstones of a strong and well-joined up model control framework are as follows:

- Clearly defined responsibilities
- An inclusive model approval process
- Adherence to a strict model-product scope
- Calibration control
- Independent model validation
- Model reserving methodologies
- Model release and change management procedures
- Decommissioning of obsolete models
- Model control committees
- Model inventory reports
- Documentary standards
- Audit oversight

In most firms, the responsibilities for the development, use, and control of models is segregated across different groups to prevent conflicts of interest and to leverage the expertise of the different functions. Front office research builds and refines models through a continuous feedback process with the trading desk and information technology (IT) deploys these models into production. The IT groups must be functionally independent from the business units to guarantee the integrity of the pricing model code and executables. Trades on these models will be booked into the systems by middle office and their valuations will automatically flow downstream to feed finance ledgers and risk management systems. On the control side, structured deal review teams carry out an analysis of new trades to ensure that they have been correctly booked; product control reconcile and explain the daily profit and loss on all trades; market and credit risk management measure, monitor and analyze the associated risks; collateral management liaises with counterparties on the need for any additional margin to be posted; client valuation teams provide prices to clients; model validation groups review the pricing models; and valuation control teams independently verify the prices of all positions. Every one of these teams gains insights into the performance of the models being used, and, if properly managed, the existence of such a diverse collection of teams involved with models can lead to a very strong model control culture in which all facets of model usage, appropriateness, and control are continuously and seamlessly monitored, discussed, and resolved as required.

The reality can be far removed from this utopia, often being one of arrogance and disdain by the front office for the back office; of almost unquestioning reverence and acquiescence in return; and of a lack of understanding and intense rivalries between the different groups on the control side. This can lead to the creation of opaque silos which may hinder the dissemination of necessary information between teams, and increases the likelihood that crucial controls do not get addressed anywhere since each group will have the ability to unilaterally formulate its own remit without having to consider the impact on the overall model control environment.

This breakdown in model control culture and processes is in itself a key source of model risk and can only be mitigated by ensuring that the responsibilities for all parties involved in model development, use, and control are clearly defined, well-joined up, and known to all parties. The challenge on the control side is to reconcile the required skills and expectations of the two main teams involved in managing model risks, namely, the model validation and valuation control groups, whilst ensuring the existence of direct oversight over both groups at a realistically manageable level. The model validation group commonly resides within risk management and the aspirations of a typical model validator would be a position as a front office research analyst or as a risk manager. The valuation control group is part of the finance division and a typical career path for a valuation controller would lead up through the finance and product control ranks. Recurring discussions tend to take place about the reporting lines of the model validation group and, in particular, whether this team should be moved into the finance division. Given the differing team aspirations, such a move could be perceived as hindering career choices and might lead to higher staff turnover. In addition, there seems no need to merge these two control teams as long as their responsibilities are clearly articulated and cover all required elements. However, this can be hindered by political rivalries and the fact that the ultimate senior management reporting line for these two teams tend to be the chief risk officer (CRO) and the chief financial officer (CFO), respectively, resulting in the fact that there is often no common reporting line below the executive committee. Given the industry focus on model risk, the time may have come for the creation of a "CMRO" position, the chief model risk officer, whose sole responsibility would be model risk and control. The heads of model validation, valuation control, and any other control team dealing with model issues would report into the CMRO, who would be jointly accountable to the CRO and CFO. This should also be mirrored in the industry with the acceptance of model risk as a separate risk category, rather than trying to consider it as a subset of market, credit, or operational risk.

AN INCLUSIVE MODEL APPROVAL PROCESS

Although a crucial component of any framework to control model risk, the model approval process is often erroneously taken to be the whole framework itself—it is hoped that this chapter has clearly articulated the need for a much more widely encompassing definition. The recommendation for an inclusive model approval process engaging all parties involved in model development, use, and control, and based on product approvals should come as no surprise given the previous discussions. In many firms, a model is considered approved once the model validation group, and only this group, has signed off on the model. Furthermore, the approval may reference a modeling framework only, or, at best, a very general product description. Given the number of groups involved in model use, and the issues of model appropriateness, this is clearly not acceptable anymore. In addition, the front office needs to take clear ownership for its models, their usage, and limitations. The front office decides on the choice of model to use for a particular product but this gets modulated by the control groups during the approval process.

The model approval process should be treated as a product approval process with the product scope of the model placed clearly at the forefront.

Please note that this is separate from approvals granted by new product committees which provide firm-wide permissions to start trading entirely new product classes and consider a much larger set of issues (from infrastructure to legal, compliance, and reputational aspects). A new product approval may be contingent on prior model approval but the model approval process usually deals with products which are already within the trading mandate of a particular business unit. The model approval process should also be used to capture changes to existing models or migrations of positions onto new models. Approval should be requested for a specific product, to be officially valued and risk managed using a particular model and referencing a clearly defined set of calibration targets. As part of this process, clear documentation should be produced by the front office detailing the product (with the payoff made explicit), model theory, numerical implementation, calibration methodology and targets, method for producing risk sensitivities, results of tests carried out, model limitations, model risks, and recommended model reserves to compensate for those limitations and risks. The approvals should be with regards to the information submitted in this documentation, and every group should be able to propose modifications to this document. The whole process should furthermore be embedded within an automated workflow system which stores the documentation, tracks the status of all approvals, notifies the next group in the chain of a new approval request, provides reporting capabilities, and allows for a fully auditable process.

The responsibilities of all approvers need to be made explicit—every group must be fully and unambiguously aware of the meaning and implications of their sign-off. Model consistency must be explicitly considered as part of the approval process; there should only be one model and calibration set approved at any one time for a particular product and the rationale for approving another model and/or calibration set for the same product (without rescinding the existing approval) must be clearly understood and the domain of applicability of the new approval emphasized. Such situations should only occur when considering trading in different economies for which the market behavior is substantially different. But the same product on the same currency should be valued using the same model and calibration set. Although this may seem obvious, the reality is that, since models and approvals are rarely decommissioned, most firms will have a legacy of models which are all approved to value exactly the same product. Full approval should also be withheld until the model/product is fully implemented in the production trading and risk management system on its own product template.

Modeling frameworks should be submitted through this approval process, but with an empty product scope and a clear disclaimer that approval for the framework does not constitute permission to use that framework for official valuations for any specific product and that separate product approvals will be necessary. For practical purposes, having one approval for the modeling framework (for example, a new stochastic volatility model) with an empty product scope then enables product approvals to reference specific sections of this framework documentation, and leads to a more streamlined documentation and approval process. Since the documentation detailing the product-model under approval needs to be disseminated to an audience which is even wider than the approvers, sensitive theoretical and numerical implementation sections can always be brief as long as they reference other more detailed documents which can then be access controlled to the relevant parties.

Model risk is inherent not only in the use of models but also in their construction. Conceptually, a pricing model can be considered as having three components, namely, the input data, the model engine, and the output data. All three model components are possible sources of model risk and the mitigation of these particular risks is typically addressed through the independent validation of the model, and we refer the reader to Whitehead (2010) for details on the necessary concepts to mitigate these risks through the model validation process.

CODE CONTROL AND REGRESSION TESTING

Change management, model release, and calibration control procedures ensure the integrity of the pricing codes, and pricing model executables. Such procedures do not aim to address the appropriateness of the pricing models and calibration choices themselves, but rather to verify that the approved models and calibrations are the ones actually used in production, that any changes to the model code are audited and have only been made by authorized personnel, and that any proposed changes to the model executables or calibration sets have to go through a rigorous process prior to release into production. These procedures are necessary to minimize the possibility of not only accidental, but especially deliberate, manipulation of the system to hide losses or inflate gains. Common features of such procedures include the requirements that code must be kept under a source control database with access levels regularly reviewed, that the rationale for code changes must be documented and that the production version of the model must reside in a secure location. Regression testing between the current production version of the model and its proposed replacement provides not only an assessment of the materiality of the proposed model changes on prices and risk sensitivities, but also serves as a tool to confirm that only those advertised changes have in fact been made. This regression testing must be carried out not only against the current portfolio of trades, but also against a static portfolio purposely built so that every material portion of code is covered by a regression test. For example, consider the situation in which the only live barrier positions are of type "up and in call"; a regression test which only uses the currently traded portfolio would not capture changes to the code used to price "down and in put" options. The construction of such a static portfolio of test cases can be onerous given the numerous features which most products contain. The author's suggestion would be to leverage the comprehensive tests carried out by the model validation group as part of their validation process. Finally, all relevant control groups must explicitly sign off on the proposed production release to certify awareness of the impact of the new release and to confirm that this release does not contravene other model control policies.[1]

MODEL GOVERNANCE

The governance structure placed around the use of models will be critical to foster a strong, transparent and consistent model control environment, and to minimize the impact of model risk-related events. The governance of model risk must not only provide comfort to senior management that model-related issues will receive the appropriate focus and escalation, but must also facilitate the awareness of senior executives to, and their active involvement in, model related issues. The need for a new position of Chief Model Risk Officer has already been highlighted to unite the different model-focused control groups under the same hierarchy and to raise the profile of model control. Oversight of the model control environment occurs through the aegis of model control committees which meet typically monthly and are meant to monitor the model control environment. The challenge with such committees is to ensure senior management involvement whilst ensuring that the committees remain at a manageable level. Creation of a CMRO can only strengthen these committees. New product committees are an integral part of all institutions and participation of the relevant model control staff in such committees is required to address model-related issues. Good governance also demands the implementation of strong documentary standards to facilitate the role of all parties, to create a corporate record and mitigate key person risk. Such documentary standards should not only include detailed policies and procedures for all model control aspects and the use of model documentation standards, but also the requirement that all model code be explicitly commented by the developer at inception. The final point is the need to have full internal audit oversight of every aspect of model use, approval, and control.

CONCLUSION

The approaches discussed in this chapter to mitigate model risk have concentrated on the overall control framework and, as such, apply equally to all trades. Any additional efforts at mitigating model risk should be focused on particular trades and given resource constraints, must be aimed at those trades which are deemed to be exposed to the highest model risk. This requires the automated calculation of model risk metrics for each trade and the existence of methodologies for the assessment and measurement of model risk. In any case, one cannot hope to fully control what is unquantifiable. Techniques for measuring model risk can be divided into qualitative approaches (assigning subjective ratings to models) and quantitative methodologies which postulate the use of specific model risk loss functions but implicitly require the production of valuations under a number of different models (please refer to Kerkhof, Melenberg, and Schumacher, 2002 for details of such an approach). Unfortunately, much progress remains to be made in this area and most firms are not able to quantify model risk on an automated basis. As a proxy, model inventory reports can be used with aggregated standard metrics such as present value, notional, and associated sensitivities. If such measurements were possible, then firms could address model risk by implementing a reserving methodology based on their model risk metrics. The difficulties involved in carrying out precisely such an automated quantification of model risk for traded positions are implicitly acknowledged by regulators who do not currently require the calculation of an explicit regulatory capital charge related to model risk. Instead, model risk is implicitly captured in the setting of the various "multipliers" applied to the regulatory capital charges for market and credit risk, reflecting an assessment of the model controls implemented in each firm.

REFERENCES

Derman, E. (2001) Markets and Models. *Risk*, July: 48–50.

Kerkhof, J., Melenberg, B., and Schumacher, H. (2002) Model Risk and Regulatory Capital. Working paper, Tilburg University, The Netherlands.

Rebonato, R. (2002) Theory and Practice of Model Risk Management. In Field, P. (ed.), *Modern Risk Management: A History*. London: RiskWaters Group.

Whitehead, P. (2010) Concepts to Validate Valuation Models. In Gregoriou, G., Hoppe C., and Wehn, C. (eds.) (2010). *The Risk Modeling Evaluation Handbook*. McGraw-Hill.

NOTES

1. For example, if a firm's policy is to require all model changes to be fully approved prior to being placed into production, then a proposed production release which would lead to unapproved changes being used in production would have to be rejected.

PART V

LIMITATIONS TO MEASURE RISK

This page intentionally left blank

CHAPTER 18

BEYOND VALUE AT RISK
Expected Shortfall and Other Coherent Risk Measures

Andreas Krause

ABSTRACT

Value at risk is the leading risk measure in financial markets, despite shortcomings that may discourage the diversification of risk and could provide incentives to make more risky investments. Coherent risk measures have been developed that have properties which are theoretically superior and should reduce model risk arising from the shortcomings in the value-at-risk measure. This chapter will provide an overview of the most commonly used coherent risk measures in financial markets as well as in insurance. We will discuss the properties of these risk measures and also evaluate the quality of their estimation.

THE PROBLEM WITH VALUE AT RISK

Value at risk (VaR) has become the most commonly used risk measure in the banking industry since its introduction in the mid–1990s. The attraction of VaR is that it focuses on potential losses—the main concern of risk managers and regulators—and its interpretation is easily accessible to non-specialists. As with all risk measures that have been proposed in the past, it is not without its flaws. Apart from econometric problems in estimating VaR and the fact that confidence intervals around the VaR estimate are substantial in most cases, there are also problems with the properties of VaR;

Krause (2003) provides an overview. To illustrate the first of the main conceptual problems, let us consider the example given in Danielsson et al. (2005) with the assets having a payoff of X_i

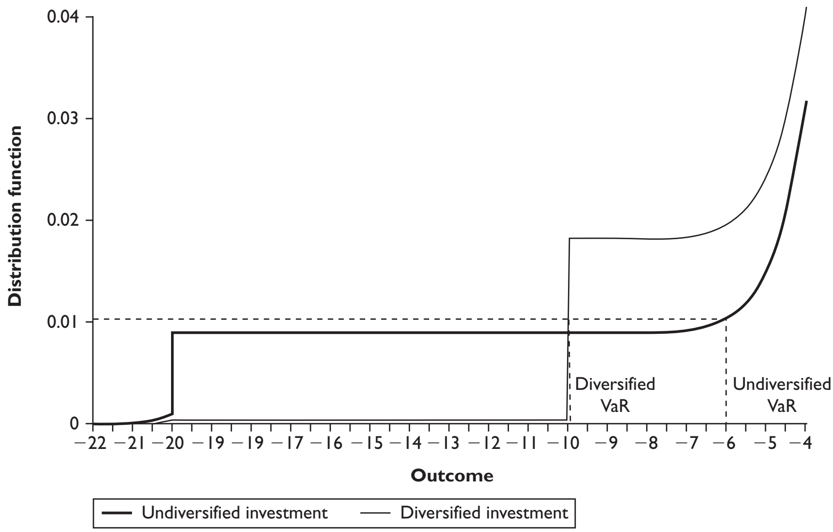
$$
\begin{aligned}
X_i &= \varepsilon_i + \eta_i \\
\varepsilon_i &\sim N(0,1) \\
\eta_i &= \begin{cases} 0 & \text{w. p. } 0.991 \\ -10 & \text{w. p. } 0.009 \end{cases} \\
Cov\left[\varepsilon_i, \varepsilon_j\right] &= Cov\left[\varepsilon_i, \eta_i\right] = Cov\left[\varepsilon_i, \eta_j\right] = Cov\left[\eta_i, \eta_j\right] = 0
\end{aligned}
\tag{18.1}
$$

Let us consider the 99 percent VaR of an investment into two units of asset X_1. In this case 0.9 percent of the potential losses are covered by the shock η_i and the outcome ε_i can only contribute 0.1 percent of the losses toward VaR, giving rise to a VaR of 6.2 for those two units of asset X_1. Suppose now that the investor diversifies the portfolio and invests into one unit of asset X_1 and one unit of asset X_2. In this case we can show that the VaR of this investment is 9.8. The reason for the larger losses in the second portfolio is that the losses from the shocks η_i are now above the 1 percent threshold, while in the first portfolio they were below this 1% level; thus the substantial increase in VaR; Figure 18.1 illustrates this problem. We observe then that holding a diversified portfolio does not necessarily reduce VaR.

This result is driven by the presence of large shocks which are relatively frequently occurring, in the above example with a probability of 0.9 percent. It can easily be argued that for the vast majority of cases in the financial industry such large shocks do not occur that frequently, if at all. However, an example given in Breuer (2006) shows that the VaR of a portfolio consisting of a short position in out-of-the-money call options and a short position in out-of-the-money put options can have a VaR that is larger than the VaR of the individual positions. In general we observe similar problems when trading nonlinear asset such as options or other derivatives. Therefore, trading derivatives positions, which obviously is a very common situation for most financial institutions, may actually result in situations where the use of VaR discourages diversification of investments. This is in contrast to the usual perception that diversification reduces risk and should thus be encouraged.

Another problem with VaR is that it might actually encourage taking larger risks than a decision maker would choose when using other risk measures. Consider two probability distributions with the same VaR but in the tail of the distribution beyond which one might actually produce potentially larger losses than one would with the other, i.e., would

Figure 18.1 Illustration of a Case Where Diversification Leads to a Higher Value at Risk



generally be regarded as riskier. With the general relationship between risk and returns being positive and the two VaRs being identical, an investor would choose the investment with the higher expected return, i.e., the more risky project. The reason for this result is that VaR ignores the size of any losses beyond VaR.

Given these shortcomings of VaR, we will in this chapter discuss a number of risk measures which have been proposed and are not subject to the problems mentioned above. By avoiding these problems, we can assure that risk is modeled adequately in all situations and thereby reducing the model risk arising from the risk measure itself. After discussing general properties risk measures should have in the coming section, we will explore in more detail the risk measure which is closest to VaR, expected shortfall, in this chapter's third section. The fourth section explores the more general class of distortion risk measures, of which expected shortfall and VaR are special cases, before we cover lower partial moments in this chapter's fifth section. We conclude our discussion in the final section of this chapter.

COHERENT RISK MEASURES

Given the problems with VaR as detailed above, Artzner et al. (1997, 1999) suggested a number of desirable properties a risk measure should have in order to be a good risk measure. They called risk measures with such properties *coherent*.

Axioms

The first of these required properties can be deduced by comparing two investments, X and Y. Suppose that for any scenario the outcome of X never exceeds the outcome of Y, i.e., $X \leq Y$, then X is more risky than Y. A risk measure $\rho(\cdot)$ has to reflect this property.

Axiom 1 (Monotonicity): If $X \leq Y$, then $\rho(X) \geq \rho(Y)$.

Suppose that we have two investments, X and Y, whose outcomes for any scenario differ only by the fixed amount α, hence $Y = X + \alpha$. The possible losses of Y, compared to X, are reduced by α, thus reducing the risk by α. Hence investing an amount of α into a riskless asset reduces the risk by that amount.

Axiom 2 (Translation invariance): For any constant α, $\rho(X + \alpha) = \rho(X) - \alpha$.

From setting $\alpha = \rho(X)$, we see that $\rho(X + \rho(X)) = 0$, thus justifying the need for a capital requirement of $\rho(X)$ to cover the risks associated with an investment X. It is reasonable to propose that the size of a position has no influence on the characteristics of the risks associated with it, all possible outcomes are transformed proportionally with the position size and so is the risk. This inference obviously neglects potential losses arising from the imperfect liquidity of markets.

Axiom 3 (Scale invariance): For all $\lambda \geq 0$, $\rho(\lambda X) = \lambda\rho(X)$.

When comparing the risk of a combination of two investments X and Y, $X + Y$, with the individual risks of each investment, it should be that no additional risks arise from this combination. The reasoning behind this requirement is the usual argument that diversification reduces the risks involved.

Axiom 4 (Sub-additivity): $\rho(X + Y) \leq \rho(X) + \rho(Y)$.

A risk measure fulfilling these four axioms is called a *coherent risk measure*. Such coherent risk measures exhibit properties that should ensure that choices which are commonly regarded as more risky, receive a higher value of the risk measure.

It was furthermore shown in Acerbi (2004) that any convex combination of coherent risk measures is again a coherent risk measure, i.e., $\rho(X) = \sum_i \lambda_i \rho_i(X)$ with $\sum_i \lambda_i = 1$ and $\lambda_i \geq 0$ is coherent if all $\rho_i(X)$ are coherent.

Coherence of Common Risk Measures

We have shown with the example above that VaR is not sub-additive; hence it is not a coherent risk measure. It can in fact be shown that VaR is fulfilling all the other axioms of coherence, but it is the lack of sub-additivity that fails to ensure the proper assessment of risks.

The other risk measure commonly used in finance is the standard deviation of returns. As we know that $\text{var}[X + \alpha] = \text{var}[X]$, it is obvious that the standard deviation is not translation invariant and therefore not a coherent risk measure. The same can easily be shown for the covariance of returns with the market and therefore the commonly used β that measures the systematic risk. The problem with these risk measures is two assets with the same standard deviation or β but very different means are supposed to have the same risk.

Given that the most commonly used risk measures VaR, standard deviation and β are not coherent and thus do not ensure a proper risk assessment, we will in the coming sections of this chapter explore a range of coherent risk measures that have been proposed.

Limitations of Coherent Risk Measures

Although coherent risk measures are mostly consistent with the preferences of their users, the coherence axioms have been subject to criticism, see, e.g., Krause (2002). The main focus of this critique has been the requirement of axiom 4, sub-additivity. The argumentation has been that in the presence of catastrophic risks, this sub-additivity is not necessarily fulfilled; see Rootzen and Klüppelberg (1999) and Danielsson (2002).

The essence of the argument is that with catastrophic losses causing the failure of companies, regulators have to include systemic risks in their assessment. Systemic risks arise when not all losses can be covered by the company, thus causing creditors to bear parts of these losses, which in turn may force them into bankruptcy. Companies treat systemic risks as externalities which are not further considered.

When a large number of catastrophic risks are covered by small independent companies, losses to creditors are also relatively small and hence systemic risks negligible, provided the risks are sufficiently independent of each other. On the other hand, if a single company faces a large number of these risks, each of which may cause the company to fail, the losses to creditors are much larger given the increased size of the company. Therefore the systemic risk is increased, despite the diversification of the company into several (catastrophic) risks, an obvious violation of sub-additivity. The above example shows the importance of an appropriate perspective on risks. Companies exclude systemic risks in their considerations, concentrating only on the risks faced by their organization. Therefore the single company above took on additional risks; while from a regulator's point of view, this was merely a reallocation of risks between organizations which obviously affected the systemic risk. There are, however, no additional risks taken as required for the applicability of sub-additivity. Hence with a proper view of risk, sub-additivity has not been violated.

Although coherent risk measures provide a framework to choose appropriate risk measures based on theoretical considerations, it has to be ensured that the definition of risk suits the perspective of its user. Once this definition has been made appropriately, coherent risk measures can provide a risk assessment in accordance with preferences.

EXPECTED SHORTFALL

One of the advantages of value at risk is that the concept is easily understood and the value itself has an instant meaning. When introducing a coherent risk measure, it would be beneficial if such properties could be retained as much as possible. Based on VaR, a similar measure, called expected shortfall, has been developed, which is a coherent risk measure. Although expected shortfall is a special case of the distortion risk measures we will discuss in the coming section, its role as one of the most prominent coherent risk measures in finance deserves a more detailed treatment.

Definition and Properties

Value at risk is commonly defined implicitly as the c-quantile of the distribution of possible outcomes X

$$\text{Prob}(X < -\text{VaR}_c) = c \tag{18.2}$$

A more formal definition of VaR would be

$$\text{VaR}_c = -\inf\left\{x \middle| \text{Prob}(X \le x) \le c\right\} \tag{18.3}$$

Based on VaR we can now define a new risk measure which does take into account the size of potential losses beyond the VaR

$$\text{ES}_c = -\text{E}\left[X \middle| X \le -\text{VaR}_c\right] = \frac{1}{c}\int_0^c VaR_i\,di \tag{18.4}$$

This measure, known as *expected shortfall* or *tail conditional expectation* for continuous outcome distributions, measures the expected losses, given that VaR has been exceeded. Using the expected shortfall overcomes one of the problems VaR faces, namely, that the size of losses beyond its value is not considered. It has furthermore been shown in Acerbi and Tasche (2002a, 2002b) that expected shortfall is actually a coherent risk measure, thus also dealing with the lack of sub-additivity of VaR.

Despite being a coherent risk measure, and thus superior to VaR, expected shortfall still retains much of the intuitive meaning that made VaR so popular. While VaR is commonly interpreted as the amount that can reasonably be lost, where "reasonable" is determined via the confidence level c, we can interpret expected shortfall as the average amount that is lost once VaR is exceeded. Thus expected shortfall represents the average loss if the losses are exceptionally high.

Tail Risk

One important aspect, apart from measuring risk appropriately, is whether the risk measure orders different risks in accordance with the preferences of individuals. In accordance with Yamai and Yoshiba (2002b, 2002c) we can define a *tail risk* as the possibility of choosing the wrong alternative due to the risk measure underestimating the risk from fat tails.

In order to evaluate the choices we can use the concept of stochastic dominance. We can recursively define nth–*order stochastic dominance*, by setting $F_X^1(z) = F_X(z)$ and $F_X^n(z) = \int_{-\infty}^{z} F_X^{n-1}(t)dt$:

$$X \geq_n Y \Leftrightarrow \forall z : F_X^n(z) \leq F_Y^n(z) .$$

One of the key results from utility theory is that if X dominates Y by nth–*order stochastic dominance*, and for all $k = 1, \ldots, n$ we have $(-1)^k U^{(k)}(x) \leq 0$, then $E[U(X)] \geq E[U(Y)]$, i.e., the choice is fully compatible with the preferences of individuals; see Ingersoll (1987, pp. 138–139). A risk is called consistent with nth–*order stochastic dominance* if $X \geq_n Y \Leftrightarrow \rho(X) \leq \rho(Y)$.
We can now more formally introduce the notion of tail risk. Suppose $\rho(X) < \rho(Y)$ and $\forall\ z \leq Z$ for some threshold Z we find that $F_X^n(z) \leq F_Y^n(z)$. In this case we say that the risk measure $\rho(\cdot)$ is free of n^{th}–order tail risk; thus a risk measure without n^{th}–order tail risk is consistent with n^{th}–order stochastic dominance.

It has been shown in Yamai and Yoshiba (2002b) that VaR is in general only consistent with first order stochastic dominance, thus it cannot be ensured that VaR ranks choices according to their downside risk. In contrast to this, expected shortfall is consistent with second-order stochastic dominance and thereby always ordering choices by their downside risk. While using expected shortfall will not ensure making choices always in accordance with preferences, it will provide the correct choice in a much wider range of situations.

Under certain restrictions, however, VaR will also be a consistent risk measure under second-order stochastic dominance. If the underlying distribution of two choices is from the same class of elliptical distributions and the distributions have the same mean, VaR will be consistent with second-order stochastic dominance. While such distributions might be appropriate for a wide range of financial assets and the means will in most cases not differ substantially to make a meaningful difference to the outcome, they are often not adequate for the use of derivatives as well as the presence of rare but extreme events similar to that used in the introductory example above.

It is thus not only that expected shortfall is superior in terms of being a coherent risk measure, it is also superior in that it allows decision making in accordance with preferences. It is, however, not the only risk measure which fulfills these conditions (as we will see in the following two sections) but generally regarded to be the most accessible and intuitively understood risk measure for financial markets.

Estimating Expected Shortfall

When estimating expected shortfall, we are interested in observations at the lower tail of the distribution, i.e., extreme values. It is therefore natural to use extreme value theory as a way to estimate expected shortfall. Extreme value theory suggests that beyond a threshold u, the distribution of outcomes follows a Generalized Pareto Distribution as $u \to \infty$:

$$G(x) = \begin{cases} 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-\frac{1}{\xi}} & \text{if } \xi \neq 0 \\ 1 - e^{-\frac{x}{\beta}} & \text{if } \xi = 0 \end{cases} \tag{18.5}$$

here β is the shape parameter, taking a similar role to the standard deviation for a normal distribution, and ξ determines the tail of the distribution; a value of $\xi > 0$ denotes fat tails.

If we define a threshold u beyond which we estimate the Generalized Pareto Distribution, Yamai and Yoshiba (2002c) derived that

$$VaR_c = u + \frac{\beta}{\xi}\left(\left(\frac{c}{\frac{N_u}{N}}\right)^{-\xi} - 1\right) \tag{18.6}$$

where N_u denotes the number of observations exceeding the threshold u and N being the sample size. The expected shortfall can then be estimated as

$$ES_c = \frac{VaR_c}{1-\xi} + \frac{\beta - \xi u}{1-\xi} \tag{18.7}$$

The standard error of the estimation of expected shortfall is generally slightly larger than the standard error of the VaR estimate. From Kendall (1994) we can derive that for the VaR estimate we get as the standard error

$$\sigma_{VaR} = \frac{1}{f\left(q_c\right)} \sqrt{\frac{c\left(1-c\right)}{N}} \tag{18.8}$$

where q_c denotes the c-quantile and $f(-)$ the density function. For the expected shortfall we need to define a quantile $d > c$ and then Yamai and Yoshiba (2002a) obtain for sufficiently large sample sizes N

$$\sigma_{ES}^2 = \frac{1}{N}\left(\frac{cq_c^2 + (1-d)q_d^2}{\left(c-d\right)^2} + \frac{1}{(c-d)^2}\int_{q_c}^{q_d} t^2 f(t)dt - \frac{\left((1-d)q_d + cq_c + \int_{q_c}^{q_d} tf(t)dt\right)^2}{(c-d)^2}\right) \tag{18.9}$$

Comparing the standard errors of VaR and expected shortfall, relative to their respective mean estimate for a student-t distribution of outcomes, we find that the relative standard errors of the expected shortfall are slightly below that of VaR, due to the expected shortfall having a larger mean estimate. As Yamai and Yoshiba (2002a) pointed out, the above standard errors are only correct for large sample sizes; for sample sizes that we consider here, they are shown to be slightly larger. Overall we can deduct that expected shortfall and VaR have comparable estimation errors, which for small quantiles can be quite substantial.

DISTORTION RISK MEASURES

Expected shortfall as well as VaR are only special cases of a wider class of risk measures, distortion risk measures, that we will introduce in this section. General properties of these risk measures to ensure that they are coherent as well as consistent with second-order stochastic dominance are also considered.

Definition and Use

A distortion risk measure, introduced by Wang (1996), adjusts the distribution function such that large losses receive a higher weight than their actual probability in the risk analysis. Define a distortion function g: $[0,1] \rightarrow [0,1]$ which is nondecreasing with $g(0) = 0$ and $g(1) = 1$ and with F denoting the distribution function we set $S = 1 - F$ as the decumulative distribution function. We can define

$$\rho(X) = \int_{-\infty}^{0} g(S(x))\,dx \tag{18.10}$$

as a risk measure of a risky variable X, where we only consider losses for simplicity.[1]

If we use different distortion functions we will obtain a wide range of risk measures, VaR and expected shortfall included; see Wirch and Hardy (1999).

Other risk measures can be constructed using different distortion functions; Table 18.1 provides an overview of the most commonly used distortion risk measures. We find that VaR as well as expected shortfall only use information on the distribution up to the c-quantile. Other distortion functions, however, use the entire distribution and thus use information also on the size of "reasonable"losses to make an optimal choice; for those risk measures that use information on the entire distribution we find that $\forall\, t < 1$: $g(t) < 1$.

Distortion risk measures can be estimated easily by using a discrete approximation of the integral in Equation (18.10) through estimating a range of quantiles and then applying the distortion function. The standard error of such an estimation can in general only be determined by using Monte Carlo simulations, although some attempts to find analytical solutions have been made; see Gourieroux and Liu (2006).

It would be important to establish which of the distortion risk measures fulfills the required properties, coherence, and consistency with at least second-order stochastic dominance.

Concave Distortion Functions

Wirch and Hardy (1999) have shown that concave distortion functions generate coherent risk measures, i.e., for any $0 \leq \lambda \leq 1$ we find $g(\lambda x + (1 - \lambda)y) \geq \lambda g(x) + (1 - \lambda)g(y)$. The concavity of the distortion function is a necessary and sufficient condition for the risk measure to be coherent.

It is obvious that the distortion function for VaR is not concave and therefore confirms that VaR is not a coherent risk measure. On the other hand, the

Table 18.1 Overview of Distortion Risk Measures

Risk Measure	Distortion Function	Restrictions for Coherent Risk Measures
Beta-distortion	$g(t)=\int_0^t \frac{1}{\beta(a,b)} v^{a-1}(1-v)^{b-1}\,dv$	$\beta(a,b)=\int_0^1 v^{a-1}(1-v)^{b-1}\,dv$ $a \leq 1, b \geq 1$
Proportional hazard transform	$g(t) = t^a$	$a \leq 1$
Dual power function	$g(t) = 1 - (1 - t)^a$	$a \geq 1$
Gini principle	$g(t) = (1 + a)t - at^2$	$0 \leq a \leq 1$
Wang transform	$g(t) = \Phi(\Phi^{-1}(t) - \Phi^{-1}(c))$	
Lookback	$g(t) = t^a(1 - a \ln t)$	$0 < a < 1$
Expected shortfall	$g(t)=\begin{cases}1 & \text{if } c<t\leq 1\\ \frac{t}{c} & \text{if } 0<t<c\end{cases}$	
Value at risk	$g(t)=\begin{cases}1 & \text{if } c<t\leq 1\\ 0 & \text{if } 0<t<c\end{cases}$	

distortion function of expected shortfall is concave, thus expected shortfall is a coherent risk measure as outlined above. From Table 18.1 we see that a wide range of other distortion risk measures also give rise to coherent risk measures; hence, expected shortfall is only one of many candidates with the desired properties to choose from. The attractiveness of expected shortfall in finance is its simplicity in interpreting the results obtained by nonspecialists while the other risk measures are mostly used in actuarial sciences to determine insurance premia. These risk measures are much less intuitively understood and for that reason not widely used outside their original remit.

Risk measures which have a distortion function that is strictly concave, i.e., for any $0 < \lambda < 1$ we find $g(\lambda x + (1 - \lambda)y) > \lambda g(x) + (1 - \lambda)g(y)$, can be shown to preserve second-order stochastic dominance strongly. If the distortion function is not strictly concave, the risk measure preserves second-order stochastic dominance only weakly, thus we may find $X \geq_2 Y \Rightarrow \rho(X) = \rho(Y)$. Clearly expected shortfall preserves second order stochastic dominance only weakly, as do the other coherent distortion risk measures presented in Table 18.1. If the restrictions on the distortion functions are tightened such that they become strict inequalities, second-order stochastic dominance is preserved strongly, except for VaR and expected shortfall. There has been some recent progress in Bellini and Caperdoni (2007) to establish consistency with higher order stochastic dominance, in particular third-order stochastic dominance. The results so far are however limited and it is not clear whether apart from the expected value, i.e., $g(t) = t$ any other distortion functions do exist that are consistent with third-order stochastic dominance.

We thus find distortion risk measures to exhibit desirable properties that can easily be verified: strictly concave distortion functions generate coherent risk measures that preserve second-order stochastic dominance strongly. Despite these desirable properties, Darkiewicz et al. (2003) show that concave distortion functions in general do not preserve the correlation order, potentially causing problems for portfolios of assets. However, they point out that for most distortion functions this has no real practical relevance and therefore distortion risk measures can be used for portfolios.

LOWER PARTIAL MOMENTS

One problem when using expected shortfall as well as a range of other coherent distortion risk measures is that they in general are only consistent with second-order stochastic dominance. We can use lower partial moments to address this problem. Define

$$\sigma_n = \int_{-\infty}^{K} \left(K - t\right)^n f(t)dt \tag{18.11}$$

as the n^{th} lower partial moment of the distribution of outcomes X with some benchmark K. It can be shown that using the n^{th} lower partial moment ensures choices are made in accordance with $(n + 1)^{th}$–order stochastic dominance. However, the n^{th} lower partial moment is not a coherent risk measure, thus missing one of the important characteristics of a good risk measure.

It has, however, been shown in Fischer (2001) that convex sums of lower partial moments are coherent risk measures. Define a risk measure as

$$\rho(X) = -\mu + \sum_{i=1}^{+\infty} a_i \sigma_i + a_\infty \sigma_\infty \tag{18.12}$$

where $\sum_{i=1}^{+\infty} a_i + a_\infty \leq 1$, which is a coherent risk measure. Please note that it is not necessary to consider all lower partial moments of the distribution as we can set a large number of coefficients equal to zero and use a weighted average of a small number of lower partial moments, usually restricted to first and second lower partial moments, sometimes also the third or fourth moment, as well as σ_∞, which is the measure for the maximum loss. Although such risk measures are coherent, it is not ensured that they are also consistent with stochastic dominance, an issue that merits further research.

CONCLUSION

Value at risk has become the most popular risk measure, not least because of the regulatory framework set in Basel II, besides the intuitive meaning it has to nonspecialists. Over time it has been shown, however, that its properties are not always desirable and that it can be manipulated, particularly when using derivatives, to show a lower risk of a position than it actually has. Furthermore, it does not necessarily encourage diversification of risks and is in general not compatible with the preferences of individuals. These problems arose as the result of VaR not being coherent and not being consistent with second-order stochastic dominance.

An alternative which is coherent and weakly preserves second-order stochastic dominance was found to be expected shortfall, the average loss once VaR has been exceeded. Expected shortfall has the advantage of being nearly as easily accessible as VaR while overcoming most of its theoretical shortcomings and at the same time its estimation errors are approximately equivalent. Expected shortfall is however only a special case of the much larger class of coherent distortion and risk measures. These wider classes of risk measures have the advantage of using information on the entire distribution of outcomes rather than only information up to the c-quantile as expected shortfall does.

A drawback of distortion and spectral risk measures is that they are not easily understood and interpreted, unlike expected shortfall. While these risk measures are widely used in actuarial sciences, they are nearly unknown in the wider financial world and by investors. These risk measures have additionally been developed to suit the specific needs of users in the insurance industry and so not surprisingly do not appeal to the needs of investors in financial markets. As the class of distortion and spectral risk measures is potentially very large, it leaves the prospect open for newly developed risk measures that are able to address the needs of investors.

Investors who are more inclined to use moments of their distribution of outcomes to assess the risk of a position can use a weighted average of lower partial moments as an alternative coherent risk measure, only requiring small adjustments to the moments commonly used in finance.

REFERENCES

Acerbi, C. (2002) Spectral Measures of Risk: A Coherent Representation of subjective Risk Aversion. *Journal of Banking and Finance*, 26(7): 1505–1518.

Acerbi, C. (2004) Coherent Representations of Subjective Risk-Aversion. In G. Szego (ed.), *Risk Measures for the 21st Century*. New York: Wiley; 147–207.

Acerbi, C. and Tasche, D. (2002a) Expected Shortfall: A Natural Coherent Alternative to Value at Risk. *Economic Notes*, 31(2): 1–10.

Acerbi, C. and Tasche, D. (2002b) On the Coherence of Expected Shortfall. *Journal of Banking and Finance*, 26(7): 1487–1503.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1997) Thinking Coherently. *Risk*, 10(11): 68–71.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999) Coherent Risk Measures. *Mathematical Finance*, 9(3): 203–228.

Bellini, F. and Caperdoni, C. (2007) Coherent Distortion Risk Measures and Higher-Order Stochastic Dominances. *North American Actuarial Journal*, 11(2): 35–42.

Breuer T. (2006) Providing Against the Worst: Risk Capital for Worst Case Scenarios. *Managerial Finance*, 32(9): 716–730.

Danielsson, J. (2002) The Emperor Has No Clothes: Limits to Risk Modelling. *Journal of Banking and Finance*, 26(7): 1273–1296.

Danielsson, J., Jorgensen, B.N., Samorodnitsky, G., Sarma, M., and de Vries, C.G. (2005) Sub-additivity Re-Examined: the Case for Value-at-Risk. Working paper, London School of Economics.

Darkiewicz, G., Dhaene, J. and Goovaerts, M. (2003) Coherent Distortion Risk Measures—A Pitfall. Working paper, KU Leuven, Belgium.

Dowd, K. and Cotter, J. (2007) Exponential Spectral Risk Measures. *ICFAI Journal of Mergers and Acquisitions*, 5(4): 57–66.

Fischer, T. (2001) Examples of Coherent Risk Measures Depending on One-Sided Moments. Working paper, Department of Mathematics, TU Darmstadt, Germany.

Gourieroux, C. and Liu, W. (2006) Sensitivity Analysis of Distortion Risk Measures. Working paper, University of Toronto, Canada.

Gzyl, H. and Mayoral, S. (2006) On a relationship Between Distorted and Spectral Risk Measures. Munich Personal RePEc Archive 1940.

Ingersoll, J.E. (1987) *Theory of Financial Decision Making*. Lanham, MD: Rowman & Littlefield.

Kendall, M. (1994) *Kendall's Advanced Theory of Statistics*. New York: Wiley.

Krause, A. (2002) Coherent Risk Measurement: An Introduction. *Balance Sheet*, 10(4): 13–17.

Krause, A. (2003) Exploring the Limitations of Value at Risk: How Good Is It in Practice' *Journal of Risk Finance*, 4(2): 19–28.

Rootzen, H. and Klüppelberg, C. (1999) A Single Number Can't Hedge Against Economic Catastrophes. *Ambio*, 28(6): 550–555.

Wang, S.S. (1996) Premium Calculation by Transforming the Layer Premium Density. *ASTIN Bulletin*, 26(1): 71–92.

Wirch, J.L. and Hardy, M.R. (1999) Distortion Risk Measures: Coherence and Stochastic Dominance. Working paper, Department of Actuarial Mathematics and Statistics, Heriot-Watt University, U.K.

Yamai, Y. and Yoshiba, T. (2002a) Comparative Analyses of Expected Shortfall and Value-at-Risk: Their Estimation Error, Decomposition, and Optimization. *Monetary and Economic Studies*, 20(1): 87–122.

Yamai, Y. and Yoshiba, T. (2002b) Comparative Analyses of Expected Shortfall and Value-at-Risk (2): Expected Utility Maximization and Tail Risk. *Monetary and Economic Studies*, 20(2): 95–116.

Yamai, Y. and Yoshiba, T. (2002c) Comparative Analyses of Expected Shortfall and Value-at-Risk (3): Their Validity under Market Stress. *Monetary and Economic Studies*, 20(3): 181–238.

NOTES

1. A related class of risk measures are *spectral risk measures*; see Acerbi (2002, 2004) and Dowd and Cotter (2007). These risk measures are defined as $\rho(X) = -\int_0^1 \varphi(p) q_p \, dp$, where q_p denotes the p-quantile and we require that $\int_0^1 \varphi(p) dp = 1$. The different spectral risk measures differ only in their risk spectrum φ. Gzyl and Mayoral (2006) have shown that spectral and distortion risk measures are identical if $g' = \varphi$.

This page intentionally left blank

CHAPTER 19

VALUE AT RISK COMPUTATION IN A NONSTATIONARY SETTING

Dominique Guégan

ABSTRACT

This chapter recalls the main tools useful to compute value at risk (VaR) associated with an m-dimensional portfolio. The limitations of the use of these tools are explained when nonstationarities are observed in time series. Indeed, specific behaviors observed by financial assets, such as volatility, jumps, explosions, and pseudo-seasonalities, provoke nonstationarities which affect the distribution function of the portfolio. Thus, a new way for computing VaR is proposed which allows the potential noninvariance of the m-dimensional portfolio distribution function to be avoided.

INTRODUCTION

Value at risk (VaR) is now a major task of much financial analysis involved in risk management. It has become the standard market measure for portfolios holding various assets. Value at Risk is defined as the potential loss which is encountered in a specified period, for a given level of probability. Hence, VaR is essentially measured by quantiles.

The main objective of the 1988 Basel Accord was to develop a risk-based capital framework that strengthens and stabilizes the banking system. In

1993, the group of 30 set out the following requirements: "Market risk is best measured as 'Value at Risk' using probability analysis based upon a common confidence interval and time horizon" (Gamrowski and Rachev, 1996). In 1996, following this recommendation, the Basel Committee decided to take into account the importance of market risks. The way to define capital requirements thus changed in order to be sufficiently sensitive to the risk profile of each institution. Until now, capital requirements are increasingly based on risk-sensitive measures, which are directly based on VaR for market risk. VaR is a common language to compare the risks of the different markets and can be translated directly into a regulatory capital requirement (Basel Committee, 1996). This measure has also permitted the financial institutions to develop their own internal model. On the other hand, this measure is based on some unrealistic assumptions that are specified later, and the expected shortfall (ES) measure appears preferable (Artzner et al., 1997).

As VaR measures appear mainly as quantiles, the different ways to compute them are specified in univariate and multivariate settings. To do so, financial data sets are used which are characterized by structural behaviors such as the volatility, jumps, explosions, and seasonality that provoke nonstationarity. The question is how to model these features through the distribution function to obtain a robust VaR. New strategies have to be defined and some of them are proposed.

RISK MEASURES

Definition

Traditionally, the risk from an unexpected change in market prices (i.e., the market risk) was based on the mean of the deviation from the mean of the return distribution: the variance. In the case of a combination of assets, risk is computed via the covariance between the pairs of investments. Using this methodology makes it possible to describe the behavior of the returns by the first two moments of the distributions and by the linear correlation coefficient $\rho(X, Y)$ between each pair of returns. This latter measure, which is a measure of dispersion, can be adopted as a measure of risk only if the relevant distribution is symmetrical (and elliptical). On the other hand, the correlation coefficient measures only the codependence between the linear components of two returns X and Y. This very intuitive method is the basis of the Markowitz (1959) portfolio theory, in which the returns on all assets, as well as their dependence structure are assumed to be Gaussian. This approach becomes incorrect as a measure of the dependence between returns, as soon as the cumulative distribution of returns is totally asymmetric, leptokurtic, and contains extreme values.

So far, since the 1996 Basel amendment, the official measure of market risk is the value at risk which is specified below.

Definition 19.1: For a given horizon and a probability level α, $0 < \alpha < 1$, VaR_α is the maximum loss that may be recorded in a specified period, with a level of confidence of $1 - \alpha$. If X is a random return with distribution function F_X, then

$$F_X\ (VaR_\alpha) = \Pr[X \leq VaR_\alpha] = \alpha \tag{19.1}$$

Thus, losses lower than Var_α occur with probability α.

It is now well known that the VaR number can provide an inadequate representation of risk because some assumptions are often unrealistic. The main problem is its incoherent property. Indeed, the VaR measure does not verify the subadditivity property, meaning that the VaR of the sum of two positions X and Y is not less or equal to the sum of the VaR of the individual positions. This situation arises with nonlinear financial instruments such as options. Alternatively, VaR can also indicate what the worst loss incurred in $(1 - \alpha)\%$ of time is, but it says nothing about the loss on the remaining $\alpha\%$. This means that during periods of turmoil, the VaR measure is unable to provide information about the largest losses. This could lead a risk manager to select the worst portfolio, thinking it to be the least risky. Finally, existence of nonstationarities inside most financial data sets makes the computation of VaR often very irrelevant. Another measure of market risk is the ES (also called conditional value at risk). This coherent measure represents the expectation of a loss, given that a threshold is exceeded, for instance VaR_α, and for a probability level α is equal to: $ES_\alpha = E[X | X \leqslant VaR_\alpha]$.

In that latter case, the ES measure is a lower bound of the VaR_α introduced in this section.

This chapter discusses the different problems encountered in computing a robust VaR from sample data sets, given nonstationarity. All these discussions can be extended without difficulty to the ES risk measure.

The Basel amendment has imposed several rules, the most important being the daily calculation of a capital charge to cover the market risk of a portfolio. This calculus is linked to the estimated VaR and has led the financial institutions to develop their internal models. The rule needs to develop methods to estimate the distribution function F_X every day in order to compute Equation (19.1). Now, assuming the invariance of the distribution for any asset during the whole period under study is not always reasonable, since the basic properties of financial assets are not the same in stable periods and during crisis. Guégan (2008) provided a recent discussion of this problem, so that specific strategies need to be developed in the context discussed below. The next section specifies the tools used to compute VaR.

Tools to Compute VaR

Assuming a portfolio is composed of a unique asset, its distribution function can be estimated analytically, using tests (Kolmogorov test, x^2 test), graphical methods (Q-Q plot, etc.) or using a nonparametrical kernel method. When the portfolios are composed of more than one asset, the joint distribution of all assets making up the portfolio needs to be composed as well. In case of independent assets, this last distribution is the product of the assets' distribution. When the assets exhibit dependence between each other, the best way to compute the distribution function of the portfolio is to use the notion of copula, if the aim is to obtain an analytical form of the distribution; if not, nonparametric techniques like the kernel method can be used, which are not studied here. A copula may be defined as follows.

Definition of a Copula

Consider a general random vector $X = (X_1, \ldots, X_m)$ which may represent m components of a portfolio measured at the same time. It is assumed that X has an m-dimensional joint distribution function $F(x_1, \ldots, x_m) = \Pr[X_1 \leq x_1, \ldots, X_m \leq x_m]$. It is further assumed that for all $i \in \{1, \ldots, m\}$, the random variables X_i have continuous margins F_i, such that $F_i(x) = \Pr[X_i \leq x]$. Accordingly, it has been shown by Sklar (1959) that:

Definition 19.2: The joint distribution F of a vector $X = (X_1, \ldots, X_m)$ with continuous margins $F_1, \ldots, F_m$ can be written as

$$F(x_1, \ldots, x_m) = C(F_1(x_1), \ldots, F_m(x_m)), \ (x_1, \ldots, x_m) \in R^m. \qquad (19.2)$$

The unique function C in Equation (19.2) is called the copula associated to the vector X.

The function C is a multivariate distribution function, generally depending on a parameter θ, with uniform margins on [0,1] and it provides a natural link between F and $F_1, \ldots, F_m$. From Equation (19.2), it may be observed that the univariate margins and the dependence structure can be separated, and it makes sense to interpret the function C as the dependence structure of the random vector X.

Estimation of a Copula

To choose a copula associated to a portfolio, we restrict ourselves to the bivariate case. Let $X = (X_1, X_2)$ be a random vector with a bivariate distribution function F, continuous univariate marginal distribution functions F_1 and F_2, and the copula C: $F(x_1, x_2; \theta) = C(F_1(x_1), F_2(x_2); \theta)$.

Here, the copula is parameterized by the vector $\theta \in R^q$, with $q \in N$. $X = \{(X_{i_1}, X_{i_2}), i = 1, 2, \ldots, n\}$ denotes a sample of n observations and the procedure to determine the copula is the following:

1. The marginal distribution functions $F_j, j = 1,2$ are estimated by the rescaled empirical distribution functions:

$$\hat{F}_{nj}(x) = \frac{1}{n+1}\sum_{i=1}^{n} I(X_{ij} \leq x),\ j = 1, 2 \tag{19.3}$$

2. The parameter θ of the copula C is estimated by a maximum log-likelihood method. It is assumed, in this case that the density c of the copula exists, and then $\hat{\theta}$ maximizes the following expression:

$$L(\theta, X) = \sum\nolimits_{i=1}^{n} \log c(\hat{F}_{n1}(X_{i1}), \hat{F}_{n2}(X_{i2}); \theta) \tag{19.4}$$

   where $\hat{F}_{nj}, j = 1,2$ is introduced in Equation (19.3) and: $c(u_1, u_2; \theta) = \dfrac{\partial C(u_1, u_2; \theta)}{\partial u_1 \partial u_2}, (u_1, u_2) \in [0,1]^2$. The estimator $\hat{\theta}$ is known to be consistent and asymptotically normally distributed under regular conditions.
3. In order to apply the maximum likelihood method to estimate θ, we need to work with independent, identically random variables. X_{ij} is known for $i = 1,2,\ldots, n$ and $j = 1,2$ to be not independent time series, so that each time series can start being filtered using an adequate filter (ARMA processes, related GARCH processes, long memory models, Markov switching models, etc.). Then, the previous step is applied to obtain the copula C_θ on the residuals $(\varepsilon_{i1}, \varepsilon_{i2})$ for $i = 1,2,\ldots, n$, associated with each time series. It should be noted that the copula which permits the dependence between (X_1, X_2) and $(\varepsilon_1, \varepsilon_2)$ to be measured will be the same.
4. In order to choose the best copula C_θ, several criteria can be used:
   a. The D^2 criteria. the D^2 distance is associated to the vector (X_1, X_2): $D_C^2 = \sum_{x_1, x_2} \left|C_{\hat{\theta}}(\hat{F}_1(x_1), \hat{F}_2(x_2)) - \hat{F}(x_1, x_2)\right|^2$. Then, the copula $C_{\hat{\theta}}$ for which the smallest D_C^2 is obtained will be chosen as the best

copula. Here $\hat{F}(x_1, x_2)$ is the empirical joint distribution function associated to the vector (X_1, X_2).

b. AIC criteria. When the parameter of the copula by maximizing the log-likelihood Equation (19.4) is obtained, the maximization provides a value of the AIC criteria. This value can be used to discriminate between different copulas. The copula for which this criterion is minimum, is retained.

c. Graphical criteria. From the definition of a copula C, it is known that if U and V are two uniform random variables then the random variables $C(V/U) = \frac{\partial C}{\partial U}(U, V)$ and $C(V/U) = \frac{\partial C}{\partial V}(U, V)$ are also uniformly distributed. This property can be used to estimate the adjustment between the empirical joint distribution and the different copulas, by way of the classical Q-Q plot method. For this, it is necessary to calculate the partial derivatives of the various copulas considered. In the case of Archimedean copulas (see below), only $C_\theta(U/V)$ are investigated, since they are symmetrical.

Classes of Copulas

The previous methods can be adjusted on many copulas. Two classes of copulas are mainly used: elliptical copulas and Archimedean copulas.

1. *Elliptical copulas*. The most commonly used elliptical distributions to model financial assets are the Gaussian and student-t distributions. Their expressions are:

$$C_\Phi(u, v) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \exp\left(-\frac{z_1^2 + z_2^2 - 2z_1z_2}{2(1-\rho^2)}\right) dz_1 dz_2, \text{ and}$$

$$C_T(u, v) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{t_v^{-1}(u)} \int_{-\infty}^{t_v^{-1}(v)} \exp\left(1+\frac{z_1^2 + z_2^2 - 2z_1z_2}{2(1-\rho^2)}\right)^{-\frac{v+2}{2}} dz_1 dz_2$$

and where $\Phi^{-1}(u)$ is the inverted Gaussian probability distribution function and t_v^{-1} is the inverted student-t probability distribution function with degrees of freedom. These copulas are both symmetrical but they have different tail dependence behavior.

2. *Archimedean copulas*. A second class of copulas which is very attractive concerns the Archimedean copulas. To define these copulas, the following class of functions is introduced: $\Phi_\theta = \{\varphi_\theta: [0,1] \to [0, \infty],$ $\varphi_\theta(1) = 0,\ \varphi_\theta'(t) < 0,\ \varphi_\theta''(t) > 0, \theta \in [-1, 1]\}$. Classical functions for $\varphi_\theta \in \Phi_\theta$ are:, $\varphi_\theta(t) = -\log t$, $\varphi_\theta(t) = (1 - t)^\theta$, $\varphi_\theta(t) = t^{-\theta} - 1$ with $\theta > 1$.

It is then easy to show that for all convex functions $\varphi_\theta \in \Phi_\theta$, a function C_θ exists such that

$$C_\theta(u,v) = \varphi_\theta^{-1}(\varphi_\theta(u) + \varphi_\theta(v)), \, if \, \varphi_\theta(u) + \varphi_\theta(v) \leq \varphi_\theta(0) \tag{19.5}$$

and $C_\theta(u, v) = 0$ otherwise. The function $C_\theta(u, v)$ is a symmetric two-dimensional distribution function whose margins are uniform in the interval[0,1]. This is called the Archimedean copula generated by Φ_θ. Amongst the Archimedean distributions, several laws exist: for instance, the Frank law, the Cook and Johnson law, the Gumbel law, the Ali-Mikhail-Haq law (Joe, 1997). The Archimedean property means that it is possible to construct a copula by way of a generator Φ_θ and that a formula exists which permits Kendall's tau to be computed from this operator, say:

$$\tau(C_\theta) = 1 + 4\int_0^1 \frac{\varphi_\theta(t)}{\varphi_\theta^{'}(t)} dt \tag{19.6}$$

M-Variate Archimedean Copulas

It is not easy to work in a multivariate setting using copula. Nevertheless a bivariate family of Archimedean copulas can be extended naturally enough to an m-variate family of Archimedean copulas, $m > 2$, under some constraints (Joe, 1997). First of all, to get this extension, all the bivariate marginal copulas which make up the multivariate copulas have to belong to the given bivariate family. Second, all multivariate marginal copulas up 3 to $m - 1$ may have the same multivariate form. This situation may be illustrated for a trivariate copula. It is assumed three markets denoted (X_1, X_2, X_3) may be observed, and for each there is an n sample. It is assumed that each bivariate margin is characterized by a dependence parameter $\theta_{i,j}$, $(i \neq j \in \{1,2,3\})$. If $\theta_2 > \theta_1$ with $\theta_{1,2} = \theta_2$, and $\theta_{1,3} = \theta_{2,3} = \theta_1$, then a trivariate Archimedean copula has the following form:

$$C_{\theta_1,\theta_2}(u_1,u_2,u_3) = \varphi_{\theta_1}^{-1}(\varphi_{\theta_1} o \varphi_{\theta_2}^{-1}(\varphi_{\theta_2}(u_1) + \varphi_{\theta_2}(u_2)) + \varphi_{\theta_1}(u_3))$$

Empirically, for two random variables X_1 and X_2, $\theta(X_1, X_2)$ denotes the dependence parameter deduced from Kendall's tau, denoted by $\tau(X_1, X_2)$, by means of Equation (19.6). For a random vector $X = (X_1, X_2\, X_3)$ with joint distribution F and continuous marginal distribution functions, F_1, F_2, F_3, Equation (19.2) becomes for all $(x_1, x_2, x_3) \in R^3$: $F(x_1, x_2, x_3) = C_{\theta_1\theta_2}(F_1(x_1), F_2(x_2), F_3(x_3)) = C_{\theta_1}(C_{\theta_2}(F_1(x_1), F_2(x_2)), F_3(x_3))$, if $\theta_1 \leq \theta_2$ with

$\theta_1 = \theta(X_1, X_2) = \theta(X_2, X_3)$ and $\theta = \theta(X_1, X_2)$. When the copulas are retained, in order to choose the trivariate copula $C_{\hat{\theta}_1\hat{\theta}_2}$ that best models the empirical joint distribution $\hat{F}$ of the series (X_1, X_2, X_3), an extension of the numerical criterion D_c^2 can be derived: $D^3{}_C = \sum_{x_1,x_2,x_3} |C_{\hat{\theta}_1}(C_{\hat{\theta}_2}(\hat{F}_1(x_1), \hat{F}_2(x_2), \hat{F}_3(x_3)) - \hat{F}(x_1, x_2, x_3)|^2$

Then, the copula $C_{\hat{\theta}_1\hat{\theta}_2}$ which yields the lowest D_C^3 value, is retained as the best copula.

Copula's Tail Behavior

The copulas are also characterized by their tail behavior, through their upper tail and lower tail coefficients. These coefficients are important for the computation of the VaR measure. Indeed, if we retain a copula whose lower tail behavior is null although there are co-movements inside the markets following negative shocks for instance, then the computation of the VaR will be biased. The tail dependence concept indicates the amount of dependence in the upper-right quadrant tail or in the lower-left quadrant tail of a bivariate distribution. The upper and lower tail dependence parameters of a random vector (X_1, X_2) with copula C can be defined as:

Definition 19.3: If a bivariate copula C is such that $\lim_{u\uparrow 1} \bar{C}(u,u)/(1-u) = \lambda_U$ exists with $\bar{C}(u, u) = 1 - 2u + C(u, u)$, then the copula C has an upper tail dependence if $\lambda_U \in (0, 1]$, and no upper tail dependence if $\lambda_U = 0$. Moreover if a bivariate copula C is such that: $\lim_{u\downarrow 0} C(u,u)/u = \lambda_L$ exists, it may be said that the copula C has lower tail dependence if $\lambda_L \in (0, 1]$, and no lower tail dependence if $\lambda_L = 0$. These tail coefficients can be computed in different ways with respect to the classes of copulas considered here.

1. *Student-t copula*. For this copula, the lower and upper tail dependence coefficients are equal to $\lambda_U = \lambda_L = 2\bar{\tau}_{v+1}\left(\sqrt{\frac{(v+1)(1-\rho)}{1+\rho}}\right)$, where $\bar{\tau}_{v+1}(x) = 1 - t_{v+1}(x)$, $t_{v+1}(x)$, is the student distribution function with $v + 1$ degrees of freedom, and ρ the linear correlation coefficient. Thus, λ_U is an increasing function of ρ. We can also observe that when v tends to infinity, λ_U tends to 0.
2. *Archimedean copulas*. If the generator function is such that $\varphi'(0)$ is finite, the copula C_θ does not have upper tail dependence. If C_θ has upper tail dependence, then $\varphi'(0) = -\infty$, the upper tail dependence parameter is $\lambda_U = 2 - 2\lim_{t\downarrow 0} \varphi_\theta^{-1'}(2t)/\varphi_\theta^{-1'}(t)$, and the lower tail dependence parameter is $\lambda_L = 2\lim_{t\to\infty} (\varphi_\theta^{-1'}(2t)/\varphi_\theta^{-1'}(t))$. If $\varphi^{-1'}$ is known,

then the tail dependence of any Archimedean copula can be estimated.

3. *Survival copulas*. λ_U, λ_L of a survival copula can also be derived from its associated copula (the survival copula of C is given by: $C(S)(u,v) = u + v - 1 + C(1-u, 1-v)$). Thus, $\lambda_U^{C(S)} = \lambda_L^C$ and $\lambda_L^{C(S)} = \lambda_U^C$. This means that if a copula has an upper tail dependence then the associated survival copula has a lower tail dependence and vice-versa. Moreover, a survival copula and its associated copula have the same Kendall's tau.
4. *Linear combinations of copulas*. In order to obtain copulas which have upper and lower tail dependences without being symmetrical, new copulas are constructed as convex linear combinations of two copulas. Hence, for $\omega \in [0, 1]$ and two Archimedean copulas $C_{\theta 1}$ and $C_{\theta 2}$ a new copula C is obtained, which is defined as: $C(u, v) = \omega C_{\theta 1}(u, v) + (1-\omega) C_{\theta 2}(u, v)$. The properties of these copulas can be derived from those of $C_{\theta 1}$ and $C_{\theta 2}$. Suppose that $C_{\theta 1}$ and $C_{\theta 2}$ have, respectively, an upper and a lower tail dependence, then $\lambda_U^C = \omega \lambda_{\theta 1U}^C$ and $\lambda_L^C = (1 - \omega)\lambda_{\theta 2L}^C$.

COMPUTATION OF THE VAR

As soon as the distribution function of a portfolio is known, the VaR is directly computed from this multivariate distribution function or from the associated copula. The VaR measure corresponds to a quantile of a distribution function associated with a small probability. Several strategies have therefore been formulated to compute it:

- The quantile can be estimated directly from the multivariate distribution function using the previous approach.
- The behavior of the distribution function above may be considered at a certain threshold to focus on the tail behavior of the joint distribution function.

VaR as a Quantile of the Whole Distribution Function

Computing the VaR from the whole sample is not simple because of nonstationarities which exist inside the financial data sets. Indeed, most of the financial data sets cover a reasonably long time period, so economic factors may induce some changes in the dependence structure. The basic properties of financial products may change in different periods (stable periods and crisis periods). Therefore, it seems important to detect changes in the dependence structure in order to adapt all the previous tools inside a nonstationary setting. Different strategies can be considered: one is based on the notion of dynamic copula, the other one on the notion of meta-distribution.

Dynamic Copulas

Dynamic copulas have recently been studied in risk management by Dias and Embrechts (2004) investigating the dynamic evolution of copulas' parameters. A change in a copula's family may also be examined, Caillault and Guégan (2005, 2009) and Guégan and Zhang (2008, 2009). Using dynamics inside a copula permits some time-varying evolutions inside the data sets to be modeled. Other nonstationary features can be modelled when a copula's family is changed, change-point techniques can be used to find the change times both for the parameters and the distribution functions. These changes can also be detected using moving windows along the data sets observed. This method makes all types of copula changes observable and makes the change trend clearer. However, how to decide the width of the moving window and the length of the time interval of movement is important and influences the accuracy of the result for the copula change. The following may be carried out:

1. Testing the changes inside the parameters when the copula family remains static.
2. Testing the changes inside the copulas.
   In order to understand clearly copula changes, a series of nested tests based on the conditional pseudo copula can be used (Fermanian, 2005). The different steps are:
   a. A test is first carried out to see whether the copula does indeed change during a specified time period
   b. If the copula seems changeless, the result in the static case continues to hold
   c. Whether the copula's family changes is then detected
   d. If the result of the test shows that the copula's family may not change, then only changes of copula parameters are dealt with
   e. Otherwise, if the result of the test tells us that the copula family may change, then the changes of copula family are examined
3. Change-point tests can be used to detect when there is change inside the parameters. Now, considering that change-point tests have less

power in case of "small" changes, it may be assumed that the parameters change according to a time-varying function of predetermined variables.

4. U-statistics can also be used to detect change point.

Finally, this sequence of steps permits a sequence of copulas to be obtained: it can be a particular copula with evolutionary parameters and/or sequences of different copulas. At each step the VaR measure is computed providing a sequence of VaR measures that evolve over time.

Meta-Distribution

In the previous approach, the complete information set was used in order to try to adapt empirically the evolution of the changes that are observed all along the trajectory. Sometimes the changes are very important corresponding to specific events and need to be clearly identified. Indeed, in finance, structural behaviors such as volatility, jumps, explosions, and seasonality provoking strong nonstationarity may be observed. Alternatively, aggregation or distortion may also be at the origin of nonstationarity. Thus, the assumptions of strong or weak stationarity fail definitively. Indeed, the existence of volatility means that the variance must depend on time. With seasonality, the covariance depends on time producing evidence of nonstationarity. Existence of jumps produces several regimes within data sets. These different regimes can characterize the level of the data or its volatility. Changes in mean or in variance affect the properties of the distribution function characterizing the underlying process. Thus, this distribution function cannot be invariant under time shifts and thus a global stationarity cannot be assumed. Distortion effects correspond to explosions that cannot be removed from any transformation. This behavior can also be viewed as a structural effect. Existence of explosions means that some higher order moments of the distribution function do not exist. Concatenated data sets used to produce specific behavior cannot have the same probability distribution function for the whole period, as soon as there is a juxtaposition of several data sets. Aggregation of independent or weakly dependent random variables is a source of specific features. All of these behaviors may provoke the nonexistence of higher order moments and noninvariance of the distribution function. Using the dynamic copula concept does not make it always possible to detect correctly the time at which changes arise, because the change point method is not always applicable. Thus, it appears necessary to work in another way, in order to integrate correctly the nonstationarities in the computation of VaR.

This research proposes to build homogeneity intervals on which the underlying distribution function is invariant almost up to the four first moments extending the works of Starica and Granger (2005), who propose a test based on the first two order moments of a distribution function. The principle of the test here is the following. It is assumed that a time series $(Y_1, \ldots, Y_n)$ is observed, and a subset $(Y_{m_1}, \ldots, Y_{m_2})$, $\forall m_1, m_2 \in N$ considered, on which the test is then applied, based on the four first moments. For this subset, the test provides a certain value and a confidence interval. Then, rolling windows are used, and another subset $(Y_{m_2+1}, \ldots Y_{m_2+p})$, for some $p \in N$, is considered, on which the test is again applied. This is extended in the same way, sequentially. For each interval, the value of the test is compared with the one obtained with the previous interval, using confidence intervals. Thus, a sequence of homogeneity intervals is constructed, for which invariance is known to exist for the fourth order moments using the following statistic:

$$\hat{T}(n,Y) = \sup_{\lambda \in [-\pi,\pi]} \left| \int_{[-\pi,\pi]^{k-1}} \left(\frac{I_{c_k,Y,n}(z)}{f_{c_k,Y}} - \frac{\hat{c}_k}{c_k} \right) dz \right| \tag{19.7}$$

where $\hat{c}_k$ is an estimate of c_k, the cumulants of order k of the process $(Y_t)_t$, $f_{ck,Y}$ denote the spectral density of cumulants of order k, and $I_{Ck,Y,n}$, its estimate using a sample $(Y_1,\ldots,Y_n)$.

It may be shown, under the null hypothesis that the cumulants of order k are invariant in the subsamples, that Equation (19.7) converges in distribution to $\frac{(2\pi)^{k-1}}{c_k} B(\sum_{j=1}^{k-1} \lambda_j)$ where $B(.)$ is the Brownian bridge, for $k = 3, 4$. The critical values associated with this test can be computed, and will permit confidence intervals to be built. Then, as soon as these homogeneity intervals have been identified, an invariant distribution function can be computed for each interval, and so a sequence of invariant distribution functions can be defined throughout the sample.

For a portfolio which is composed of m assets, this is calculated for each asset. Therefore, a copula linking these different assets using the margins can be estimated for a specific homogeneity interval, for instance on the last one. But other strategies can also be developed. This approach provides two kinds of results:

1. Working with only one asset: this method associating this nonstationary time series with a distribution function obtained through a copula, permits to link a sequence of invariant distribution functions detected all along the trajectory. Indeed, as soon as the

asset $(Y_t)_t$ is characterized by a sequence of r stationary subsamples $Y^{(1)}, \ldots, Y^{(r)}$, each characterized by an invariant distribution function $F_{Y^{(i)}}$, $i = 1, \ldots, r$, the copula linking these margins permits the distribution function of $(Y_t)_t$ to be estimated, and provides an analytical expression of this distribution function that is called a meta-distribution function (this copula can also be characterized by sequence of parameters θ evolving over time): $F(Y^{(1)}, \ldots, Y^{(r)}) = C_\theta(F(Y^{(1)}), \ldots, F(Y^{(r)}))$, where $Y_t^{(i)}$ is the process defined in each subsample, $i = 1, \ldots, r$.

2. Working with m assets: if a portfolio which has m assets $(X_1, X_2, \ldots, X_m)$ is considered next, the same procedures as used before may be used again. This means that for each asset, a sequence of invariant distribution functions $(F_{X_i}^{(1)}, \ldots, F_{X_i}^{(r)})$ is defined, for $i = 1, \ldots, m$ (assuming that r homogeneity intervals are detected for each asset). Then, in order to obtain a robust value of the VaR measure associated with this portfolio, the best copula C_θ is estimated which permits, for instance, the invariant distribution function associated to each market to be linked to the last homogeneity interval. This provides the following multivariate distribution function:

$$F(X_1, \ldots, X_m \,|\, I_r) = C_\theta(F_{X_1}^{(r)}(X_1), \ldots, F_{X_m}^{(r)}(X_m)) \tag{19.8}$$

where I_r is the r-th homogeneity interval. For a given α, the VaR_α is computed as the quantile of the expression (19.8) for this α.

The Pot Method

To compute the VaR associated with a portfolio, it is also possible to consider an approach based on the behavior of the tails of the empirical joint distribution of the assets, using the peak-over-threshold method. This method computes the associated distribution of excesses over a high threshold u, for a random variable X whose distribution function is F, as $F_u(y) = P[X - u \leq y \mid X > u]$

$$F_u(y) = \frac{F(y-u) - F(u)}{1 - F(u)} \tag{19.9}$$

for $0 \leq y < x_+ - u$, where $x_+ \leq \infty$ is the upper endpoint of F. For a large class of distribution functions F (including all the common continuous distribution functions), the excess function F_u converges on a generalized Pareto distribution (GPD), denoted $G_{\xi\beta}$, as the threshold u rises. Furthermore, it may be assumed that the GPD models can approximate the unknown excess distribution function F_u. For a certain threshold u and for some ξ and β (to be estimated):

$$F_u(y) = G_{\xi\beta}(y) \tag{19.10}$$

By setting $x = u + y$ and combining Equations (19.9) and (19.10) the following is obtained: $F(x) = (1 - F(u))\,G_{\xi\beta}(x - u) + F(u)$, $x > u$, which permits an approximation of the tail of the distribution F to be obtained. From an empirical point of view, the following steps are taken:

1. If dealing with a time series with an unknown underlying distribution F, an estimate for $F(u)$ may be constructed, using the N_u data exceeding the fixed threshold u and the parameters ξ and β of the GPD may be estimated. Then the following estimator for the tail distribution is obtained

$$\hat{F}(x) = 1 - \frac{N_u}{N}\left(1 + \hat{\xi}\frac{x-u}{\hat{\beta}}\right)^{-1/\hat{\xi}} \tag{19.11}$$

   which is only valid for $x > u$.
2. Next, using the tail estimator from Equation (19.11) with the estimated values of $\hat{\xi}$ and $\hat{\beta}$, the tail of the empirical marginal distribution $\hat{F}_i$ may be computed for each market X_i for $x_i > u_i$, $i = 1, 2$.
3. In order to find the copula associated with these markets, the empirical values $\hat{\tau}$ of the Kendall's tau between the two markets X_i, $i = 1,2$ may be computed. This $\hat{\tau}$ is computed in the tails (that are defined by the points on which the GPD is adjusted). The parameter $\hat{\theta}$ of the Archimedean copula is computed using the estimation $\hat{\tau}$.
4. Using the empirical distribution $\hat{F}_i$ computed on the tails of each market X_i for $x_i > u_i$, $i = 1,2$, the following relationship may be obtained for the market (X_1, X_2): $\hat{F}(x_i, x_j) = C_{\hat{\theta}}(\hat{F}(x_i), \hat{F}(x_j))$, $x_i > u_i$, $x_j > u_j$, where $C_{\hat{\theta}}$ denotes a copula.
5. Finally the diagnosis proposed in this chapter's second section is employed to retain the best copula.

To use this method, the threshold u needs to be chosen, which can be a limitation on the method. One way to solve this problem is to use the approach developed in the third section of this chapter, Guégan and Ladoucette (2004).

CONCLUSION

In this chapter, we discussed extensively the influence of the presence of nonstationarity within data sets in order to compute for VaR. To detect the existence of local or global stationarity on data sets, a new test based on the empirical moments more than two is presented. Then, the concept of meta-distribution is introduced to characterize the joint distribution function of a nonstationary sample. This approach provides interesting solutions to some current, open questions. It is likely that more robust values for the VaR measure may be obtained using this approach, as well as for ES.

Other points still need to be developed to improve the computations of the risk measures in a nonstationary setting.

- The use of the change point theory has to be developed to get the exact date at which the homogeneity intervals begin.
- The notion of "extreme" copulas need to be investigated in details, in order to build robust estimates for VaR and ES measures.
- The knowledge of the computation of VaR measures in an *m*-dimensional setting is still open. An approach has been proposed by Aas et al. (2009) based on cascades method. Nevertheless the choice of the best copulas inside so many permutations is not clear and the computation of VaR depends strongly of the choice of these permutations. Some new proposals have recently been put forward by Guégan and Maugis (2008), using vines.

REFERENCES

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009) Pair-copula Constructions of Multiple Dependence. *Insurance: Mathematics and Economics*, 44(2): 182–198.

Artzner, P., Delbaen, F., Eber, J., and Heath, D. (1997) Thinking Coherently. *Risk*, 10(11): 68–71.

Basel Committee on Banking Supervision. (1996) Amendment to the Basel Capital Accord to Incorporate Market Risk. Basel, Switzerland: BIS.

Caillault, C. and Guégan, D. (2005) Empirical Estimation of Tail Dependence Using Copulas: Application to Asian Markets. *Quantitative Finance*, 5(5): 489–501.

Caillault, C. and Guégan, D. (2009) Forecasting VaR and Expected Shortfall Using Dynamical Systems: A Risk Management Strategy. *Frontiers in Finance and Economics*, 6(1): 26–50.

Dias, A. and Embrechts, P. (2004) Dynamic Copula Models for Multivariate High-Frequency Data in Finance. In: G. Szegoe (ed.), *Risk Measures for the 21st Century*, Chapter 16, pp. 321–335. Hoboken, NJ: Wiley Finance Series.

Fermanian, J-D. (2005) Goodness of Fit Tests for Copulas. *Journal of Multivariate Analysis*, 95(1): 119–152.

Gamrowski, B. and Rachev, S.T. (1996) Testing the Validity of Value-at-Risk Measures. *Lecture Notes in Statistics*, 114(1): 307–320.

Guégan, D. (2008) Non-Stationarity and Meta-distribution. Working paper 2008–26, Publications du CES Paris 1 Panthéon, Sorbonne, France.

Guégan, D. and Maugis, P.A. (2008) Note on New Prospects on Vines. Working paper 2008–95, Publications du CES Paris 1 Panthéon, Sorbonne, France.

Guégan, D. and Zhang, J. (2008) Pricing Bivariate Option under GARCH-GH Model with Dynamic Copula. *Insurance: Mathematics and Economics*, 42(3): 1095–1103.

Guégan, D. and Zhang, J. (2009) Change Analysis of Dynamic Copula for Measuring Dependence in Multivariate Financial Data. *Quantitative Finance*, 8: 1–10.

Guégan, D. and Ladoucette, S. (2004) Dependence Modelling of the Joint Extremes in a Portfolio Using Archimedean Copulas: Application to MSCI Indices. Note de Recherche IDHE-MORA, 05–2004, France.

Joe, H. (1997) *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

Markovitz, H. (1959) *Portfolio Selection: Efficient and Diversification of Investments*. New York: John Wiley & Sons.

Sklar, A. (1959) Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8(5): 229–231.

Starica, C. and Granger, C.W.J. (2005) Nonstationarities in Stock Returns. *The Review of Economics and Statistics*, 87(3): 3–23.

CHAPTER 20

COPULA-VAR AND COPULA-VAR-GARCH MODELING: Dangers for Value at Risk and Impulse Response Functions

Carluccio Bianchi, Maria Elena De Giuli, Dean Fantazzini, and Mario A. Maggi

ABSTRACT

Copulas have been recently proposed as a statistical tool to build flexible multivariate distributions, since they allow for a rich dependence structure and more flexible marginal distributions that better fit the features of empirical financial and economic data. Our extensive simulation studies investigate how misspecification in the marginals may affect the estimation of the dependence function represented by the copula and the effects of these biases for value-at-risk and impulse response functions analyses. We show that the use of normal marginals, when the true data generating process is leptokurtic, produces biased estimates of the correlations. This may results in more aggressive value-at-risk estimates or in smaller confidence bands when computing impulse response functions.

INTRODUCTION

The need for multivariate models beyond the multivariate normal distribution has been demonstrated by a wide literature showing evidence against the normality assumption for economic variables, starting with Mills (1927) and continuing to the present; see, e.g., Nelsen (2006) and references therein. Using copulas allows the reconsideration of the assumptions made in the analysis of jointly dependent random variables: while there are many univariate distributions that can be used, there are much fewer multivariate distributions and this problem becomes dramatic in very large dimensions; see, e.g., Fantazzini (2009, 2010). Furthermore, even though one can use pseudo-maximum likelihood techniques where the marginal distributions are known, but the joint is not, Greene (2002) showed that in some cases this will result in inconsistent estimates. Similar problems may arise when dealing with quasi-maximum likelihood estimation, too, as recently discussed in details in Prokhorov and Schmidt (2009), who showed that estimates can be inconsistent when independence has been incorrectly assumed.

Indeed, using copulas allows to factor a joint distribution into the marginals and the dependence function represented by the copula. The dependence relationship is modeled by the copula, while position, scaling, and shape (mean, standard deviation, skewness, and kurtosis) are modeled by the marginals. Copulas have been used extensively in biology, finance, and statistics, and we refer the readers to the textbooks by Cherubini, Luciano, and Vecchiato (2004) and Nelsen (2006) for a detailed discussion of copulas and their financial applications. Recently, copulas have also been applied to model operational risks (cf. Chernobai, Rachev, and Fabozzi, 2007; Fantazzini, Dalla Valle, and Giudici, 2008).

The first contribution of this chapter is a Monte Carlo study of the finite sample properties of the marginals and copula estimators, under different hypotheses about the data generating process (DGP). We find the interesting result that, when the true DGP is leptokurtic, using normal marginals without generalized autoregressive conditional heteroskedasticity (GARCH) effects cause the dependence parameters to be biased. Particularly, the correlations show a negative bias that increases in absolute value with the degree of leptokurtosis. Moreover, if small samples are concerned, this bias generally increases, except for strongly leptokurtic data, where the small sample distortion has no clear sign. We also remark that, when the DGP has GARCH effects, the small sample biases of the variance parameters are very large, even though they vanish as the sample size increases. The same behavior is shown by the degrees of freedom of the Student's t.

The second contribution is the analysis, applying a Monte Carlo study, of misspecified margins on the estimation of multivariate value at risk (VaR) for equally weighted portfolios. When there are no GARCH effects, the GARCH models (both with normally and Student's t distributed errors) deliver very conservative VaR estimates; while if there are GARCH effects with normally distributed errors, the use of normal-GARCH marginals underestimates the true VaR. In both cases, the biases are due to poor marginals estimates in small/medium datasets. When the true DGP is leptokurtic, the models employing normal marginals (with or without GARCH effects) underestimate the true VaR, particularly in the case of extreme quantiles. However, in this case, the negative biases are mainly due to the negative correlations biases.

The third contribution of the chapter is the proposal of a general methodology to compute impulse response functions (IRFs) with a non-normal joint multivariate distribution. We show that using normal marginals, when data are leptokurtic, produces confidence bands narrower than they should be, while a shock may be considered statistically more persistent than it actually is. Moreover, negative biases in the marginals and in the dependence structure can underestimate the true IRFs. That is why some caution should be taken when using IRFs estimated with a multivariate normal vector autoregression (VAR) model implemented in standard software packages. In this perspective, we propose here a general methodology to compute IRFs for a general copula-VAR-GARCH model, where we accommodate for both non-normality and nonlinearities, and we allow for the imposition of structural restrictions. Furthermore, we suggest the use of bootstrapped standard errors to gauge the precision of the estimated IRFs.

The rest of the chapter is organized as follows. This chapter's second section presents the copula-VAR and copula-VAR-GARCH models, while in the third section we perform simulation studies in order to assess the finite sample properties of these models under different DGPs. We investigate the effects of model misspecifications on VaR in the fourth section of this chapter, while we examine the effects of misspecifications on IRFs analysis in the fifth section, where we also describe a procedure to construct generalized IRFs together with bootstrapped standard errors. The final section concludes.

COPULA-VAR-GARCH MODELING

We present here a copula-VAR model, where there are n endogenous variables $x_{i,t}$ explained by an intercept μ_i, autoregressive terms of order p, and an heteroskedastic error term $\sqrt{h_{i,t}}\eta_{i,t}$,

$$
\begin{aligned}
x_{i,t} &= \mu_1 + \sum_{i=1}^{n}\sum_{l=1}^{p}\phi_{1,i,l}x_{i,t-l} + \sqrt{h_{1,t}}\eta_{1,t} \\
\vdots \qquad & \qquad\quad \vdots \qquad\qquad\qquad \vdots \\
x_{n,t} &= \mu_n + \sum_{i=1}^{n}\sum_{l=1}^{p}\phi_{n,i,l}x_{i,t-l} + \sqrt{h_{n,t}}\eta_{n,t}
\end{aligned}
\tag{20.1}
$$

We suppose the standardized innovations $\eta_{i,t}$ to have mean zero and variance one, while $\sqrt{h_{i,t}}$ can be time varying as in GARCH models or constant, depending on the data at hand

$$\begin{aligned} h_{1,t} &= \omega_1 + \alpha_1(\eta_{1,t-1}\sqrt{h_{1,t-1}})^2 + \beta_1 h_{1,t-1} \\ &\vdots \qquad\qquad \vdots \qquad\qquad\qquad \vdots \\ h_{n,t} &= \omega_n + \alpha_n(\eta_{n,t-1}\sqrt{h_{n,t-1}})^2 + \beta_n h_{n,t-1} \end{aligned} \tag{20.2}$$

The innovations $\eta_{i,t}$ have a multivariate joint distribution $H_t\,(\eta_{1,t}, \ldots, \eta_{n,t};\, \theta)$ with the parameters vector θ, which can be rewritten by using Sklar's theorem (1959):

$$(\eta_{1,t}, \ldots, \eta_{n,t}) \sim H_t(\eta_{1,t}, \ldots, \eta_{n,t}; \theta) = C_t\,[F_{1,t}(\eta_{1,t}; \alpha_n), \ldots, F_{n,t}(\eta_{\mathrm{n},t}; \alpha_n); \gamma] \tag{20.3}$$

Particularly, the joint distribution H_t is the copula $C_t[\cdot;\, \gamma]$ of the cumulative distribution functions (cdf) of the innovations marginals $F_{1,t}(\eta_{1,t};\, \alpha_1), \ldots,$ $F_{n,t}(\eta_{1,t};\, \alpha_n)$, where γ, $\alpha_1, \ldots, \alpha_n$ are the copula and the marginals parameters, respectively.

Thanks to Sklar's theorem, using copulas has the important benefit of linking together two or more marginals distributions, not necessarily identical, to form a well-defined multivariate joint distribution, which is known in statistics as meta-distribution.[1]

Sklar's theorem can be used to find the analytical density functions for many important copulas. Among these, the normal-copula density function can be computed as follows:

$$\begin{aligned} c^{Normal}(\Phi_1(x_1), \ldots, \Phi_n(x_n); \theta_0) &= \frac{f^{Normal}(x_1, \ldots, x_n)}{\prod_{i=1}^{n} f_i^{Normal}(x_i)} \\ &= \frac{\dfrac{1}{(2\pi)^{n/2}\,|\Sigma|^{1/2}} \exp\left(-\dfrac{1}{2}x'\Sigma^{-1}x\right)}{\prod_{i=1}^{n} \dfrac{1}{\sqrt{2\pi}} \exp\left(-\dfrac{1}{2}x_i^2\right)} = \\ &= \frac{1}{|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\zeta'(\Sigma^{-1} - I)\zeta\right) \end{aligned} \tag{20.4}$$

where $\zeta = (\Phi^{-1}(u_1), \ldots, \Phi^{-1}(\mathrm{u_n}))'$ is the vector of univariate Gaussian inverse distribution functions, $u_i = \Phi(x_i)$, while Σ is the correlation matrix. We remark that this copula belongs to the class of elliptical copulas. As an alternative to elliptical copulas we could use Archimedean copulas, but they present the serious limitation of modeling only positive dependence (or only partial negative dependence), while their multivariate extensions involve strict restrictions on dependence parameters. As a consequence we do not consider them here.

We can follow a procedure similar to Equation (20.4) to derive the copula of the multivariate Student's t distribution (the *t-copula*). Moreover, the copula parameters can be made time varying, too. However, recent literature (see Chen, Fan, and Patton, 2004) has shown that a simple time constant normal copula is, in most cases, sufficient to describe the dependence structure of daily financial data. Instead, when the number of variables is larger than 20, more complicated copulas than the normal are needed. Similar evidence has been found recently by Fantazzini et al. (2008) who analyzed monthly operational risk data. Actually, macroeconomic analysis usually works with a small number of endogenous variables sampled at a monthly or lower frequency, which are known to have a simpler dependence structure than daily financial data. For this reason, we examine here only a constant normal copula $C_t^{\text{Normal}} = C^{\text{Normal}}$.

SIMULATION STUDIES

This section discusses the results of the simulation studies concerning a trivariate copula-VAR(1)-GARCH(1,1) model specified as follows:

$$Y_t = \mu + \Phi_Y^{(1)} Y_{t-1} + h_t \eta_t \tag{20.6}$$

where the matrix h is diagonal and contains the square roots of the variances: $h_{i,t} = \omega_i + \alpha_i(\eta_{i,t-1}\sqrt{h_{i,t-1}})^2 + \beta_i h_{i,t-1}$, $i = 1, 2, 3$. The innovations η_t are modeled by a Gaussian copula, with Gaussian or Student's t marginals, and the correlation matrix Σ:

$$(\eta_{1,t},\ldots,\eta_{n,t}) \sim C^{\text{Normal}}[F_{1,t}(\eta_{1,t};\alpha_1),\ldots,F_{n,t}(\eta_{1,t};\alpha_n);\Sigma] \tag{20.7}$$

We consider the following possible DGPs:

- We consider different marginals specifications for the true DGP in Equation (20.6):
    - Student's t with 3 degrees of freedom
    - Student's t with 10 degrees of freedom
    - Normal distribution
- Distributions are used with and without GARCH volatility

- Time series of 500, 2,000, and 20,000 observations are chosen for our simulations

This allows us to study the two-step inference functions from marginals (IFMs, e.g., see Cherubini, Luciano, and Vecchiato, 2004) estimation method also in the case of short time series. For this reason, the Gaussian specification is also estimated by a two-step procedure, rather than by ordinary least squares.

For each marginals and variance specification, we generate 1,000 Monte Carlo samples and estimate the following models: **Normal**, **Normal-GARCH**, and **Student's t-GARCH**.

We do not report the numerical results, which are available upon request. We focus our discussion on correlation and variance parameters because, with regard to the coefficients of the autoregressive terms, we remark that the estimated parameters are very similar across the various specifications and their biases are very small. Besides, the variance and correlations are the most important parameters with respect to various economic and financial applications, such as VaR estimates. Moreover, as Fantazzini (2008) clearly shows in a wide empirical exercise, "the AR specification of the mean is not relevant in all cases."

The simulation studies performed yield some interesting results:

- *Correlation parameters* $\rho_{i,j}$: The normal innovations distribution generally underestimates the absolute value of correlations. This bias increases with the leptokurtosis of the DGP. For example, for very fat-tailed cases, this negative bias is approximately equal to 14 percent of the true value. This bias stabilizes around these percentages also with a larger sample of 20,000 observations. As expected, the bias decreases when the degrees of freedom of the Student's t increase and when the volatility is constant. When the time series becomes shorter, the correlation biases may either increase or decrease, but they generally remain negative. Obviously, the well specified Student's t model estimates correlations with greater precision, instead. Indeed, the Student's t is very precise when the true DGP is strongly leptokurtic, while it behaves like the normal when the degrees of freedom increase.
- *Variance parameters*: As expected, when there are no GARCH effects, the model employing normal marginals works best, while the Student's t model is not efficient. However, it is interesting to note that *if* there are GARCH effects with normally distributed errors, the use of normal-GARCH marginals underestimates the variance when small samples are used. In this case, a copula-Student's t-GARCH

model seems a better choice. Furthermore, the Student's t is more precise when the true DGP is strongly leptokurtic ($\upsilon = 3$), while the normal again underestimates the variance parameters.

IMPLICATIONS FOR VALUE-AT-RISK ESTIMATION

Value at risk is a concept developed in the field of risk management that is defined as the maximum amount of money that one could expect to lose with a given probability over a specific period of time. While the VaR approach is widely used, its notion is nonetheless controversial, primarily as a consequence of the diverse methods used in computing VaR, producing widely divergent results. If the cdf of the joint distribution is known, then the VaR is simply its p-th quantile times the value of the financial position; however, the cdf is not known in practice and must be estimated. Jorion (2007) provided an introduction to VaR as well as a discussion of its estimation, while the www.gloriamundi.org website comprehensively cites the VaR literature as well as providing other VaR resources.

We explore here the potential impact of misspecified marginals on the estimation of multivariate VaR for equally weighted portfolios, by using the same DGPs discussed at the beginning of this chapter's third section. For the sake of simplicity, we suppose to invest an amount $M_i = 1$, $i = 1, \ldots, n$ in every asset (where $n = 3$ in our simulation studies).

We consider eight different quantiles to better highlight the overall effects of the estimated copula parameters on the joint distribution of the losses: 0.25, 0.50, 1.00, 5.00, 95.00, 99.00, 99.50, 99.75 percent, that is we consider both the "loss tail" and the "win tail." Tables 20.1 to 20.6 report

Table 20.1 VaR Estimates when the True Marginals Are Normal without/with GARCH Effects; Employed Marginals for Estimation: Normal *without* GARCH Effects

True Model: Normal		Estimated Marginals: Normal			True Model: Normal GARCH		Estimated Marginals: Normal		
	VaR		Bias (%)	RMSE		VaR		Bias (%)	RMSE
0.25%	0.0027	0.0027	−0.49	0.0000	**0.25%**	0.0161	0.0158	−1.74	0.0001
0.50%	0.0018	0.0018	0.83	0.0000	**0.50%**	0.0141	0.0138	−1.62	0.0001
1.00%	0.0008	0.0008	0.64	0.0000	**1.00%**	0.0119	0.0117	−1.27	0.0000
5.00%	−0.0018	−0.0018	−0.32	0.0000	**5.00%**	0.0060	0.0059	−2.25	0.0000
99.75%	−0.0191	−0.0191	−0.36	0.0000	**99.75%**	−0.0327	−0.0325	−0.69	0.0000
99.50%	−0.0182	−0.0182	−0.15	0.0000	**99.50%**	−0.0306	−0.0305	−0.56	0.0000
99.00%	−0.0173	−0.0172	−0.04	0.0000	**99.00%**	−0.0285	−0.0283	−0.50	0.0000
95.00%	−0.0146	−0.0146	0.09	0.0000	**95.00%**	−0.0226	−0.0224	−0.64	0.0000

Table 20.2 VaR Estimates when the True Marginals Are Normal without/ with GARCH Effects; Employed Marginals for Estimation: Normal *with* GARCH Effects

True Model: Normal		Estimated Marginals: Normal			True Model: Normal GARCH		Estimated Marginals: Normal		
	VaR		Bias (%)	RMSE		VaR		Bias (%)	RMSE
0.25%	0.0027	0.0147	441.29	0.1398	**0.25%**	0.0161	0.0077	−52.20	0.0682
0.50%	0.0018	0.0128	613.33	0.1147	**0.50%**	0.0141	0.0064	−54.58	0.0580
1.00%	0.0008	0.0108	1213.46	0.0989	**1.00%**	0.0119	0.0050	−57.96	0.0458
5.00%	−0.0018	0.0052	−385.48	0.0492	**5.00%**	0.0060	0.0011	−81.62	0.0240
99.75%	−0.0191	−0.0312	63.27	0.1412	**99.75%**	−0.0327	−0.0242	−26.02	0.0705
99.50%	−0.0182	−0.0294	61.42	0.1238	**99.50%**	−0.0306	−0.0229	−25.41	0.0599
99.00%	−0.0173	−0.0274	58.61	0.1010	**99.00%**	−0.0285	−0.0214	−24.77	0.0482
95.00%	−0.0146	−0.0217	48.93	0.0488	**95.00%**	−0.0226	−0.0175	−22.19	0.0240

Table 20.3 VaR Estimates when the True Marginals Are Normal without/ with GARCH Effects; Employed Marginals for Estimation: Student's t *with* GARCH Effects

True Model: Normal		Estimated Marginals: Student's t-GARCH			True Model: Normal-GARCH		Estimated Marginals: Student's t-GARCH		
	VaR		Bias (%)	RMSE		VaR		Bias (%)	RMSE
0.25%	0.0027	0.0320	1076.44	0.8395	**0.25%**	0.0161	0.0158	−1.89	0.0001
0.50%	0.0018	0.0286	1499.18	0.6949	**0.50%**	0.0141	0.0138	−1.73	0.0001
1.00%	0.0008	0.0251	2955.51	0.5619	**1.00%**	0.0119	0.0117	−1.37	0.0000
5.00%	−0.0018	0.0153	−937.55	0.2846	**5.00%**	0.0060	0.0059	−2.22	0.0000
99.75%	−0.0191	−0.0490	156.01	0.8608	**99.75%**	−0.0327	−0.0324	−1.08	0.0001
99.50%	−0.0182	−0.0457	150.55	0.7390	**99.50%**	−0.0306	−0.0304	−0.78	0.0001
99.00%	−0.0173	−0.0420	143.48	0.6037	**99.00%**	−0.0285	−0.0283	−0.80	0.0001
95.00%	−0.0146	−0.0321	119.92	0.2930	**95.00%**	−0.0226	−0.0224	−0.69	0.0000

Table 20.4 VaR Estimates when the True Marginals Are Student's t ($\upsilon = 3, \upsilon = 10$) with GARCH; Employed Marginals for Estimation: Normal *without* GARCH Effects

True Model: Student's t (υ = 3) GARCH		Estimated Marginals: Normal			True Model: Student's t (υ = 10) GARCH		Estimated Marginals: Normal		
	VaR		Bias (%)	RMSE		VaR		Bias (%)	RMSE
0.25%	0.0263	0.0164	−37.60	0.0970	**0.25%**	0.0180	0.0159	−11.90	0.0045
0.50%	0.0195	0.0144	−25.86	0.0252	**0.50%**	0.0153	0.0138	−9.74	0.0021
1.00%	0.0139	0.0123	−11.48	0.0024	**1.00%**	0.0127	0.0117	−7.38	0.0008
5.00%	0.0042	0.0063	51.97	0.0046	**5.00%**	0.0059	0.0059	−1.04	0.0000
99.75%	−0.0434	−0.0333	−23.40	0.1024	**99.75%**	−0.0347	−0.0325	−6.17	0.0045
99.50%	−0.0365	−0.0311	−14.87	0.0292	**99.50%**	−0.0320	−0.0305	−4.80	0.0023
99.00%	−0.0307	−0.0289	−6.03	0.0033	**99.00%**	−0.0293	−0.0283	−3.25	0.0009
95.00%	−0.0207	−0.0228	10.08	0.0042	**95.00%**	−0.0225	−0.0224	−0.25	0.0000

Table 20.5 VaR Estimates when the True Marginals Are Student's t ($\upsilon = 3, \upsilon = 10$) with GARCH; Employed Marginals for Estimation: Normal *with* GARCH Effects

True Model: Student's t (υ = 3) GARCH		Estimated Marginals: Normal GARCH			True Model: Student's t (υ = 10) GARCH		Estimated Marginals: Normal GARCH		
	VaR		Bias (%)	RMSE		VaR		Bias (%)	RMSE
0.25%	0.0263	0.0156	−40.60	0.1097	**0.25%**	0.0180	0.0150	−16.76	0.0087
0.50%	0.0195	0.0137	−29.82	0.0329	**0.50%**	0.0153	0.0131	−14.33	0.0048
1.00%	0.0139	0.0115	−16.93	0.0054	**1.00%**	0.0127	0.0111	−12.58	0.0025
5.00%	0.0042	0.0058	39.32	0.0026	**5.00%**	0.0059	0.0054	−8.24	0.0002
99.75%	−0.0434	−0.0324	−25.53	0.1173	**99.75%**	−0.0347	−0.0317	−8.71	0.0087
99.50%	−0.0365	−0.0303	−17.04	0.0375	**99.50%**	−0.0320	−0.0297	−7.16	0.0052
99.00%	−0.0307	−0.0282	−8.42	0.0064	**99.00%**	−0.0293	−0.0276	−5.58	0.0026
95.00%	−0.0207	−0.0223	7.86	0.0026	**95.00%**	−0.0225	−0.0220	−2.32	0.0003

Table 20.6 VaR Estimates when the True Marginals Are Student's t ($\upsilon = 3, \upsilon = 10$) with GARCH; Employed Marginals for Estimation: Student's t *with* GARCH Effects

True Model: Student's t ($\upsilon = 3$) GARCH		Estimated Marginals: Student's t-GARCH			True Model: Student's t ($\upsilon = 10$) GARCH		Estimated Marginals: Student's t-GARCH		
	VaR		Bias (%)	RMSE		VaR		Bias (%)	RMSE
0.25%	0.0263	0.0218	−17.24	0.0204	**0.25%**	0.0180	0.0165	−8.44	0.0022
0.50%	0.0195	0.0176	−9.64	0.0034	**0.50%**	0.0153	0.0143	−6.95	0.0011
1.00%	0.0139	0.0139	−0.10	0.0000	**1.00%**	0.0127	0.0119	−5.70	0.0005
5.00%	0.0042	0.0057	37.80	0.0024	**5.00%**	0.0059	0.0058	−2.22	0.0000
99.75%	−0.0434	−0.0384	−11.53	0.0249	**99.75%**	−0.0347	−0.0332	−4.15	0.0020
99.50%	−0.0365	−0.0342	−6.38	0.0053	**99.50%**	−0.0320	−0.0309	−3.47	0.0012
99.00%	−0.0307	−0.0304	−1.05	0.0001	**99.00%**	−0.0293	−0.0286	−2.37	0.0005
95.00%	−0.0207	−0.0222	7.20	0.0022	**95.00%**	−0.0225	−0.0224	−0.58	0.0000

the true VaR, the mean across simulations, the bias in percentage, and the root mean square error. For the sake of interest and space, we report only the results for $n = 500$.

In general, the estimated quantiles show a very large degree of under/overestimation, which depends heavily both on the sample dimension and the underlying joint distribution. The major findings are reported below:

- *When there are no GARCH effects, the GARCH models (both with normal and Student's t errors) deliver very conservative VaR estimates.* This is due to poor estimates of the marginals variance and confirms previous evidence found by Hwang and Valls Pereira (2006) concerning small sample properties of GARCH models.
- *If there are GARCH effects with normally distributed errors, the use of normal-GARCH marginals underestimates the true VaR.* This is again due to poor estimates of the marginals variance.
- *When the true DGP is leptokurtic* ($\upsilon = 3$ *or* $\upsilon = 10$), *the normal models (with or without GARCH effects) underestimate the true VaR*, particularly in the case of extreme quantiles. In this case, the negative biases are not due to the marginals problems (since they are more or less correctly estimated), but to the negative correlations biases. As expected, the negative biases decrease when the degrees of freedom υ increase from 3 to 10.

The Monte Carlo evidence highlights that one should first check for GARCH effects, and then for the type of marginals distribution. If the null hypothesis of homoskedasticity is accepted, normal or Student's t marginals with constant volatility are a much more efficient choice than trying an (unrestricted) model with time-varying volatility, particularly when dealing with small/medium datasets with a number of observations less than 1,000.

IMPLICATIONS FOR IMPULSE RESPONSE FUNCTIONS ANALYSIS

An IRF traces the effect of a one-time shock to one of the innovations in current and future values of the endogenous variables. If the innovations η_t are contemporaneously uncorrelated, the interpretation of the IRF is straightforward: the i-th innovation $\eta_{i.t}$ is simply a shock to the i-th endogenous variable $x_{i.t}$. Innovations, however, are usually correlated, and may be viewed as having a common component which cannot be associated with a specific variable: hence a shock to variable i cannot be identified on that variable only. The identification problem is the subject of the structural VAR literature (e.g., see Sims, 1980, Blanchard and Quah, 1989, and Amisano and Giannini, 1997). For example, it is possible to use the inverse of the Cholesky factor of the normal copula correlation matrix $\mathbf{\Sigma}$ to orthogonalize the impulses and transform the innovations so that they become uncorrelated. However, other identification schemes can be imposed.

Following Fang, Kotz, and Ng (1987), we know that given a matrix $\boldsymbol{A}$ such as $\mathbf{\Sigma} = \boldsymbol{AA}'$, and assuming a set of i.i.d standard normal random variables $(Z_1,\dots, Z_n)'$, the random vector $\boldsymbol{AZ}$ is multi-normally distributed with mean zero and covariance matrix $\mathbf{\Sigma}$. Estimates of the impulse response vectors can then be derived by combining the parameter estimates of the copula VAR, together with the structural matrix $\boldsymbol{A}$ derived by imposing restrictions on the estimated copula correlation matrix $\mathbf{\Sigma}$. The full procedure is the following one:

1. We consider an initial disturbance vector $\boldsymbol{u_0} = (0\dots,0, 1, 0,\dots,0)'$, i.e., a vector of all zeroes apart from the k-th element, and we premultiply this vector by the estimate of the structural matrix $\boldsymbol{A}$, to get the impact response $\boldsymbol{S} = \boldsymbol{A}\ \boldsymbol{u_0}$, as if the innovations η_t were multinormally distributed with mean zero and covariance matrix $\mathbf{\Sigma}$ (which coincides with the correlation matrix).
2. Since we work with the normal copula, which is the copula of the multivariate normal distribution, and recalling (20.4), we have to

transform the vector $\boldsymbol{S}$ by using the standard normal cdf, i.e., $Y_i = \Phi(S_i)$, $i = 1, \ldots, n$.

3. In order to get the standardized innovations $\eta_{i.t}$, we have to use the inverse of the different marginals cdfs F_i, $i = 1, \ldots, n$, that is, $\boldsymbol{Z} = (Z_1, \ldots, Z_n)' = (F_1^{-1}(Y_1), \ldots, F_n^{-1}(Y_n))$, with F_1^{-1} not necessarily identical.
4. Finally, we have to rescale these standardized innovations by using the square roots of the variances $\sqrt{h_{i,t}}$ to get the response vector $\theta_{t,k,0} = Z_1\sqrt{h_{i,t}}, \ldots, Z_n\sqrt{h_{n,t}}$, where $\sqrt{h_{i,t}}$ may be constant or time varying.
5. The remaining response vectors $\theta_{k,h}$ at time $t + h$ can be estimated by solving forwards for the endogenous variables in Equations (20.1)–(20.3).

We want to remark that if we use normal marginals $F_i = \Phi_i$ together with a normal copula, we are back to the joint normal distribution and steps *2* and *3* cancel out. Moreover, if we also assume constant variances, $\sqrt{h_{i,t}} = \omega_i$, we get the IRFs for a standard VAR model. In this sense, our approach is a general one which nests many standard cases.

In order to gauge the precision of the estimated IRFs, we can employ bootstrapping techniques. This procedure involves creating artificial histories for the endogenous variables of the model and then submitting these histories to the same estimation procedure as real data. The artificial histories are created by replacing the parameters in the model with their estimated values, drawing residuals whose moments are determined by the estimated copula-marginals functions and then calculating the endogenous variables. Since the artificial histories are finite samples, their estimates will not coincide exactly with those from the original data; by creating a large number of artificial histories, we can then make a bootstrapped approximation to the distribution of the estimated parameters. This distribution forms the basis for adding confidence intervals bands to the central estimate of the IRFs. The full procedure is reported below.

1. *Draw an* $n \times T$ vector of standardized innovations η_i for $i = 1, \ldots, n$, from the joint normal copula-marginals density $C^{\text{Normal}}[F_{1,t}(\eta_{1,t}; \hat{\alpha}_1), \ldots, F_{n,t}(\eta_{1,t}; \hat{\alpha}_n); \hat{\Sigma}]$.
2. *Create an artificial history for the endogenous variables*: replace all parameters in Equations (20.1)–(20.3) by their estimated values, together with the standardized innovations $\eta_{i,t}$ drawn in the previous step, which have to be rescaled by the square roots of the variances

$\sqrt{h_{i,t}}$. Equations (20.1)−(20.3) can then be applied recursively to get an artificial sample.

3. *Estimate a copula-VAR model*, using the data from the artificial history. The estimation gives bootstrapped estimates of the marginals parameters $\{\tilde{\mu}_i, \tilde{\phi}_{i,j,l}, \tilde{\omega}_i, \tilde{\alpha}_i, \tilde{\beta}_i\}$ for $i =, \ldots, n, j = 1, \ldots, n$ and $l = 1, \ldots, p$ and of the copula correlation matrix $\tilde{\mathbf{\Sigma}}$.
4. *Estimate the matrix* $\tilde{\mathbf{A}}$, by imposing identification restrictions on the estimated correlation matrix $\tilde{\Sigma}$, where $\tilde{\mathbf{A}}\tilde{\mathbf{A}}' = \tilde{\mathbf{\Sigma}}$.
5. *Calculate the bootstrapped estimates of the IRFs*. Use the new parameters $\{\tilde{\mu}_i, \tilde{\phi}_{i,j,l}, \tilde{\omega}_i, \tilde{\alpha}_i, \tilde{\beta}_i\}$ for $i =, \ldots, n, j = 1, \ldots, n$ and $l = 1, \ldots, p$ together with $\tilde{\mathbf{A}}$, to get a bootstrapped estimate of the IRFs $\tilde{\theta}_{t,k,0}, \ldots, \tilde{\theta}_{t,k,h}$.
6. *Repeat the above five steps for a large number of times*, in order to get a numerical approximation of the distribution of the original estimates $\tilde{\theta}_{t,k,0}, \ldots, \sim \hat{\theta}_{t,k,h}$. This distribution forms the basis for adding confidence intervals bands to the central estimate of the IRFs.

We present the IRFs of the first endogenous variable to its own noise impulse, where we use the DGPs and the simulated parameters employed in the section Simulation Studies, for the first 10 lags together with the 90 percent bootstrapped confidence intervals. We report only the results for $n = 500$ for sake of interest and space, since macroeconomic empirical analyses deal with short time series.

Table 20.7 IRFs when the True Marginals Are Normal without/with GARCH Effects; Employed Marginals for Estimation: Normal *without* GARCH Effects (n = 500)

True Model: Normal		Estimated Marginals: Normal			True Model: Normal-GARCH		Estimated Marginals: Normal		
IRFs	True Values	Lower C. B.	IRF	Upper C. B.	IRFs	True Values	Lower C. B.	IRF	Lower C. B.
IRF _1	0.2240	0.2114	0.2331	0.2362	*IRF _1*	0.5000	0.3809	0.4817	0.6593
IRF _2	0.1120	0.0888	0.1116	0.1273	*IRF _2*	0.2500	0.1693	0.2385	0.3500
IRF _3	0.0112	−0.0450	0.0069	0.0226	*IRF _3*	0.0250	−0.0358	0.0231	0.0861
IRF _4	0.0168	−0.0111	0.0129	0.0277	*IRF _4*	0.0375	−0.0190	0.0339	0.0896
IRF _5	0.0118	−0.0160	0.0092	0.0282	*IRF _5*	0.0263	−0.0244	0.0224	0.0668
IRF _6	−0.0014	−0.0211	−0.0026	0.0135	*IRF _6*	−0.0031	−0.0457	−0.0054	0.0353
IRF _7	−0.0011	−0.0084	−0.0009	0.0066	*IRF _7*	−0.0024	−0.0187	−0.0019	0.0164
IRF _8	0.0007	−0.0022	0.0014	0.0077	*IRF _8*	0.0015	−0.0075	0.0017	0.0128
IRF _9	−0.0003	−0.0034	0.0005	0.0044	*IRF _9*	−0.0006	−0.0082	0.0002	0.0093
IRF _10	−0.0004	−0.0006	0.0004	0.0026	*IRF _10*	−0.0008	−0.0026	0.0011	0.0073

Table 20.8 IRFs when the True Marginals Are Normal without/with GARCH Effects; Employed Marginals for Estimation: Normal *with* GARCH Effects

True Model: Normal		Estimated Marginals: Normal-GARCH			True Model: Normal-GARCH		Estimated Marginals: Normal-GARCH		
IRFs	True Values	Lower C. B.	IRF	Upper C. B.	IRFs	True Values	Lower C. B.	IRF	Lower C. B.
IRF _1	0.2240	0.2117	0.2350	0.2364	*IRF_1*	0.5000	0.3815	0.4491	0.4605
IRF _2	0.1120	0.0923	0.1124	0.1275	*IRF_2*	0.2500	0.1269	0.1690	0.2660
IRF _3	0.0112	−0.0462	0.0068	0.0226	*IRF_3*	0.0250	−0.0862	−0.0203	−0.0129
IRF _4	0.0168	−0.0115	0.0130	0.0277	*IRF_4*	0.0375	−0.0611	−0.0135	−0.0082
IRF _5	0.0118	−0.0160	0.0092	0.0282	*IRF_5*	0.0263	−0.0193	0.0242	0.0664
IRF _6	−0.0014	−0.0213	−0.0026	0.0135	*IRF_6*	−0.0031	−0.0411	−0.0048	0.0313
IRF _7	−0.0011	−0.0084	−0.0009	0.0066	*IRF_7*	−0.0024	−0.0162	−0.0025	0.0123
IRF _8	0.0007	−0.0022	0.0014	0.0077	*IRF_8*	0.0015	−0.0063	0.0018	0.0113
IRF _9	−0.0003	−0.0035	0.0005	0.0044	*IRF_9*	−0.0006	−0.0071	0.0001	0.0085
IRF _10	−0.0004	−0.0006	0.0004	0.0026	*IRF_10*	−0.0008	−0.0026	0.0011	0.0073

Table 20.9 IRFs when the True Marginals Are Normal without/with GARCH Effects; Employed Marginals for Estimation: Student's t *with* GARCH Effects

True Model: Normal		Estimated Marginals: Student's t-GARCH			True Model: Normal-GARCH		Estimated Marginals: Student's t-GARCH		
IRFs	True Values	Lower C. B.	IRF	Upper C. B.	IRFs	True Values	Lower C. B.	IRF	Lower C. B.
IRF_1	0.2240	0.2100	0.2350	0.2848	*IRF_1*	0.5000	0.4647	0.5047	0.5484
IRF_2	0.1120	0.0819	0.1124	0.1285	*IRF_2*	0.2500	0.2085	0.2516	0.3010
IRF_3	0.0112	−0.0825	0.0068	0.0243	*IRF_3*	0.0250	−0.0224	0.0234	0.0680
IRF_4	0.0168	−0.0254	0.0130	0.0289	*IRF_4*	0.0375	−0.0088	0.0352	0.0785
IRF_5	0.0118	−0.0350	0.0092	0.0401	*IRF_5*	0.0263	−0.0198	0.0245	0.0681
IRF_6	−0.0014	−0.0242	−0.0026	0.0138	*IRF_6*	−0.0031	−0.0418	−0.0048	0.0318
IRF_7	−0.0011	−0.0127	−0.0009	0.0067	*IRF_7*	−0.0024	−0.0165	−0.0026	0.0126
IRF_8	0.0007	−0.0024	0.0014	0.0112	*IRF_8*	0.0015	−0.0063	0.0019	0.0115
IRF_9	−0.0003	−0.0041	0.0005	0.0059	*IRF_9*	−0.0006	−0.0073	0.0001	0.0086
IRF_10	−0.0004	−0.0045	0.0004	0.0033	*IRF_10*	−0.0008	−0.0025	0.0005	0.0049

Table 20.10 IRFs when the True Marginals Are Student's t (υ= 3, υ= 10) with GARCH Effects; Employed Marginals for Estimation: Normal *without* GARCH Effects

True Model: Student's t (υ= 3) GARCH		Estimated Marginals: Normal			True Model: Student's t (υ= 10) GARCH		Estimated Marginals: Normal		
IRFs	True Values	Lower C. B.	IRF	Upper C. B.	IRFs	True Values	Lower C. B.	IRF	Lower C. B.
IRF_1	0.5984	0.3801	0.4774	0.6605	*IRF_1*	0.5263	0.4483	0.4955	0.5454
IRF_2	0.2992	0.1712	0.2349	0.3280	*IRF_2*	0.2631	0.1999	0.2452	0.2976
IRF_3	0.0299	−0.0297	0.0226	0.0748	*IRF_3*	0.0263	−0.0209	0.0225	0.0685
IRF_4	0.0449	−0.0191	0.0333	0.0872	*IRF_4*	−0.0395	−0.0115	0.0359	0.0815
IRF_5	0.0314	−0.0251	0.0233	0.0739	*IRF_5*	0.0276	−0.0182	0.0241	0.0666
IRF_6	−0.0037	−0.0450	−0.0040	0.0348	*IRF_6*	−0.0033	−0.0412	−0.0051	0.0323
IRF_7	−0.0029	−0.0185	−0.0015	0.0161	*IRF_7*	−0.0026	−0.0162	−0.0024	0.0129
IRF_8	0.0018	−0.0083	0.0023	0.0149	*IRF_8*	0.0015	−0.0065	0.0020	0.0122
IRF_9	−0.0007	−0.0081	0.0005	0.0106	*IRF_9*	−0.0006	−0.0071	0.0001	0.0081
IRF_10	−0.0009	−0.0027	0.0010	0.0069	*IRF_10*	−0.0008	−0.0026	0.0004	0.0052

Table 20.11 IRFs when the True Marginals Are Student's t (υ = 3, υ = 10) with GARCH Effects; Employed Marginals for Estimation: Normal *with* GARCH Effects

True Model: Student's t (υ= 3) GARCH		Estimated Marginals: Normal-GARCH			True Model: Student's t (υ= 10) GARCH		Estimated Marginals: Normal-GARCH		
IRFs	True Values	Lower C. B.	IRF	Upper C. B.	IRFs	True Values	Lower C. B.	IRF	Lower C. B.
IRF_1	0.5984	0.4767	0.5819	0.7676	*IRF_1*	0.5263	0.4490	0.4963	0.5463
IRF_2	0.2992	0.2240	0.2881	0.3764	*IRF_2*	0.2631	0.1999	0.2456	0.2981
IRF_3	0.0299	−0.0150	0.0330	0.0708	*IRF_3*	0.0263	−0.0212	0.0226	0.0686
IRF_4	0.0449	−0.0021	0.0559	0.0866	*IRF_4*	0.0395	−0.0119	0.0359	0.0816
IRF_5	0.0314	−0.0100	0.0327	0.0695	*IRF_5*	0.0276	−0.0183	0.0242	0.0667
IRF_6	−0.0037	−0.0443	0.0071	0.0410	*IRF_6*	−0.0033	−0.0413	−0.0051	0.0323
IRF_7	−0.0029	−0.0121	0.0067	0.0194	*IRF_7*	−0.0026	−0.0163	−0.0024	0.0129
IRF_8	0.0018	0.0002	0.0059	0.0181	*IRF_8*	0.0015	−0.0065	0.0020	0.0122
IRF_9	−0.0007	0.0004	0.0085	0.0126	*IRF_9*	−0.0006	−0.0071	0.0001	0.0081
IRF_10	−0.0009	0.0068	0.0188	0.0083	*IRF_10*	−0.0008	−0.0026	0.0004	0.0052

Table 20.12 IRFs when the True Marginals Are Student's t (υ = 3, υ = 10) with GARCH Effects; Employed Marginals for Estimation: Student's t *with* GARCH Effects

True Model: Student's t (υ= 3) GARCH		Estimated Marginals: Student's t-GARCH			True Model: Student's t (υ= 10) GARCH		Estimated Marginals: Student's t-GARCH		
IRFs	True Values	Lower C. B.	IRF	Upper C. B.	IRFs	True Values	Lower C. B.	IRF	Lower C. B.
IRF_1	0.5984	0.4562	0.5732	0.7970	*IRF_1*	0.5263	0.4520	0.5216	0.5754
IRF_2	0.2992	0.2047	0.2820	0.4002	*IRF_2*	0.2631	0.2103	0.2582	0.3128
IRF_3	0.0299	−0.0353	0.0271	0.0896	*IRF_3*	0.0263	−0.0220	0.0237	0.0719
IRF_4	0.0449	−0.0236	0.0400	0.1077	*IRF_4*	0.0395	−0.0121	0.0378	0.0859
IRF_5	0.0314	−0.0295	0.0280	0.0894	*IRF_5*	0.0276	−0.0191	0.0254	0.0702
IRF_6	−0.0037	−0.0540	−0.0048	0.0415	*IRF_6*	−0.0033	−0.0435	−0.0054	0.0340
IRF_7	−0.0029	−0.0225	−0.0019	0.0195	*IRF_7*	−0.0026	−0.0171	−0.0025	0.0136
IRF_8	0.0018	−0.0103	0.0028	0.0182	*IRF_8*	0.0015	−0.0068	0.0021	0.0128
IRF_9	−0.0007	−0.0097	0.0006	0.0126	*IRF_9*	−0.0006	−0.0075	0.0001	0.0086
IRF_10	−0.0009	−0.0033	0.0012	0.0085	*IRF_10*	−0.0008	−0.0027	0.0005	0.0054

We note that, as expected, confidence bands are wider with Student's t noise distributions than with normal noise distributions. This evidence clearly highlights the fact that neglecting leptokurtosis in empirical data may result in confidence bands which do not consider the greater uncertainty displayed by the underlying economic process. A policy maker may therefore consider a shock to be more persistent than it actually is. Besides, the use of wrong marginals underestimates the true IRFs, because of the negative biases in the dependence structure highlighted in the section Simulation Studies. That is why some caution should be taken when using the IRFs estimated with a multivariate normal VAR model implemented in all standard software packages.

CONCLUSION

This chapter analyzed copula-VAR and copula-VAR-GARCH modeling and highlighted some pitfalls that may emerge if wrong marginals are used, with particular attention to value at risk and impulse response functions analysis.

The simulation studies performed highlighted the fact that the use of normal marginals (such as in ordinary least squares estimates of normal VAR models), when the true data generating process is leptokurtic, produces biased estimates of the copula parameters vector. Particularly, we found evidence of a negative bias in the correlation parameters. However, this bias decreases as the number of degrees of freedom of the true DGP increases, and when the variance is constant. Besides, we also found evidence that small samples may increase such a bias, because of poor estimates of the marginals variance parameters.

We then assessed the potential impact of misspecified marginals on the estimation of multivariate VaR and we found that the marginals and dependence biases may result in lower VaR estimates, particularly for extreme quantiles, which are fundamental for risk management.

The chapter also developed a procedure to construct a set of IRFs for a copula-VAR-GARCH model with a multivariate distribution different from the normal one and with structural identifying assumptions. A procedure for computing bootstrapped standard errors was also presented. We found that the use of wrong marginals underestimates the IRFs and may result in narrower confidence bands. That is why some caution should be taken when using the IRFs estimated with a multivariate normal VAR model, as implemented in standard software packages.

An avenue of future research might be to perform a simulation analysis with higher dimensional portfolios and dynamic copulas.

REFERENCES

Amisano, G. and Giannini, C. (1997) *Topics in Structural VAR Econometrics*, 2nd ed. Berlin: Springer-Verlag.

Blanchard, O.J. and Quah, D. (1989) The Dynamic Effects of Aggregate Demand and Supply Disturbances. *American Economic Review*, 79(4): 655–673.

Chen, X., Fan, Y., and Patton, A. (2004) Simple Tests for Models of Dependence Between Multiple Financial Time Series with Applications to U.S. Equity Returns and Exchange Rates. FMG Technical Report No. 483, London School of Economics, London, UK.

Chernobai, A.S., Rachev, S.T., and Fabozzi, F.J. (2007) *Operational Risk: A Guide to Basel II Capital Requirements, Models and Analysis*. New York: Wiley.

Cherubini, U., Luciano, E., and Vecchiato, W. (2004) *Copula Methods in Finance*. New York: Wiley.

Fang, K.T., Kotz, S., and Ng, K.W. (1987) *Symmetric Multivariate and Related Distributions*. London: Chapman & Hall.

Fantazzini, D. (2008) Dynamic Copula Modelling for Value at Risk. *Frontiers in Finance and Economics*, 5(2): 72–108.

Fantazzini, D. (2009) The Effects of Misspecified Marginals and Copulas on Computing the Value at Risk: a Monte Carlo Study. *Computational Statistics & Data Analysis*, 53(6): 2168–2188.

Fantazzini, D. (2010) Three-Stage Semi-parametric Estimation of T-Copulas: Asymptotics, Finite-Sample Properties and Computational Aspects. *Computational Statistics & Data Analysis*, forthcoming.

Fantazzini, D., Dalla Valle, L., and Giudici, P. (2008) Copulae and Operational Risks. *International Journal of Risk Assessment and Management*, 9(3): 238–257.

Greene, W.H. (2002) *Econometric Analysis*, 5th ed. New York: Prentice Hall.

Hwang, S. and Valls Pereira, P.L. (2006) Small Sample Properties of GARCH Estimates and Persistence. *The European Journal of Finance*, 12(6–7): 473–494.

Jorion, P. (2007) *Value at Risk: The New Benchmark for Managing Financial Risk*, 3rd ed. New York: McGraw-Hill.

Mills, F.C. (1927) *The Behaviour of Prices*. New York: National Bureau of Economic Research.

Nelsen, R.B. (2006) *An Introduction to Copulas (Lecture Notes in Statistics)*, 2nd ed. New York: Springer.

Prokhorov, A.B. and Schmidt, P. (2009) Redundancy, Robustness, and Validity of Copulas in Likelihood Models. *Journal of Econometrics*, 153(1), 93–104

Sklar, A. (1959) Fonctions de Répartition à n Dimensions et Leurs Marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229–231.

Sims, C.A. (1980) Macroeconomics and Reality. *Econometrica*, 48(1): 1–48.

NOTES

1. As an example of copula flexibility, if we have three marginals, the first $F_{1,t}$ $(\eta_{1,t};\ \nu_1)$ may follow a Student's t distribution with ν_1 degrees of freedom, the second $F_{2,t}$ $(\eta_{2,t})$ a standard normal distribution, while the third $F_{3,t}$ $(\eta_{3,t};\ \lambda_3,\ \nu_3)$ may be a skewed Student's t distribution with ν_3 degrees of freedom, where λ_3 is the skewness parameter. Moreover, the copula function may be, for example, a normal one.

CHAPTER 21

SMALL SAMPLES AND EVT ESTIMATORS FOR COMPUTING RISK MEASURES

Simulation and Empirical Evidences

Dean Fantazzini and Alexander Kudrov

ABSTRACT

Extreme value theory (EVT) deals with the analysis of rare events and it has been recently used in finance to predict the occurrence of such events, or, at least, to build more robust models for unexpected extreme events. Particularly, EVT has been used to model the loss severities in operational risk management, while the use of GARCH-EVT models has gained popularity when computing value at risk (or other risk measures) in market risk management. To date, little attention has been devoted to the analysis of the small-sample properties of EVT estimators and their effects on the computation of financial risk measures. In this chapter we present and discuss the results of a Monte Carlo study of the small sample properties of these estimators in an operational risk setting, together with an empirical analysis dealing with market risk management.

INTRODUCTION

Extreme value theory (EVT) has become a valuable tool to assess the likelihood of rare but large events in financial risk management. EVT deals with the modeling of extreme events and aims at modeling the tails of a distribution. The first pioneering work about EVT dates back to Fisher and Tippett (1928), while Balkema and de Haan (1974) and Pickands (1975) presented the foundation for threshold-based extreme value methods. EVT methods have been applied to different areas, from hydrology to engineering, and recently made applicable to finance and insurance. Textbook level presentation can be found in Embrechts, Kluppelberg, and Mikosh (1997) and Reiss and Thomas (1997).

A well-known result in market risk management (MRM) is that absolute or squared financial log-returns are strongly autocorrelated, due to the empirical fact that a large absolute movement tends to be followed by a large absolute movement. Such phenomena are typically modeled as ARCH or GARCH processes (see Bollerslev, 1986), and risk measures are then computed based on the conditional forecasts for the mean and the variance. As an alternative way to improve relevant measures for market risk management, one can consider the two-step procedure proposed by McNeil and Frey (2000).

Recently, EVT methods have also been proposed in operational risk management (ORM) to model the tail of the severity distributions, in order to build more robust models for unexpected extreme events. This is extremely important when calculating a risk measure such as value at risk (VaR) or the expected shortfall (ES) at high confidence levels, such as the case of operational risk, where VaR at the 99.9 percent level is required. See Cruz (2002) for a review of EVT for operational risk.

Nevertheless, little attention has been devoted to the analysis of the small-sample properties of EVT estimators and their effects on the computation of financial risk measures, with the exception of Jalal and Rockinger (2008), who "investigate the consequences of using GARCH filtered returns when the data is generated either by some GARCH but with non-Gaussian innovations or some non-GARCH type process such as a switching regime process or a stochastic volatility with jumps model" (p. 876).

In this work we first present and discuss the results of a Monte Carlo study of the small-sample properties of EVT estimators, where the simulation data-generating processes (DGPs) are designed to reflect the stylized facts about real operational risk. Then, we compare different EVT estimators to compute risk measures for market risk management using very recent U.S. data, including also the global financial crisis. The rest of the chapter is organized as follows: the second section in this chapter shows how EVT can be used in operational risk management and discusses the small-sample properties of EVT estimators for this case. This chapter's third section compares different EVT estimators for market risk management by using recent U.S. data, while the final section concludes.

EVT AND OPERATIONAL RISK MANAGEMENT: SMALL-SAMPLE PROPERTIES

Theoretical Setup: The Standard LDA Approach

The standard loss distribution approach (LDA) for ORM employs two types of distributions: the one that describes the frequency of risk's events and the one that describes the severity of the losses that arise for each considered event. The frequency represents the number of loss events in a time horizon, while the severity is the loss associated to the *k*-th loss event. Formally, for each type of risk *i* (for example, business lines) and for a given time period, operational losses could be defined as a sum (S_i) of the random number (n_i) of the losses (X_{ij}): $S_i = X_{i1} + X_{i2} + \ldots + X_{i\,ni}$.

A widespread statistical model is the actuarial model. In this model, the probability distribution of S_i could be described as follows: $F_i(S_i) = F_i(n_i)^* F_i(X_{ij})$, where $F_i(S_i)$ is the probability distribution of the expected loss for risk *i*, $F_i(n_i)$ the probability of event (frequency) for the risk *i*, while $F_i(X_{ij})$ is the loss given event (severity) for the risk *i*. The underlying assumptions for the actuarial model are that the losses are random variables independent and identically distributed (i.i.d.), and the distribution of n_i (frequency) is independent of the distribution of X_{ij} (severity). In the actuarial model, the frequency of a loss event in a certain temporal horizon could be modeled by a *Poisson* distribution or a negative binomial. For the severity, we could use an *exponential*, a *pareto*, a *gamma* distribution, or the *generalized pareto distribution* (GPD). However, due to the extremely difficulty in estimating the parameter θ in small samples, as highlighted by Fantazzini et al. (2008) where, with $T = 72$, 40 percent of the simulated samples resulted in a negative θ, and even with a dataset of $T = 2{,}000$ observations the estimates of θ were not stable, we consider here only the Poisson distribution.

The distribution F_i of the losses for each intersection business line/event type *i* is obtained by the convolution of the frequency and severity distributions; nevertheless, the analytic representation of this distribution is computationally difficult or impossible. For this reason it is common to use a Monte Carlo simulation. A risk measure such as VaR or ES is then estimated to evaluate the capital requirement for that particular intersection *i*. See Fantazzini, Dallavalle, and Giudici (2008) and Rachedi and Fantazzini (2009) for more details about frequency and severity modeling.

Given the extreme behavior of operational risk losses, we can analyze the tail of the severity distributions using EVT and the GPD. In short, EVT affirms that the losses exceeding a given high threshold *u* converge asymptotically to the GPD, whose cumulative function is usually expressed as follows:

$$GPD_{\xi,\beta(y)} = \begin{cases} 1-\left(1+\xi\dfrac{y}{\beta}\right)^{-1/\xi} & \xi \neq 0 \\ 1-\exp\left(-\dfrac{y}{\beta}\right) & \xi = 0 \end{cases}$$

where $y = x - u$, $y \geq 0$ if $\xi \geq 0$ and $0 \leq y \leq -\beta/\xi$ if $\xi \leq 0$, and where y are called *excesses*, whereas x are *exceedances*. It is possible to determine the conditional distribution function of the excesses, i.e., y as a function of

$$x,\ F_u\left(y\right) = P\left(X - u \leq y \middle| X > u\right) = \frac{F_x\left(x\right) - F_x\left(u\right)}{1 - F_x\left(u\right)}$$

In these representations the parameter ξ is crucial: when $\xi = 0$, we have an exponential distribution; when $\xi < 0$, we have a Pareto distribution type II and when $\xi > 0$ we have a Pareto distribution type I. Moreover this parameter has a direct connection with the existence of finite moments of the losses distributions. We have that $E(x^k) = \infty$ if $k \geq 1/\xi$.

Hence, in the case of a GPD as a Pareto type I, when $\xi \geq 1$, we have infinite mean models, as also shown by Neslehova, Embrechts, and Chavez-Demoulin (2006). Following Di Clemente and Romano (2004) and Rachedi and Fantazzini (2009), we suggest to model the loss severity using the log-normal for the body of the distribution and EVT for the tail, in the following way:

$$F_i\left(x\right) = \begin{cases} \Phi\left(\dfrac{\ln x - \mu\left(i\right)}{\sigma\left(i\right)}\right) & 0 < x < u(i) \\ 1 - \dfrac{N_u\left(i\right)}{N_i}\left(1 + \xi\left(i\right)\dfrac{x - u\left(i\right)}{\beta\left(i\right)}\right)^{-1/\xi(i)} & u(i) \leq x, \end{cases}$$

where Φ is the standardized normal cumulative distribution functions, $N_u(i)$ is the number of losses exceeding the threshold $u(i)$, $N(i)$ is the number of the loss data observed in the i-th ET, whereas $\beta(i)$ and $\xi(i)$ denote the scale and the shape parameters of GPD.

Simulation Studies

In this section we present the results of a Monte Carlo study of the small-sample properties of different frequency-severity marginal estimators discussed in the previous sections. The simulation data-generating processes are designed to reflect the stylized facts about real operational risks and we chose the parameters of DGPs among the ones estimated in Fantazzini, Dallavalle, and Giudici (2008). We consider a total of 24 possible DGPs in Table 21.1.

In addition to these DGPs, we consider two possible data situations: (1) $T = 50$ and (2) $T = 500$. We generated 1,000 Monte Carlo samples for each marginal specification described in Table 21.1 and we estimated the VaR at the 99 and 99.9 percent levels, together with ES at the 99 and 99.9 percent levels. We then considered the following marginal estimators: (1) *Poisson-exponential*,

Table 21.1 Parameters of Frequency-Severity Models Used for 24 Simulated DGPs

DGP	Poisson	Exponential	
(1)	1.40	9844.48	
(2)	2.19	21721.43	
(3)	0.08	153304.55	
(4)	0.46	206162.38	
(5)	0.10	96873.25	
(6)	0.63	7596.41	
(7)	0.68	12623.41	
(8)	0.11	35678.13	
DGP	**Poisson**	**Gamma**	
(9)	1.40	0.15	64847.81
(10)	2.19	0.20	109320.57
(11)	0.08	0.20	759717.47
(12)	0.46	0.11	1827627.20
(13)	0.10	0.20	495700.99
(14)	0.63	0.38	19734.01
(15)	0.68	0.06	211098.10
(16)	0.11	0.26	135643.25
DGP	**Poisson**	**Pareto**	
(17)	1.40	2.36	13368.41
(18)	2.19	2.50	32493.69
(19)	0.08	2.51	230817.02
(20)	0.46	2.25	258587.73
(21)	0.10	2.49	143933.30
(22)	0.63	3.25	17104.96
(23)	0.68	2.13	14229.16
(24)	0.11	2.71	61145.58

(2) *Poisson-gamma*, (3) *Poisson-Pareto*, (4) *Poisson-log-normal-GPD*, and (5) *GPD* directly on the marginal loss S_i.

Tables 21.2–21.4 report the true VaR/ES at the 99 and 99.9 percent levels, the mean bias in percentage, the median bias in percentage, the relative root mean square error, i.e, the RMSE with respect to the true value, and the t test for the null hypothesis that the empirical mean across simulations is equal to the true value. For sake of interest and due to space limits, we report here only the results for $T = 50$ for the Poisson–log-normal–GPD model and for GPD directly fitted to the losses, while the results for $T = 500$ and for the remaining models are available from the authors upon request.

We estimated the Poisson and the exponential distribution by maximum likelihood methods, while we resorted to the method of moments for gamma and Pareto. Instead, we use the probability weighted moments for GPD, given the better small sample properties; see Rachedi and Fantazzini (2009) and Fantazzini, Dallavalle, and Giudici (2008) for more details.

- *True DGP-Poisson-Exponential (DGPs 1–8) (Table 21.2)*: If we use the Poisson-exponential or the Poisson-gamma marginal models, the VaR/ES estimates are already precise with $T = 50$ observations, where the former model is the most efficient, as expected. The use of the Poisson-Pareto model results in overestimated risk measures around 5 to 10 percent of the true values, while the Poisson–log-normal–GPD model delivers slightly underestimated VaR at the 99 and 99.9 percent level (around −10 percent when $T = 50$), whereas the ES at the 99.9 percent level is usually quite precise or slightly overestimated. Instead, using GPD directly on the marginal losses S_i results in strongly underestimated risk measures, between −20 and −40 percent when $T = 50$, while the empirical estimates are very close to the theoretical VaR and ES when $T = 500$.
- *True DGP-Poisson-Gamma (DGPs 9–16) (Table 21.3)*: The Poisson-exponential model results in strongly underestimated risk measures, around −50 and −70 percent, and the degree of underestimation remains constant over the time dimension. The correct Poisson-gamma model delivers negatively biased estimates when $T = 50$, between −10 and −20 percent of the true value and the median biases are slightly higher, thus highlighting a skewed empirical distribution when small samples are of concern. When $T = 500$, the risk measures are close to the true value and t statistics are already not statistically significant. The Poisson-Pareto model produces strongly

Table 21.2 VaR Estimation: Monte Carlo Results for DGPs 1–8 (Poisson-Exponential), $T = 50$

	T=50	Poisson_LogNormal_GPD					T=50	GPD_ONLY				
		True VaR	Mean %	Median %	RRMSE	T-stat		True VaR	Mean %	Median %	RRMSE	T-stat
DGP 1	VaR 99%	101615	−6.38	−8.06	0.19	−10.53	VaR 99%	101615	−14.19	−19.82	0.34	−42.03
	VaR 99.9%	185344	3.13	−3.34	0.41	2.42	VaR 99.9%	185344	−17.22	−31.11	0.54	−32.07
	ES 99%	137452	1.50	−2.72	0.32	1.47	ES 99%	137452	−14.59	−24.49	0.44	−33.08
	ES 99.9%	228411	28.30	9.41	1.06	8.47	ES 99.9%	228411	−2.96	−33.81	0.92	−3.23
DGP 2	VaR 99%	313211	−5.41	−6.92	0.20	−8.75	VaR 99%	313211	−14.27	−19.19	0.32	−44.86
	VaR 99.9%	553840	4.72	−2.66	0.36	4.13	VaR 99.9%	553840	−17.05	−29.65	0.51	−33.55
	ES 99%	415997	2.78	−1.70	0.29	3.04	ES 99%	415997	−14.57	−23.12	0.41	−35.22
	ES 99.9%	674750	31.11	10.87	0.90	10.95	ES 99.9%	674750	−3.48	−31.57	0.87	−4.02
DGP 3	VaR 99%	330012	−19.85	−18.72	0.35	−17.69	VaR 99%	330012	−16.43	−24.96	0.53	−30.92
	VaR 99.9%	723369	−10.50	−13.93	0.26	−12.78	VaR 99.9%	723369	−19.39	−36.29	0.68	−28.71
	ES 99%	501148	−11.13	−11.45	0.27	−12.82	ES 99%	501148	−17.12	−28.82	0.60	−28.69
	ES 99.9%	928222	0.23	−5.96	0.35	0.21	ES 99.9%	928222	−6.52	−40.79	1.06	−6.17
DGP 4	VaR 99%	1018803	−11.19	−12.26	0.22	−16.09	VaR 99%	1018803	−14.93	−21.54	0.38	−39.15
	VaR 99.9%	2025206	−2.95	−6.66	0.27	−3.50	VaR 99.9%	2025206	−19.96	−36.36	0.59	−33.95
	ES 99%	1443464	−3.55	−6.60	0.24	−4.64	ES 99%	1443464	−16.09	−28.55	0.50	−32.43
	ES 99.9%	2555860	13.77	1.36	0.55	7.98	ES 99.9%	2555860	−5.41	−38.70	0.98	−5.52
DGP 5	VaR 99%	225753	−17.85	−17.87	0.34	−16.38	VaR 99%	225753	−17.69	−25.84	0.51	−34.57
	VaR 99.9%	480685	−10.19	−12.93	0.27	−12.01	VaR 99.9%	480685	−18.77	−34.81	0.65	−28.86
	ES 99%	336726	−9.82	−11.22	0.28	−11.14	ES 99%	336726	−17.25	−28.64	0.57	−30.11
	ES 99.9%	616740	2.49	−4.43	0.40	1.95	ES 99.9%	616740	−6.00	−38.31	1.01	−5.95

(Continued)

Table 21.2 VaR Estimation: Monte Carlo Results for DGPs 1–8 (Poisson-Exponential), $T = 50$ (*Continued*)

		Poisson_LogNormal_GPD						GPD_ONLY				
	T=50	**True VaR**	**Mean %**	**Median %**	**RRMSE**	**T-stat**	**T=50**	**True VaR**	**Mean %**	**Median %**	**RRMSE**	**T-stat**
DGP 6	VaR 99%	45414	−8.42	−10.50	0.21	−12.52	VaR 99%	45414	−14.77	−21.67	0.38	−39.23
	VaR 99.9%	88368	−1.62	−5.56	0.26	−1.97	VaR 99.9%	88368	−18.18	−34.28	0.59	−30.79
	ES 99%	63756	−2.21	−5.11	0.23	−3.00	ES 99%	63756	−15.22	−27.25	0.49	−30.80
	ES 99.9%	110906	16.53	3.89	0.54	9.67	ES 99.9%	110906	−2.51	−36.25	1.00	−2.52
DGP 7	VaR 99%	79651	−7.97	−9.50	0.21	−12.20	VaR 99%	79651	−15.27	−21.44	0.37	−41.65
	VaR 99.9%	153445	−0.90	−6.48	0.29	−0.98	VaR 99.9%	153445	−18.96	−34.65	0.58	−32.96
	ES 99%	111252	−1.77	−6.10	0.25	−2.26	ES 99%	111252	−15.98	−27.03	0.48	−33.29
	ES 99.9%	192598	17.81	3.51	0.60	9.37	ES 99.9%	192598	−4.00	−37.13	0.97	−4.10
DGP 8	VaR 99%	88382	−15.41	−15.81	0.32	−15.27	VaR 99%	88382	−18.35	−26.09	0.48	−37.88
	VaR 99.9%	184445	−8.23	−11.35	0.28	−9.33	VaR 99.9%	184445	−18.18	−34.42	0.65	−27.96
	ES 99%	130202	−8.08	−10.02	0.27	−9.42	ES 99%	130202	−17.14	−28.69	0.56	−30.41
	ES 99.9%	237017	5.11	−2.99	0.45	3.59	ES 99.9%	237017	−5.06	−38.00	1.02	−4.94

Table 21.3 VaR Estimation: Monte Carlo Results for DGPs 9–16 (Poisson-Gamma), $T = 50$

		Poisson_LogNormal_GPD						GPD_ONLY				
	T=50	**True VaR**	**Mean %**	**Median %**	**RRMSE**	**T-stat**	**T=50**	**True VaR**	**Mean %**	**Median %**	**RRMSE**	**T-stat**
DGP 9	VaR 99%	225010	1.39E+03	3.56E+01	1.87E+02	2.35	VaR 99\%	225010	−19.70	−30.69	0.52	−38.02
	VaR 99.9%	560773	5.42E+07	1.22E+04	1.27E+07	1.35	VaR 99.9\%	560773	−26.51	−46.61	0.71	−37.57
	ES 99%	366154	7.58E+15	8.68E+05	2.39E+15	1.00	ES 99\%	366154	−21.53	−38.77	0.64	−33.58
	ES 99.9%	756196	3.64E+16	4.15E+06	1.15E+16	1.00	ES 99.9\%	756196	−10.02	−47.98	1.14	−8.82
DGP 10	VaR 99%	641189	1.99E+02	2.23E+01	1.61E+01	3.92	VaR 99\%	641189	−18.41	−27.44	0.47	−39.03
	VaR 99.9%	1442803	3.86E+05	2.35E+03	8.52E+04	1.43	VaR 99.9\%	1442803	−23.19	−41.86	0.65	−35.44
	ES 99%	979505	6.69E+08	2.27E+04	1.10E+08	1.92	ES 99\%	979505	−19.20	−33.63	0.58	−33.18
	ES 99.9%	1877655	3.46E+09	1.14E+05	5.71E+08	1.92	ES 99.9\%	1877655	−6.46	−42.82	1.05	−6.16
DGP 11	VaR 99%	382391	−3.74E+01	−5.38E+01	6.82E−01	−17.31	VaR 99\%	382391	−23.24	−55.21	0.98	−23.61
	VaR 99.9%	1629975	3.98E+02	6.50E+00	4.60E+01	2.74	VaR 99.9\%	1629975	−38.73	−66.91	0.89	−43.37
	ES 99%	905140	1.42E+07	2.06E+03	1.99E+06	2.26	ES 99\%	905140	−29.99	−60.15	0.93	−32.18
	ES 99.9%	2336919	5.46E+07	7.91E+03	7.64E+06	2.26	ES 99.9\%	2336919	−10.70	−62.05	1.39	−7.67
DGP 12	VaR 99%	2266368	4.16E+02	6.47E+00	8.58E+01	1.53	VaR 99\%	2266368	−24.07	−40.06	0.67	−35.98
	VaR 99.9%	6756605	1.04E+09	1.41E+04	2.13E+08	1.55	VaR 99.9\%	6756605	−29.72	−52.25	0.82	−36.39
	ES 99%	4155346	1.98E+20	7.52E+07	5.97E+19	1.05	ES 99\%	4155346	−24.42	−45.12	0.79	−30.92
	ES 99.9%	9640343	8.45E+20	3.21E+08	2.55E+20	1.05	ES 99.9\%	9640343	−9.52	−52.64	1.35	−7.07
DGP 13	VaR 99%	283012	−2.91E+01	−4.33E+01	6.86E−01	−13.41	VaR 99\%	283012	−27.50	−55.93	0.91	−30.14
	VaR 99.9%	1114108	5.46E+02	1.71E+01	4.42E+01	3.90	VaR 99.9\%	1114108	−38.22	−65.36	0.87	−43.88
	ES 99%	632332	1.37E+07	2.66E+03	1.71E+06	2.53	ES 99\%	632332	−30.75	−59.22	0.90	−34.29
	ES 99.9%	1581744	5.42E+07	1.06E+04	6.77E+06	2.53	ES 99.9\%	1581744	−10.01	−60.62	1.36	−7.34

(Continued)

Table 21.3 VaR Estimation: Monte Carlo Results for DGPs 9–16 (Poisson-Gamma), $T = 50$ *(Continued)*

		Poisson_LogNormal_GPD						GPD_ONLY				
	T=50	**True VaR**	**Mean %**	**Median %**	**RRMSE**	**T-stat**	**T=50**	**True VaR**	**Mean %**	**Median %**	**RRMSE**	**T-stat**
DGP 14	VaR 99%	65074	−5.06E−02	−4.23E+00	3.23E−01	−0.05	VaR 99\%	65074	−17.39	−25.94	0.45	−38.61
	VaR 99.9%	147027	2.51E+02	5.19E+01	8.48E+00	9.36	VaR 99.9\%	147027	−24.10	−42.19	0.65	−37.15
	ES 99%	99513	6.54E+02	9.22E+01	2.66E+01	7.76	ES 99\%	99513	−19.30	−33.29	0.57	−33.84
	ES 99.9%	193359	3.16E+03	4.13E+02	1.31E+02	7.64	ES 99.9\%	193359	−9.14	−44.63	1.05	−8.67
DGP 15	VaR 99%	218604	2.96E+06	4.13E+01	6.76E+05	1.39	VaR 99\%	218604	−25.62	−45.06	0.74	−34.75
	VaR 99.9%	757972	5.08E+18	6.74E+07	1.23E+18	1.30	VaR 99.9\%	757972	−35.29	−59.68	0.82	−42.85
	ES 99%	7.30E+37	1.86E+16	2.28E+37	1.01	ES 99\%	444283	−28.45	−52.19	0.82	−34.73	
	ES 99.9%	1122525	2.86E+38	7.29E+16	8.96E+37	1.01	ES 99.9\%	1122525	−15.42	−59.42	1.27	−12.17
DGP 16	VaR 99%	115926	−2.26E+01	−3.03E+01	5.53E−01	−12.95	VaR 99\%	115926	−27.23	−47.62	0.75	−36.17
	VaR 99.9%	371620	1.70E+02	7.42E+00	1.45E+01	3.71	VaR 99.9\%	371620	−31.16	−55.00	0.81	−38.23
	ES 99%	224627	6.61E+05	2.67E+02	1.53E+05	1.37	ES 99\%	224627	−26.38	−49.32	0.80	−32.87
	ES 99.9%	514208	2.86E+06	1.15E+03	6.62E+05	1.37	ES 99.9\%	514208	−4.58	−51.25	1.32	−3.48

Table 21.4 VaR Estimation: Monte Carlo Results for DGPs 17–24 (Poisson-Pareto), $T = 50$

	Poisson_LogNormal_GPD						GPD_ONLY					
	T=50	True VaR	Mean %	Median %	RRMSE	T-stat	T=50	True VaR	Mean %	Median %	RRMSE	T-stat
DGP 17	VaR 99%	144665	−7.86	−19.63	0.47	−5.31	VaR 99%	144665	−13.10	−30.05	0.72	−18.20
	VaR 99.9%	448613	−19.64	−41.42	0.77	−8.07	VaR 99.9%	448613	−31.12	−58.18	1.19	−26.10
	ES 99%	278714	−14.72	−34.58	0.71	−6.55	ES 99%	278714	−23.63	−48.56	1.16	−20.37
	ES 99.9%	799700	−21.86	−53.12	1.12	−6.15	ES 99.9%	799700	−28.73	−69.33	1.92	−14.94
DGP 18	VaR 99%	454143	−9.69	−19.76	0.43	−7.06	VaR 99%	454143	−13.64	−28.60	0.63	−21.73
	VaR 99.9%	1327934	−21.68	−40.38	0.72	−9.54	VaR 99.9%	1327934	−30.20	−55.62	1.06	−28.51
	ES 99%	837643	−16.58	−33.19	0.65	−8.10	ES 99%	837643	−22.97	−45.99	1.01	−22.85
	ES 99.9%	2280756	−22.86	−50.25	1.03	−7.04	ES 99.9%	2280756	−26.87	−66.79	1.71	−15.71
DGP 19	VaR 99%	309696	−17.81	−23.39	0.45	−12.60	VaR 99%	309696	−8.98	−33.99	0.97	−9.22
	VaR 99.9%	1165924	−14.50	−29.42	0.57	−8.11	VaR 99.9%	1165924	−33.68	−63.37	1.13	−29.89
	ES 99%	686919	−14.51	−25.36	0.53	−8.59	ES 99%	686919	−25.26	−54.08	1.14	−22.21
	ES 99.9%	2128319	−19.79	−42.92	0.92	−6.81	ES 99.9%	2128319	−33.09	−74.25	1.62	−20.49
DGP 20	VaR 99%	1324492	−9.25	−17.53	0.40	−7.29	VaR 99%	1324492	−11.33	−31.61	1.19	−9.56
	VaR 99.9%	4444021	−20.65	−37.59	0.61	−10.78	VaR 99.9%	4444021	−28.99	−60.54	2.19	−13.24
	ES 99%	2730023	−17.67	−31.74	0.55	−10.20	ES 99%	2730023	−22.01	−51.97	2.23	−9.87
	ES 99.9%	8406870	−28.78	−51.22	0.81	−11.20	ES 99.9%	8406870	−26.47	−71.97	3.76	−7.04
DGP 21	VaR 99%	219709	−19.81	−26.00	0.45	−13.94	VaR 99%	219709	−14.34	−36.77	0.90	−15.87
	VaR 99.9%	802824	−18.02	−32.20	0.57	−10.07	VaR 99.9%	802824	−34.70	−63.07	1.19	−29.12
	ES 99%	475132	−17.27	−28.27	0.54	−10.03	ES 99%	475132	−27.00	−54.12	1.21	−22.33
	ES 99.9%	1450893	−22.27	−43.10	1.01	−7.00	ES 99.9%	1450893	−32.59	−72.86	1.84	−17.75

(Continued)

Table 21.4 VaR Estimation: Monte Carlo Results for DGPs 17–24 (Poisson-Pareto), $T = 50$ (*Continued*)

		Poisson_LogNormal_GPD						GPD_ONLY				
	T=50	**True VaR**	**Mean %**	**Median %**	**RRMSE**	**T-stat**	**T=50**	**True VaR**	**Mean %**	**Median %**	**RRMSE**	**T-stat**
DGP 22	VaR 99%	55870	−9.20	−14.73	0.35	−8.25	VaR 99%	55870	−15.49	−28.25	0.57	−27.32
	VaR 99.9%	149545	−12.38	−26.53	0.59	−6.65	VaR 99.9%	149545	−28.85	−51.59	0.96	−29.93
	ES 99%	96083	−9.04	−20.19	0.51	−5.55	ES 99%	96083	−22.02	−42.31	0.90	−24.54
	ES 99.9%	232701	−7.87	−34.10	0.91	−2.74	ES 99.9%	232701	−21.47	−60.18	1.66	−12.92
DGP 23	VaR 99%	108813	−5.18	−18.69	0.57	−2.86	VaR 99%	108813	−12.22	−31.81	0.78	−15.62
	VaR 99.9%	376892	−15.02	−41.55	1.00	−4.77	VaR 99.9%	376892	−32.77	−62.25	1.19	−27.61
	ES 99%	230677	−12.14	−35.63	0.91	−4.23	ES 99%	230677	−25.75	−53.59	1.17	−22.09
	ES 99.9%	736264	−21.40	−56.84	1.38	−4.91	ES 99.9%	736264	−33.37	−74.22	1.74	−19.19
DGP 24	VaR 99%	89317	−13.82	−19.34	0.41	−10.76	VaR 99%	89317	−15.66	−34.99	1.35	−11.62
	VaR 99.9%	302301	−16.05	−28.96	0.57	−8.93	VaR 99.9%	302301	−31.18	−59.61	2.50	−12.50
	ES 99%	181472	−13.34	−23.46	0.49	−8.60	ES 99%	181472	−23.75	−50.74	2.62	−9.05
	ES 99.9%	516015	−19.22	−38.50	0.83	−7.36	ES 99.9%	516015	−23.83	−69.07	4.87	−4.89

underestimated estimates (-10 to -20 percent) which do not improve when the time dimension T increases. Interestingly, the Poisson-log-normal-GPD model results in exploding risk measures, particularly the ES which is often over e^{16}, which decrease very slowly when the sample dimension T increases. Unreported results show that only when $T > 10{,}000$ this model starts delivering realistic risk measures. Instead, if we apply the GPD directly on the losses S_i, we obtain strongly underestimated risk measures (mean bias, -20 to -30 percent; median bias, -40 to -50 percent) when $T = 50$, whereas much better estimates are obtained with $T = 500$ (around -10 percent), even though they are still significantly different from the true estimates.

- *True DGP-Poisson-Pareto (DGPs 17–24):* The use of the Poisson-exponential marginal model result in strongly underestimated risk measures, particularly the ES which can reach a mean bias of -70 percent, and the degree of underestimation remain constant over the time dimension, similarly to what we saw when the true DGP is a Poisson-gamma model. The Poisson-gamma model results in an overestimated VaR at the 99 percent level, while the other risk measures are found to be underestimated, particularly the ES at the 99.9 percent level. A similar pattern is found for the correct Poisson-Pareto model, too, which shows in some cases worse results than the misspecified Poisson-gamma model when $T = 50$. When the sample dimension increases, both models improve their performances, but still some problems remain with ES at the 99.9 percent level which is strongly underestimated (-20 to -30 percent). Interestingly, the Poisson–log-normal–GPD model is now the one delivering the most precise estimates when $T = 500$, particularly when ES is of concern, while its performance is similar (if not better) to the Poisson-gamma and Poisson-Pareto models when dealing with small samples ($T = 50$). Finally, the GPD strongly underestimates all the considered risk measures when $T = 50$ (up to ms70 percent), while it delivers rather precise estimates when $T = 500$.

To sum up, the previous simulation studies show that EVT works fine with medium to large datasets ($T > 500$), particularly if the GPD is directly fitted to the losses S_i, while it is rather problematic when dealing with small samples. In this case, the Poisson-gamma model represents a good compromise if both Var and ES are of concern, while the Poisson-Pareto model is a more suitable model for extremely high ES estimates.

EVT AND MARKET RISK MANAGEMENT: EMPIRICAL EVIDENCE WITH RECENT U.S. DATA

In this section, we present four different EVT estimators and two standard methods, and we compare them by computing VaR at different confidence levels for the S&P 500 index over the period from January 2, 2003, to May 5, 2009, thus including the recent global financial crisis.

EVT Estimators

Proposition 1: *Estimation Algorithm for VaR Based on AR(1)-T-GARCH(1,1) Model*

This is the benchmark model for financial returns; see, for example, Hansen and Lunde (2005) and references therein.

- *Step 1:* For log-returns during any fixed consecutive 1,000 days interval (within the considered period of time), which we denote as $[T-999, T]$, to estimate the parameters of and AR(1)-T-GARCH(1,1) model, which is specified as

$$\begin{aligned} X_t &= \mu + \phi X_{t-1} + \varepsilon_t, \\ \varepsilon_t &= \sigma_t \eta_t, \ \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \ t \in [T-999, T] \end{aligned} \tag{21.1}$$

  where (η_t) are i.i.d. random variables, which have student distribution with some ν degrees of freedom; μ, φ as the real-valued parameters of the model and $\omega, \alpha, \beta, \nu$ as the positive real-valued parameters of the model (to ensure positive conditional variance).
- *Step 2:* Using the model specification and its estimated parameters, we calculate VaR at 1 percent level (lower tail) and 99 percent level (upper tail) for the distribution function of X_{T+1}, respectively, as follows:

$$\begin{aligned} VaR_{0.01}(X_{T+1}) &= \hat{\mu} + \hat{\phi} X_T + st_{\hat{\nu}}^{-1}(0.01)\sqrt{\hat{\omega} + \hat{\alpha}\hat{\varepsilon}_T^2 + \hat{\beta}\hat{\sigma}_T^2}, \\ VaR_{0.99}(X_{T+1}) &= \hat{\mu} + \hat{\phi} X_T + st_{\hat{\nu}}^{-1}(0.99)\sqrt{\hat{\omega} + \hat{\alpha}\hat{\varepsilon}_T^2 + \hat{\beta}\hat{\sigma}_T^2}, \end{aligned} \tag{21.2}$$

  where for $0 < \alpha < 1$ the value $st_{\hat{\nu}}^{-1}(\alpha)$ is defined as α-th quantile of the student distribution with $\hat{\nu}$ degrees of freedom.[1] Herein, we will denote the estimates by the letters with hats.

Proposition 2: *Estimation Algorithm for VaR Based on AR(1)-T-GJR(1,1) Model*[2]

- *Step 1:* We consider the log-returns during any fixed consecutive 1,000 days interval (within the considered period of time), which denote as $[T-999,T]$, and for them we estimate the AR(1)-T-GJR(1,1) model, which has the following specification:

$$\begin{aligned} X_t &= \mu + \phi X_{t-1} + \varepsilon_t, \\ \varepsilon_t &= \sigma_t \eta_t, \\ \sigma_t^2 &= \omega + \alpha\varepsilon_{t-1}^2 + \xi I_{(\varepsilon_{t-1}>0)}\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2,\ t \in [T-999, T] \end{aligned} \tag{21.3}$$

  where (η_t) are i.i.d. random variables, which have student distribution with some ν degrees of freedom; $I_{(\varepsilon_{t-1}>0)}$ as the indicator-function of the event $(\varepsilon_{t-1} > 0)$; μ,φ as the real-valued parameters of the model and ω, α, ξ, β, ν as the positive real-valued parameters of the model (to ensure the positivity of the conditional variance). As compared with the previous AR(1)-T-GARCH(1,1) model, in which the positive and negative shocks have the same effects on conditional mean, the AR(1)-T-GJR(1,1) model takes into account the asymmetries in impacts of positive and negative shocks on the volatility (or equivalently, on conditional volatility), which are proper for financial time series.
- *Step 2:* Using the model specification and its estimated parameters, we calculate the VaR at 1 percent level (lower tail) and 99 percent level (upper tail) for the distribution function of X_{T+1}, respectively, as follows:

$$\begin{aligned} VaR_{0.01}(X_{T+1}) &= \hat{\mu} + \hat{\phi}X_T + st_{\hat{\nu}}^{-1}(0.01)\hat{\sigma}_{T+1}, \\ VaR_{0.99}(X_{T+1}) &= \hat{\mu} + \hat{\phi}X_T + st_{\hat{\nu}}^{-1}(0.99)\hat{\sigma}_{T+1}, \end{aligned} \tag{21.4}$$

  where $\sigma_{T+1} = \sqrt{\hat{\omega} + \hat{\alpha}\hat{\varepsilon}_T^2 + \hat{\xi} I_{(\hat{\varepsilon}_T>0)}\hat{\varepsilon}_T^2 + \hat{\beta}\hat{\sigma}_T^2}$; and for $0 < \alpha < 1$ the value $st_{-1\nu}(\alpha)$ is defined as α-th quantile of the student distribution with ν degrees of freedom.

Proposition 3: *McNeil and Frey's Approach for Estimation of VaR*[3]

- *Step 1:* By using the pseudo-maximum likelihood approach, we fit an AR(1)-GARCH(1,1) model (with normal standardized innovations) to the log-returns for the fixed period $[T-999,T]$, where $T = 1000,\ldots, N$

(N is the number of the last day in our sample) and calculate the implied model's standardized residuals.[4]

$$\begin{aligned} X_t &= \mu + \phi X_{t-1} + \varepsilon_t, \\ \varepsilon_t &= \sigma_t \eta_t, \quad \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \ t \in [T-999, T] \end{aligned} \tag{21.5}$$

The procedure of Step 1 is called *AR(1)-N-GARCH(1,1) filtration*.

- *Step 2:* We calculate the levels from which extremely positive or extremely negative standardized residuals will be defined. For the extremely positive standardized residuals, the level equals 90 percent-order statistics (denoted as $U(T)$) and every standardized residual, which excesses this level is considered as extremely high. Then, we calculate the values of excesses relative to them (which we call *the upper tail exceedances*). For the extremely negative standardized residuals, the level equals 10 percent-order statistics (denote it as $L(T)$) and every standardized residual, which is lower than this level, is considered as extremely low. Then, we calculate the differences between the 10 percent-order statistics and extremely low residuals (which we call *the lower tail exceedances*).
- *Step 3:* We fit the generalized Pareto distribution to the upper tail exceedances and to the lower tail exceedances, respectively: $G(y) = 1 - (1 + \xi \frac{y}{\beta})^{-1/\xi}$, using the method of the maximum likelihood. Denote ξ_U, β_U as the estimators of GPD for the upper tail exceedances and ξ_L, β_L as the estimators of GPD for the lower tail exceedances.
- *Step 4:* Finally, we calculate VaR at 99 percent level (lower tail) and at 1 percent level (upper tail) for the distribution function of X_{T+1}. Let's denote with $F(x)$ the distribution function of the standardized residuals Z. We have that:

$$F(x) = 1 - [1 - F(U(T))]\left[1 - \frac{F(x) - F(U(T))}{1 - F(U(T))}\right].$$

For $x \leq U(T)$ we approximate $F(x)$ by the empirical distribution function of the standardized residuals. Thus, $1 - F(U(T))$ is approximated by 0.1. And for $x > U(T)$ we approximate

$$\left[1 - \frac{F(x) - F(U(T))}{1 - F(U(T))}\right] = \left[\frac{1 - F(x)}{1 - F(U(T))}\right] = P(Z > x \mid Z > U(T)), \text{ by}$$

$$G(x - U(T)) = \left(1 + \xi_U \frac{x - U(T)}{\beta_U}\right)^{-1/\xi_U}.$$

So that the p-th quantile (where $0.9 < p < 1$) of the distribution function $F(x)$ approximation is calculated as:

$$z_p = U(T) + \frac{\beta_U}{\xi_U}\left(\left(\frac{1-p}{0.1}\right)^{-\xi_U} - 1\right).$$

Similarly we get that p-th quantile (where $0 < p < 0.1$) of the distribution function $F(x)$ approximation is calculated as:

$$z_p = L(T) - \frac{\beta_L}{\xi_L}\left(\left(\frac{1-p}{0.1}\right)^{-\xi_L} - 1\right).$$

For more detail see Frey and McNeil (2000) and references therein. Therefore, we have

$$\begin{aligned} VaR_{0.01}(X_{T+1}) &= \hat{\mu} + \hat{\phi}X_T + z_{0.01}\sqrt{\hat{\omega} + \hat{\alpha}\hat{\varepsilon}_T^2 + \hat{\beta}\hat{\sigma}_T^2}, \\ VaR_{0.99}(X_{T+1}) &= \hat{\mu} + \hat{\phi}X_T + z_{0.99}\sqrt{\hat{\omega} + \hat{\alpha}\hat{\varepsilon}_T^2 + \hat{\beta}\hat{\sigma}_T^2}. \end{aligned} \tag{21.6}$$

Proposition 4: *Gonzalo and Olmo's Approach for Estimation of VaR*[5]

- *Step 1:* We make *AR(1)-N-GARCH(1,1) filtration* as in the first step of *Proposition 3*.
- *Step 2:* We compute the set of Hill's estimators extremal indexes[6] $(\gamma^l(k))$, $(\gamma^u(k))$ both for the lower and upper tail of the distribution function for standardized residuals. Suppose that we have m_u positive standardized residuals $\hat{\eta}_1^u, \ldots, \hat{\eta}_{m_u}^u$ and let $0 < \hat{\eta}_{(m_u)}^u \leq \ldots \leq \hat{\eta}_{(1)}^u$ be their ordered statistics. Then the set of Hill's estimators $(\gamma^u(k))$ is defined as follows:

$$\gamma^u(k) = \frac{1}{k}\sum_{i=1}^{k}(\log(\hat{\eta}_{(i)}^u) - \log(\hat{\eta}_{(k+1)}^u)),$$

  where $1 \leq k \leq m_u - 1$. We will call the order statistics $\hat{\eta}_{(k+1)}^u$ *the* $k+1$*-th upper threshold*. For the negative standardized residuals, we consider their absolute values and construct for them the set of Hill's estimators $(\gamma^l(k))$ as for the positive standardized residuals.
- *Step 3:* In order to get the estimator for the upper tail quantiles, we consider the sequence of functions corresponding to the upper thresholds $(\hat{\eta}_{(k+1)}^u)$:

$$G_k(y) = 1 - \left(\frac{y}{\hat{\eta}_{(k+1)}^u}\right)^{-\frac{1}{\gamma^u(k)}}, \text{ for } y > \hat{\eta}_{(k+1)}^u \text{ and } G_k(y) = 0, \text{otherwise.}$$

Suppose $\hat{F}_{\hat{\eta}^u_{(k+1)}}(y)$ is the empirical distribution function for the $\hat{\eta}^u_1, \ldots, \hat{\eta}^u_{m_u}$ exceeding the threshold $\hat{\eta}^u_{(k+1)}$. We regard $\gamma^u(k_0)$ as optimal estimator of the upper tail extremal index, if the following equity holds:

$$k_0 = \text{argmin}_k \sqrt{k} \sup_y \left| F_{\hat{\eta}^u_{(k+1)}}(y) - G_k(y) \right|.$$

Let's denote with $F(x)$ the distribution function of the standardized residuals Z. For all x we have:

$$F(x) = 1 - [1 - F(\hat{\eta}^u_{(k_0+1)})]\left[1 - \frac{F(x) - F(\hat{\eta}^u_{(k_0+1)}))}{1 - F(\hat{\eta}^u_{(k_0+1)}))}\right]. \text{When } x \leq \hat{\eta}^u_{(k_0+1)}),$$

we approximate $F(x)$ by the empirical distribution function of the standardized residuals (not only positive standardized residuals). Thus $1 - F(\hat{\eta}^u_{(k_0+1)})$ is approximated by k_0/N, where N is the size of the standardized residuals sample. And for $x > (\hat{\eta}^u_{(k_0+1)})$, we approximate

$$\left[1 - \frac{F(x) - F(\hat{\eta}^u_{(k_0+1)})}{1 - F(\hat{\eta}^u_{(k_0+1)})}\right] = \left[\frac{1 - F(x)}{1 - F(\hat{\eta}^u_{(k_0+1)})}\right] = P(Z > x \mid Z > \hat{\eta}^u_{(k_0+1)}) \text{ by } G_{k_0}(x).$$

So that the p-th quantile (where $0.9 < p < 1$) of the distribution function $F(x)$ approximation is calculated as:

$$z_p = \hat{\eta}^u_{(k_0+1)} \left(\frac{1-p}{k_0/N}\right)^{-\gamma^u(k_0)}.$$

Analogously, we have for the lower tail, that the p-th quantile (where $0 < p < 0.1$) of the distribution function $F(x)$ approximation is calculated as:

$$z_p = -\hat{\eta}^u_{(k_1+1)} \left(\frac{1-p}{k_1/N}\right)^{-\gamma^u(k_1)},$$

where k_1 corresponds to the optimal threshold $\eta^{\hat{u}}_{(k_1+1)}$ for absolute values of the negative standardized residuals. For more detail see Gonzalo and Olmo (2004). Hence, we have:

$$\begin{aligned} VaR_{0.01}(X_{T+1}) &= \hat{\mu} + \hat{\phi}X_T + z_{0.01}\sqrt{\hat{\omega} + \hat{\alpha}\hat{\varepsilon}_T^2 + \hat{\beta}\hat{\sigma}_T^2}, \\ VaR_{0.99}(X_{T+1}) &= \hat{\mu} + \hat{\phi}X_T + z_{0.99}\sqrt{\hat{\omega} + \hat{\alpha}\hat{\varepsilon}_T^2 + \hat{\beta}\hat{\sigma}_T^2}. \end{aligned} \quad (21.7)$$

Proposition 5: *Huisman, Koedijk, and Pownell's Approach for Estimation of VaR*[7]

- *Step 1:* We make *AR(1)-N-GARCH(1,1) filtration* as in the first step of *Proposition 3*.
- *Step 2:* We compute the set of Hill's estimators extremal indexes[8] $(\gamma^l(k))$, $(\gamma^u(k))$ both for the lower and upper tail of the distribution function for standardized residuals. Suppose that we have m_u positive standardized residuals $\hat{\eta}_1^u, \ldots, \hat{\eta}_{m_u}^u$ and let $0 < \hat{\eta}_{(m_u)}^u \leq \ldots \leq \hat{\eta}_{(1)}^u$ be their ordered statistics. Then the set of Hill's estimators $(\gamma^u(k))$ is defined as follows:

$$\gamma^u(k) = \frac{1}{k}\sum_{i=1}^{k}(\log(\hat{\eta}_{(i)}^u) - \log(\hat{\eta}_{(k+1)}^u)),$$

  where $k = 1, \ldots, m_u - 1$. For the negative standardized residuals we consider their absolute values and construct for them the set of Hill's estimators $(\gamma^l(k))$ similarly as for the positive standardized residuals.
- *Step 3:* Under certain regularity conditions of the distribution function with heavy tails, we can consider the following model for $(\gamma^u(k))$ (and for $\gamma^l(k)$):

$$\gamma^u(k) = \gamma^u + b_1^u k + \varepsilon_k^u, \quad k = 1, \ldots, \kappa, \tag{21.8}$$

  where $E(\gamma^u(k)) = \gamma^u + b_1^u k$, $Var(\varepsilon_k^u) = \sigma^2/k$ and γ^u is the true value of the extremal index for the upper tail of the distribution function. Then we can estimateγ^u by using the method of weighted least squares with a weighting $\kappa \times \kappa$ matrix W (that has $(\sqrt{1}, \ldots, \sqrt{\kappa})$ on the main diagonal and zeroes elsewhere). Analogously we may estimate γ^l as the true value of the extremal index for the lower tail of the distribution function.
- *Step 4:* We compute the upper tail p-th quantiles ($0.9 < p < 1$) for the standardized innovations as follows:

$$z_p = \sqrt{\frac{1/\gamma^u - 2}{1/\gamma^u}} st_{1/\gamma^u}^{-1}.$$

  And the lower tail p-th quantiles ($0 < p < 0.1$) in the following way:

$$z_p = \sqrt{\frac{1/\gamma^l - 2}{1/\gamma^l}} st_{1/\gamma^l}^{-1},$$

where γ^l is the estimator for the lower tail extremal index. Multiplication by

$$\sqrt{\frac{1/\gamma^u - 2}{1/\gamma^u}} \text{ or by } \sqrt{\frac{1/\gamma^l - 2}{1/\gamma^l}}$$

is needed for the standardization. Finally, we calculate the VaR at 1 percent level (lower tail) and 99 percent level (upper tail) for the distribution function of X_{T+1}, respectively, as follows:

$$\begin{aligned} VaR_{0.01}(X_{T+1}) &= \hat{\mu} + \hat{\phi}X_T + z_{0.01}\sqrt{\hat{\omega} + \hat{\alpha}\hat{\varepsilon}_T^2 + \hat{\beta}\hat{\sigma}_T^2}, \\ VaR_{0.99}(X_{T+1}) &= \hat{\mu} + \hat{\phi}X_T + z_{0.99}\sqrt{\hat{\omega} + \hat{\alpha}\hat{\varepsilon}_T^2 + \hat{\beta}\hat{\sigma}_T^2}, \end{aligned} \tag{21.9}$$

Empirical Analysis

In order to illustrate the comparative analysis of the algorithms described in the preceding subsections, we use the S&P 500 index over the period from January 2, 2003 to May 5, 2009. We analyze the quality of 1 percent (lower tail) and 99 percent (upper tail) VaR measure for the log-returns X_t of the daily prices P_t estimation, for which we have $X_t = \log(P_t) - \log(P_{t-1})$, $t - 1001, \ldots, 1604$ ($t = 1604$ corresponds to the date May 5, 2009). If we use the perfect VaR models, the violations should not be predictable and the $VaR_{0.99}(X_t\ VaR_{0.01}(X_t)$, violations should be simply percent and percent every day, respectively. We employ both the Kupiec's unconditional coverage test (1995) and the Christoffersen's conditional coverage test (1998), given their importance in the empirical literature. However, we remark that their power can be very low. Alternatively, one can consider tests based on the whole distribution, such as Berkowitz (2001) or Granger, Patton, and Terasvirta (2006).

Table 21.5 highlights some mixed results and the outcomes are not as satisfactory as one may have expected, particularly for the standard EVT estimators discussed in propositions 3 and 5.

CONCLUSION

We have presented and discussed the results of a Monte Carlo study of the small-sample properties of EVT estimators, where the simulation DGPs were designed to reflect the stylized facts about real operational risk. We found out that EVT works fine with medium to large datasets ($T > 500$), particularly if the GPD is directly fitted to the losses S_i, while it is rather problematic when dealing with small samples. Next, we compared different

Table 21.5 Results of Conditional, Unconditional, and Independence Tests

Proposition Number	Unconditional Test	Independence Test	Conditional Test
Proposition 1 ($VaR_{0.01}$)	H_0 is *not* rejected at 95% c.l.	H_0 is *not* rejected at 95% c.l.	H_0 is *not* rejected at 95% c.l.
Proposition 1 ($VaR_{0.99}$)	H_0 is rejected at 99% c.l.	H_0 is rejected at 99% c.l.	H_0 is rejected at 99% c.l.
Proposition 2 ($VaR_{0.01}$)	H_0 is *not* rejected at 95% c.l.	H_0 is *not* rejected at 95% c.l.	H_0 is *not* rejected at 99% c.l.
Proposition 2 ($VaR_{0.99}$)	H_0 is rejected at 99% c.l.	H_0 is rejected at 99% c.l.	H_0 is rejected at 99% c.l.
Proposition 3 ($VaR_{0.01}$)	H_0 is rejected at 99% c.l.	H_0 is *not* rejected at 95% c.l.	H_0 is rejected at 99% c.l.
Proposition 3 ($VaR_{0.99}$)	H_0 is *not* rejected at 95% c.l.	H_0 is *not* rejected at 95% c.l.	H_0 is *not* rejected at 95% c.l.
Proposition 4 ($VaR_{0.01}$)	H_0 is rejected at 99% c.l.	H_0 is *not* rejected at 95% c.l.	H_0 is rejected at 99% c.l.
Proposition 4 ($VaR_{0.99}$)	H_0 is *not* rejected at 95% c.l.	H_0 is *not* rejected at 95% c.l.	H_0 is *not* rejected at 95% c.l.
Proposition 5 ($VaR_{0.01}$)	H_0 is rejected at 99 c.l.	H_0 is rejected at 99% c.l.	H_0 is rejected at 99% c.l.
Proposition 5 ($VaR_{0.99}$)	H_0 is *not* rejected at 95% c.l.	H_0 is *not* rejected at 95% c.l.	H_0 is *not* rejected at 95% c.l.

c.l., confidence level.

EVT estimators to compute risk measures for market risk management using very recent U.S. data up to May 2009. Standard EVT estimators were not completely satisfactory.

REFERENCES

Bollerslev, T. (1986) Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31(3): 307–327.

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

Berkowitz, J. (2001) Testing Density Forecasts with Applications to Risk Management, *Journal of Business and Economic Statistics*, 19: 465–474.

Christoffersen, P. (1998) Evaluating Interval Forecasts. *International Economic Review*, 39(4): 841–862.

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

Di Clemente, A. and Romano, C. (2004) A Copula-Extreme Value Theory Approach for Modelling Operational Risk. In M.G. Cruz (ed.), *Operational Risk Modelling and Analysis: Theory and Practice*. London: Risk Books.

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

Fantazzini, D., Dallavalle, L., and Giudici, P. (2008) Copulae and Operational Risks. *International Journal of Risk Assessment and Management*, 9(3): 238–257.

Fisher, R. and Tippett, L. (1928) Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample. *Proceedings of the Cambridge Philosophical Society*, 24: 180–190.

Glosten L.R., Jagannathan, R., and Runkle, D. (1993) On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance*, 48(5): 1779–1801.

Granger, C., Patton, A., and Terasvirta, T. (2006) Common Factors in Conditional Distributions for Bivariate Time Series. *Journal of Econometrics*, 132: 43–57.

Gonzalo, J. and Olmo, J. (2004) Which Extreme Values Are Really Extreme? *Journal of Financial Econometrics*, 3(2): 349–369.

Hansen, P.R. and Lunde, A. (2005) A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)? *Journal of Applied Econometrics*, 20: 873–889.

Hill, B.M. (1975) A Simple General Approach to Inference about the Tail of a Distribution. *Annals of Statistics*, 3: 1163–1174.

Huisman R., Koedijk, K.G., and Pownall, R. (1998). VaR-x: Fat Tails in Financial Risk Management. *Journal of Risk*, 1: 47–61.

Kupiec, P. (1995) Techniques for Verifying the Accuracy of Risk Measurement Models. *Journal of Derivatives*, 2: 173–184.

Jalal, A. and Rockinger, M. (2008) Predicting Tail-related Risk Measures: The Consequences of Using GARCH Filters for Non-GARCH Data. *Journal of Empirical Finance*, 15: 868–877.

McNeil, A.J. and Frey, R. (2000) Estimation of Tail-related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach. *Journal of Empirical Finance*, 7: 271–300.

Neslehova, J., Embrechts, P., and Chavez-Demoulin, V. (2006) Infinite Mean Models and the LDA for Operational Risk, *Journal of Operational Risk*, 1(1): 3–25.

Pickands, J. (1975) Statistical Inference Using Extreme Order Statistics. *Annals of Statistics*, 3: 119–131.

Rachedi, O. and Fantazzini, D. (2009) Multivariate Models for Operational Risk: A Copula Approach Using Extreme Value Theory and Poisson Shock Models. In G. Gregoriou (ed.), *Operational Risk towards Basel III:*

Best Practices and Issues in Modelling, Management and Regulation (pp. 197–216). New York: Wiley.

Reiss, R.-D. and Thomas, M. (1997) *Statistical Analysis of Extreme Values*. Basel, Switzerland: Birkhäuser.

NOTES

1. We remark that for each $0 < \alpha < 1$ we have $st{-}1._{\nu}(\alpha) = -st{-}1_{\nu}(1 - \alpha)$.
2. This model was firstly proposed by Glosten, Jagannathan, and Runkle (1993).
3. The detailed description of the model is presented in McNeil and Frey (2000).
4. If the model is tenable, the residuals should behave like the realizations of the i.i.d. random variables.
5. A detailed description of the model is presented in Gonzalo and Olmo (2004).
6. For more detail about Hill's estimator, see Hill (1975).
7. The detailed description of the model is presented in Huisman, Koedijk, and Pownall (1998).
8. For more detail about Hill's estimator, see Hill (1975).

This page intentionally left blank

PART VI

MODELING MODEL RISK FOR RISK MODELS

This page intentionally left blank

CHAPTER 22

MODEL RISK IN COUNTERPARTY EXPOSURE MODELING

Marcus R. W. Martin

ABSTRACT

In this chapter the model risks inherent to the first two steps of generating counterparty exposure profiles will be discussed. First, future market scenarios have to be generated. Under these scenarios the over-the-counter derivatives portfolio of a counterparty has to be revalued to measure its exposure; different evolution dynamics have to be chosen. Each choice carries, therefore, a certain portion of model risk. The actual revaluation of the derivative positions in the second step usually relies on simplified methods (at least compared with typical market risk and P&L calculations). Hence, another portion of model risk is inherent to the choice of these simplifications needed to keep the task tractable within a reasonable timeframe.

Consequently, the chapter aims at identifying the different sources of model risk and tries to classify and identify potential issues of these risks in counterparty exposure measurement.

INTRODUCTION

Modeling of counterparty exposures is one of the most challenging tasks on the borderline between market and credit risk. Even under mild assumptions on the methods used to produce so-called counterparty exposure profiles, various assumptions and asset dynamics have to be specified, as well as the choice of fast but still adequate pricing models that lead to an important portion of model risk which is essential knowledge for each risk manager.

Counterparty exposure profiles contain the information of the amount of money the financial institution is exposed to a certain counterparty at a certain point in time in the future. If the market value of all outstanding contracts, e.g., over-the-counter (OTC) derivatives or securities financing transactions, is positive, a default of the counterparty at a particular point in time will lead to an unwinding of these positions, producing a loss. Since the exposure is nothing else but the positive market value of the (possibly netted) counterparty portfolio, it depends on the market environment at that particular time. Despite the situation of classical credit exposures (say, loans) a direct modeling of potential future market scenarios is necessary under which all contracts have to be revalued. Thereafter, according to applicable margin agreements and credit support annexes (CSA), the resulting market values have to be aggregated (in the respective netting sets) to produce the overall counterparty exposure at that particular point in time.

Hence, counterparty exposure profiles are generated along three steps in which we follow the same lines as Pykhtin and Zhu (2006); a somewhat refined view is given by Tang and Li (2007) and excellent brand new monographs of Cesari et al. (2009) and Gregory (2009), who also deal with credit valuation adjustments (CVA), i.e., incorporating counterparty risk into the pricing, which is not a subject we deal with in this article. We aim at discussing the model risks inherent to measuring counterparty credit risk for risk management purposes only. In particular, we focus on those model risks that arise in the first two steps in scenario generation and instrument pricing in detail, while the third step on aggregation (which is necessary for collateralized exposures only) will be dealt with in a follow-up publication.

In the first step, future market scenarios have to be generated. Under these scenarios, a counterparty's OTC derivatives portfolio has to be revalued to measure its exposure. Different evolution dynamics for the driving risk factors have to be chosen and each choice carries, therefore, a certain portion of model risk.

Additionally, the actual revaluation of the OTC derivative instrument usually relies on simplified methods (at least compared with typical market risk and P&L calculations). Hence, another portion of model risk is inherent in the choice of these simplifications needed to keep the task tractable within a reasonable timeframe.

Consequently, the chapter aims at identifying the different sources of model risk and attempts to classify and identify potential issues of these risks on counterparty exposure measurement.

BASIC INGREDIENTS OF A CREDIT RISK MEASUREMENT SYSTEM

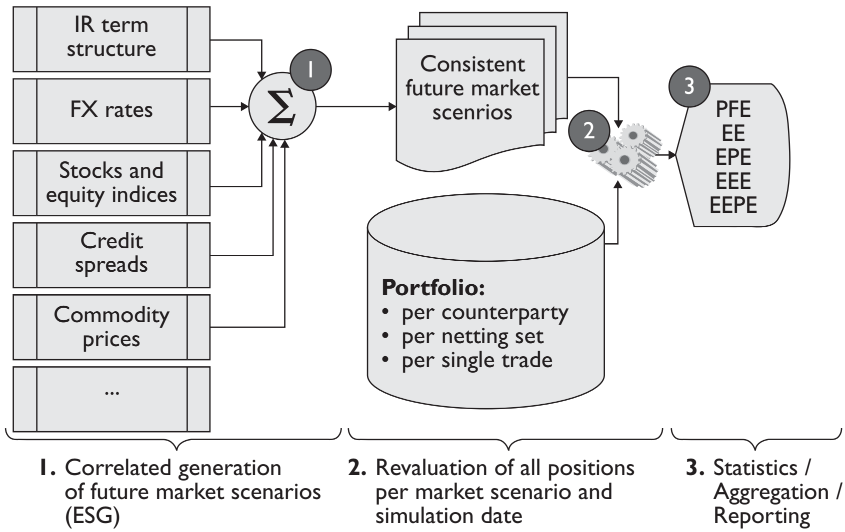
A counterparty credit risk measurement system (CRMS) usually consists of three main components already sketched in the introduction to calculate a total exposure profile per counterparty portfolio.

- First, an economic scenario generator (ESG) is used to produce future realizations of the risk factors relevant to the institution's positions on a certain set of dates defined by a prespecified time grid.
- Second, these risk factors are used to revalue the positions at certain future points in time to calculate the relevant exposure.
- Finally, a collateral model is superposed to come up with an aggregated view on the total exposure of a counterparty portfolio, taking into account any collateral and margin agreements that are in place to reduce the exposure. At this step, all exposures are aggregated and a statistical measure, e.g., the mean of the future distributions of market values (which yields the so-called expected exposure (EE), expected positive exposure (EPE), or effective expected exposure (EEPE)), or a quantile (for the so-called potential future exposure (PFE)[1]) is calculated on portfolio, counterparty, netting set, or single trade level.

Each of these steps (cf. also Figure 22.1) is subject to different aspects of model risk and, therefore, has to be treated separately in the following discussion. For this purpose, we will focus on the first two steps in this article and begin with short descriptions of typical implementations following Pykhtin and Zhu (2006).

In particular, calculation time constraints have to be actively taken into account since, in contrast to a single present value calculation for P&L and market risk purposes, all revaluations have to be performed at any future time step to calculate a present value (PV) at that point in time given the scenario produced by ESG. Hence, complex products can often not or at least not completely be treated by CRMS because such products would require Monte-Carlo-on-Monte-Carlo calculations.

Figure 22.1 Three Steps to Generate Exposure Profiles and Counterparty Measure



For these reasons, model risk inherent to CRMS is by far more complex than in classical market risk models. In order to analyze the different sources of model risk, we start with describing the respective steps in more detail.

Methods to Generate Future Economic Risk Factor Scenarios (Economic Scenario Generation)

Potential market scenarios are generated in all relevant market risk factors at different times in the future on a prespecified time grid of dates. This simulation time grid is usually chosen to have daily and, after a few weeks, weekly or monthly time-steps in the short end while the grid points become sparser in the mid- and long term. The fixed simulation time grid can also be enriched by additional, portfolio- or trade-specific time buckets. By this, risk-relevant cash flows can be incorporated according to the underlying payment schedule(s) because certain cash flows may significantly influence the total balance of the counterparty exposure.

Typically, the evolution of interest rates, FX rates, equity prices, credit spreads, and commodity prices can be generated in various ways. Sometimes, factor models are used, e.g., by modeling equity indexes and using beta regressions for deriving single share prices; while in other cases evolution dynamics similar to those used in models for instrument pricing are used, e.g., by choosing short rate or LIBOR market models to model the evolution of the term-structure of interest rates in the needed currencies (cf., e.g., Reimers and Zerbs (1999), Zerbs (1999), or Hegre (2006) and Beck (2010)).

The dynamics of risk factors are in most cases based on log-normal or mean-reverting stochastic processes governed by Brownian motions evolving under the empirical measure. This modeling approach is in concordance with the usual dynamics specified for risk measurement (e.g., in market risk measurement systems) which are generically built on the real measure based on historical data (and not necessarily constrained to a risk neutral framework; see the section "Methods for Pricing Financial Instruments"). For example, in a front-office pricing model, the interest rate scenarios are usually generated by construction of zero rates or discount factors using the market prices of government bonds or swap rates. However, such construction is in most cases computationally expensive and the forward rates or forward volatility surfaces implied from these may be too insensitive to mimic possible real-world changes over the considered time horizon as a consequence of arbitrage-free constraints.

Thereafter, the dependent evolution of risk factors is usually modeled by correlating the driving stochastic processes governing the single risk factor evolutions. Typically, correlated Brownian motions are based on empirical covariance matrices which were estimated from historical time series for the governing risk factors. This is aimed at ensuring the consistency and reasonableness of (real) market scenarios generated by the different stochastic processes (see later on in this chapter for a deeper discussion of this subject).

Methods for Pricing Financial Instruments

The methods used for pricing financial instruments are in principle the same as for market risk purposes. But due to the problem to revaluate the whole OTC derivative portfolio at a large number of future simulation dates the possibility to run complex and calculation power and time consuming algorithms is rather limited. In contrast to the pricing for P&L purposes, computationally intensive models have to be replaced by analytical approximations or simplified models which might be less accurate but efficient (see also Gregory, 2008).

Using a Brownian bridge technique, the so-called conditional valuation approach by Lomibao and Zhu (2005), even path-dependent exotics (of the first generation, e.g., barrier options, Asian options, or swap-settled swaptions) can be revalued in an efficient manner consistent to the paths generated under the economic scenario generators and the plain vanilla derivatives priced based on these scenarios.

SPECTRA OF MODEL RISKS IN A COUNTERPARTY CREDIT RISK MEASUREMENT SYSTEM

As seen in the previous section we need to distinguish the three fundamental steps in generating exposure profiles before we can measure some kind of total model risk of a CRMS. According to the above description of the measurement process we define:

- *Scenario generation risks* as those model risks (including all aspects of specification, estimation and implementation risks) that occur by specifying the evolution of risk factors and all relevant assumptions on the choice of risk factors and time buckets as well as generating overall consistent future market scenarios.
- *Pricing risks* as those model risks (including all aspects of specification, estimation, and implementation risks) inherent to the specific requirements on instrument model pricing routines that are needed to evaluate the counterparty's positions in a fast but adequate manner.
- *Aggregation risks* as those model risks (including all aspects of specification, estimation, and implementation risks) occurring in the process of aggregating the counterparty's exposures, recognizing any eligible netting agreements, and modeling the margin process including collateral valuations, etc. We will consider these risks in a forthcoming publication in more detail (see Martin, 2009a).

Furthermore, any model risks inherent to bilateral counterparty risk valuation models or credit valuation adjustments (CVA) measurement routines (for an introduction, see the new monographs of Cesari et al. 2009 or Gregory 2009 and the work by Brigo and his collaborators 2008a, 2008b, 2009; Gregory, 2009; or Crépey, Jeanblanc, and Zargari, 2009) will also not be covered by this article. The corresponding model risks might be classified as those model risks inherent to pricing and hedging collateral risk on a single trade level (see Martin, 2009b). Finally, any model risks occurring in economic capital calculations via the so-called alpha factor and due to an inadequate assessment of wrong way risks can not be covered by this article (for an introduction to the alpha factor and wrong way risk issues, see Wilde (2005)).

Sources of Model Risk in Economic Scenario Generation

The overwhelming portion of model risk dominating the scenario generation risk is the specification risk of risk factor evolution dynamics in scenario generation, including the choice of risk factors and adequate future simulation dates, the stochastic dynamics used to model interest and FX rates, stock and index prices, credit spreads, commodity futures, and so on which directly influences the overall level of counterparty risk measure.

For example, using a simple log-normal process for describing the evolution of interest rate or FX rates might lead to unrealistically large or small future rates and, consequently, to an overestimation of counterparty exposures resulting from instruments sensitive to these risk factors. On the other hand, too simplistic modeling approaches might lead to an underestimation. Take the case when a term structure model would only be able to generate future market scenarios which are due to parallel shifts of today's term structure or when an important (class of) risk factor(s) is not even stochastically modeled while being an important contributor to the total counterparty risk.

A possible solution to a reduced way to model future market scenarios can in some cases be achieved by Martingale resampling methods proposed, e.g., by Tang and Li (2007). These techniques could also be of great importance for the generation of realistic and consistent future market scenarios which often carry a further portion of model risk. Given the idealized situation that we have perfectly chosen the evolution dynamics for all relevant risk factors as well as the simulation time grid, one has to define a way to generate a correlated scenario at each future simulation date. To do so, the driving stochastic processes have to be modeled in a correlated fashion, which is an ambitious task because of the diversity of processes needed to describe the risk factor evolutions. The easiest way is to assume a multivariate normal distribution or normal copula to specify this dependency based on covariance matrices derived empirically. Both choices carry an enormous portion of model risk or, more precisely, specification risk (see, e.g., Frey, McNeil, and Nyfeler, 2001).

But the consistency within a simulated future market scenario has to be checked additionally, for example, the well-known link between credit spreads and stock prices implies that a defaulting or close-to-defaulting corporate cannot have a high equity share value while the credit spread indicates the closeness-to-default and vice-versa. Hence, consistency checks have to be established to ensure that the simulated scenario makes sense economically and, therefore, to reduce the specification risks.

Unfortunately, this is still not the full picture of model risks that have to be subsumed to scenario generation risks. When it comes to generating the actual market scenarios over a long time horizon, it is essential to use continuous-time models (of path-dependent simulation or direct-jump to simulation date type as proposed by Pykhtin and Zhu, 2006) or sophisticated discretization schemes to ensure the strong convergence of the sampled scenario paths to the prespecified risk factor distributions (see Kloeden and Platen, 1999, for an overview on different methods). Clearly, these estimation risks will be of second order compared with the overall noise in risk factor evolutions on a long simulation horizon for plain vanilla instruments but might become of higher importance for more complex an exotic deals with long-term maturities.

Sources of Model Risk in Pricing Models

Only to a small extent are the pricing risks similar and of the same spirit as in classical market risk modeling as far as the accuracy of pricing models is considered as an isolated goal (see Bossy et al., 2001; Branger and Schlag, 2004a, 2004b; Courtadon, Hirsa, and Madan, 2002; Gibson et al., 1998; Hull and Suo, 2001; Kerkhof, Melenberg, and Schumacher, 2002; Rebonato, 2005; Derman, 1997; Martin (2005); as well as Ammann and Zimmermann (2002), which is the only study of pricing risk in the context of CRMS). Nevertheless, the counterparty risk modeling specific aspects have to be explicitly discussed since the choice of a certain instrument model for counterparty risk measurement purposes is in almost all cases a trade-off between accuracy and computational efficiency. This topic has never been treated in academic literature before such that even no measures are publicly available supporting the risk manager in her decision. Since the particular decision process has to be taken into account to thoroughly quantify the model risks, we can only provide a rather vague framework on the model risks arising in this context.

In principle, a recalibration of each pricing model (or change to the risk neutral or pricing measure) would be necessary at every future simulation date given the simulated real market scenario (i.e., the market scenario generated under the real or empirical measure). Hence, a simple way to assess the first-order model risk is to compare the market value distributions generated under the empirical or real measure and under the risk neutral measure (instead of performing numerous recalibrations). The use of so-called "stochastic deflator techniques" that enable a pricing of financial instruments under the empirical measure is a relatively new development in actuarial mathematics and has to be further investigated for these purposes as well.

Another question refers to the consistency of dynamics specified in ESG for the universe of tradable plain vanilla instruments and those stochastic evolution models used in pricing models which we can easily demonstrate by an example as follows: assume a jump-diffusion dynamics is deemed necessary for the stochastic evolution of a certain share. On the other hand, when it comes to pricing an option on this share a Black-Scholes–type formula is used that assumes a complete market and a log-normal distribution of the underlying asset. Hence, the evolutionary dynamics of the option pricing model and the simulated asset distribution of the underlying are not consistent, which might lead to over- or underestimation of risks in one or the other direction (depending on the financial institution's positions). In this particular situation even the elsewhere very helpful Martingale resampling techniques by Tang and Li (2007) won't apply such that at least an empirical study on a set of benchmark trades has to be performed to get a better idea on whether the resulting effects have a conservative or progressive impact on the counterparty exposure.

Aggregation of Model Risks

Overall, the two steps described above describe the model risks occurring in portfolios with uncollateralized trades only. Even in this simplified case the aggregation of scenario generation risks and pricing risks to an overall model risk is extremely difficult to obtain. A rather general framework to assessing the model risk of (general) pricing models was given by Cont (2006). In this section, we aim at adjusting these requirements for a measure of model uncertainty (which is equivalent to our term "model risk" used throughout this chapter) for our purposes.

1. For liquidly traded plain vanilla instruments whose driving risk factors are modeled stochastically in the CRMS, the price is determined by the market within a bid-ask spread, i.e., there is no (market risk) model uncertainty on the value of a liquid instrument which is a basic ingredient of simulation of the future market scenarios. The modeling exercise therefore should emphasize the importance of these instruments and adequacy of the corresponding risk factor evolutions which should be tested in-depth historically (in-sample and out-of-sample).
2. Any measure of model uncertainty for CRMS must take into account the possibility of setting up (total or partial) hedging strategies in a model-free way. If an instrument can be replicated in a model-free way, then its value involves no model uncertainty; if it can be partially

hedged in a model-free way, this should also reduce the model uncertainty on its value. Hence, the CRMS should be designed to recognize these possibilities in order to reduce the computational burden and the pricing model risks as well.

3. When some instruments (typically, European-style call or put options for short maturities and strikes near the money) are available as liquid instruments on the market, they can be used as hedging instruments for more complex derivatives, but also carry important information on the risk-neutral counterparty exposure of the underlying (as their market value is a measure for the counterparty exposure of the plain vanilla underlying for the time to maturity of the option). This information can be used to challenge the corresponding risk factor evolution (in particular concerning the short-term, at-the-money, volatility assumptions for the driving risk factor(s)). This procedure will ensure a consistent short-term prognosis given today's information which is essential for prudent risk management purposes.
4. If one intends to compare model uncertainty with other, more common, measures of counterparty exposure risk of a portfolio, the model uncertainty on the value of a portfolio today as well as on a future simulation date should be expressed in monetary units and normalized to make it comparable to today's market value and the future exposure measure of the portfolio.
5. The consistency, mathematically and economically, of the generated scenarios and techniques applied to revalue the counterparty portfolio, netting sets, or single trades has to be ensured by additional manually or automatically performed checks of the simulated future market scenarios. This includes any estimation techniques for assessing discrepancies in the dynamics of risk factor evolutions and pricing models.

Given these guidelines a basic framework might be set up for studying the model risk of counterparty risk measurement systems following broadly the lines of Cont (2006). A thorough study of these ideas has to be carried out with supporting empirical studies which go beyond the scope of this chapter.

Nevertheless, the model risk measurement process for counterparty exposure measurement systems has to be embedded into the overall risk management and risk controlling processes that are of vital importance for the internal validation process of CRMS. Certain additional processes should also be incorporated into the daily market and counterparty risk management and measurement; e.g., one can compare the market values of all instruments that were delivered on one hand by daily P&L calculation processes with those market values obtained from (possibly simplified) pricing models in the CRMS. These differences might be used to set up deviation limits which might trigger an in-depth analysis of a single-pricing routine or the CRMS as a whole.

In market risk measurement systems the well-established backtesting processes and methodologies usually also provides insights into the performance of pricing models and the overall measurement system. Up to now there seems to be no comparable standard for backtesting exposure measurement systems which, again, are harder to assess than market risk measurement models (in particular due to the extremely long forecast horizon, which in turn means that the influence of correlations and auto-correlations must not be neglected).

CONCLUSION

This chapter aimed at providing a first assessment of possible approaches to measure and manage the model risks of CRMS. In fact, it provides a formal but practical classification structure for identifying the sources of model risks according to the process of generating future market value distributions and exposure measures.

This also points to several fields of potential future research. Currently there is no aggregation scheme known for measuring the overall effect of the different sources of model risks. Furthermore, empirical studies should be undertaken to tackle the distinct sources in more detail; e.g., the use of analytical approximations in pricing of more complex OTC derivatives is computationally extremely attractive such that there is a trade-off between pricing accuracy, computational efficiency, and the resulting model risk which has to be analyzed carefully. Finally, the question of adequately modeling collateral processes adds an additional layer of complexity and hence model risks which will be analyzed and classified in a forthcoming research work.

REFERENCES

Ammann, M. and Zimmermann, H. (2002) The Credit Model Risk of Interest Rate Derivatives and Regulatory Implications. Working paper, Swiss Institute of Banking and Finance.

Beck, T. (2010) Interest Rate Counterparty Credit Risk Modeling. Forthcoming Master's Thesis, University of Applied Sciences, Darmstadt, Germany.

Bossy, M., Gibson, R., L'Habitant, F.-S., Pistre, N., and Talay, D. (2001) A Methodology to Analyse Model Risk with an Application to Discount Bond Options in a Heath-Jarrow-Morton Framework. Working paper, University of Zürich, Switzerland.

Branger, N. and Schlag, C. (2004a) Model Risk: A Conceptual Framework for Risk Measurement and Hedging. Working paper, Goethe-Universität, Frankfurt, Germany.

Branger, N. and Schlag, C. (2004b) Can Tests Based on Option Hedging Errors Correctly Identify Volatility Risk Premia? Working paper, Goethe-Universität, Frankfurt, Germany.

Brigo, D. and Capponi, A. (2009) Bilateral Counterparty Risk Valuation with Stochastic Dynamical Models and Application to Credit Default Swaps. Working paper, Imperial College, Department of Mathematics, London, and California Institute of Technology, Division of Engineering and Applied Sciences, Pasadena.

Brigo, D. and Pallavicini, A. (2008a) Counterparty Risk and Contingent CDS Valuation under Correlation between Interest-Rates and Defaults. *RISK*: April: 39–44.

Brigo, D. and Chourdakis, K. (2008b) Counterparty Risk Valuation for Energy-Commodities Swaps—Impact of Volatilities and Correlation. Quantitative Research Special Report, Fitch Solutions, London, UK. Accepted for publication in the *International Journal of Theoretical and Applied Finance*.

Cesari, G., Aquilina, J., Charpillon, N., Filipovic, Z., Lee, G., and Manda, I. (2009) *Modelling, Pricing and Hedging Counterparty Credit Exposure—A Technical Guide*. London: Springer.

Cont, R. (2006) Model Uncertainty and Its Impact on the Pricing of Derivatives Instruments. *Mathematical Finance*, 16(3): 519–547.

Courtadon, G., Hirsa, A., and Madan, D. (2002) The Effect of Model Risk on the Valuation of Barrier Options. Working paper, University of Maryland.

Crépey, S., Jeanblanc, M., and Zargari, B. (2009) Counterparty Risk on a CDS in a Markov Chain Copula Model with Joint Defaults. Working paper, Université d'Évry Val d'Essone, Évry Cedex, France.

Derman, E. (1997) *Model Risk*. New York: Goldman Sachs Publications.

Frey, R., McNeil, A.J., and Nyfeler, M. (2001) Copulas and Credit Models. *RISK*, October: 111–114.

Gibson, R., L'Habitant, F.-S., Pistre, N., and Talay, D. (1998) Interest Rate Model Risk: What Are We Talking About? Working paper, RiskLab, Zürich.

Gregory, J. (2008) Credit Risk Models and the Sub-Prime Crisis: Rating Agency Models, Structured Credit and CDO Pricing. PRMIA London Credit Risk Symposium.

Gregory, J. (2009) Being Two-faced over Counterparty Credit Risk. *RISK*, February: 86–90.

Gregory, J. (2009) Counterparty credit risk—The new challenge for global financial markets. London: Wiley Finance.

Hegre, H. (2006) Interest Rate Modeling with Applications to Counterparty Risk. Master's thesis, Norwegian University of Science and Technology.

Hull, J. and Suo, W. (2001) A Methodology for Assessing Model Risk and Its Application to the Implied Volatility Function Model. *Journal of Financial and Quantitative Analysis*, 37(2): 297–318.

Kloeden, P.E. and Platen, E. (1999) Numerical Solution of Stochastic Differential Equations. Applications of Mathematics. *Stochastic Modelling and Applied Probability No. 23*, 3rd ed. New York: Springer.

Kerkhof, J., Melenberg, B., and Schumacher, H. (2002) Model Risk and Regulatory Capital. Working paper, Tilburg University, The Netherlands.

Lomibao, D. and Zhu, S. (2005) A Conditional Valuation Approach for Path-Dependent Instruments. Working paper, Capital Market Risk Management, Bank of America, New York.

Martin, M.R.W. (2005) Prüfung der Validierung von Instrumentmodellen. In: *Handbuch Bankgeschäftliche Prüfungen*. Stuttgart, Germany: Schaeffer-Pöschel-Verlag.

Martin, M.R.W. (2009a) Model Risks According to Modeling the Aggregation. Forthcoming Working paper, University of Applied Sciences, Darmstadt, Germany.

Martin, M.R.W. (2009b) Model Risks Inherent to Bilateral Counterparty Risk Valuation and CVA. Forthcoming Working paper, University of Applied Sciences, Darmstadt, Germany.

Pykhtin, M. (ed.). (2005) *Counterparty Credit Risk Modeling*. London: RISK Books.

Pykhtin, M. and Zhu, S. (2006) Measuring Counterparty Credit Risk for Trading Products under Basel II. Working paper, Risk Architecture, Bank of America, New York.

Rebonato, R. (2005) Theory and Practice of Model Risk Management. Quantitative Research Centre (QUARC) of the Royal Bank of Scotland. Oxford Financial Research Centre—Oxford University.

Reimers, M., Zerbs, M. (1999) A Multi-factor Statistical Model for Interest Rates. *ALGO Research Quarterly*, 2(3): 53–64.

Tang, Y. and Li, B. (2007) *Quantitative Analysis, Derivatives Modeling, and Trading Strategies—in the Presence of Counterparty Credit Risk for the Fixed Income Markets.* Hackensack, NJ: World Scientific.

Wilde, T. (2005) Analytic Methods for Portfolio Counterparty Credit Risk. In M. Pykhtin (ed.), *Counterparty Credit Risk Modeling*. London: RISK Books.

Zerbs, M. (1999) Mark-to-Future in Practice. *ALGO Research Quarterly*, 2(2): 15–18.

NOTES

1. For a discussion and concise definitions of these counterparty exposure measures, see Pykhtin and Zhu (2006) or Pykhtin (2005).

CHAPTER 23

MODEL RISK IN CREDIT PORTFOLIO MODELING

Matthias Gehrke and Jeffrey Heidemann

ABSTRACT

The loss distribution of a credit portfolio depends on default probabilities, exposures, and recovery rates. Moreover, certain background factors impact significant parts of the portfolio. The dependence of loans to those factors is often modeled in a Gaussian framework. For homogeneous portfolios with Gaussian background factors, Vasicek (1991 and 2002) provided an analytic formula for the overall loss distribution. His formula turned out to be useful for many practical situations. Starting from that, we shall see that main input factors are not well known in a financial institution's database.

In this chapter, we will study the impact of misjudging the model's parameters on the risk estimate. Each important parameter will be examined separately. Furthermore, we try to identify possible indicators of wrong estimates.

INTRODUCTION

In today's credit portfolio models, portfolio behavior depends on default probabilities, exposures, and recovery rates. Moreover, certain background factors (macroeconomic factors) have an impact on significant parts of the portfolio. The dependence of loan defaults to those factors is often modeled using a Gaussian framework. For homogeneous portfolios, Vasicek (1991 and 2002) provided an analytic formula for the overall loss distribution. His formula turned out to be useful for many practical situations. It has even been made the heart of the Basel II accord for regulatory capital. Starting from this formula, we shall see that the main input factors are not well known in a financial institution's database.

In this chapter, we will study the impact of misjudging the model's parameters on the risk estimate. Each important parameter will be examined separately. Furthermore, we will try to identify possible indicators of wrong estimates.

VASICEK'S ONE-FACTOR MODEL

The underlying idea is a large homogenous portfolio. Every obligor has an ability-to-pay or "asset value" which can be decomposed into a systematic and specific part.[1]

$$X_i = G\sqrt{\rho_i} + Z_i\sqrt{1-\rho_i} \tag{23.1}$$

G and all of Z_i are assumed to be independent standard normal distributed random variables. The only source of dependence between two obligors is via their respective exposures to the global factor G, which can be interpreted as an economic factor.

An obligor i defaults, if X_i falls below a default threshold c_i. This makes i's default probability equal to $\Phi(c_i)$.

In a large homogenous portfolio, all exposures and all default probabilities are equal, so are all the asset correlations ρ_i. The number of obligors is assumed to be large. In fact, the following approximations are derived for n approaching infinity.

Vasicek (2002) showed that under these circumstances, assumptions for the overall portfolio loss (called L) has the following distribution function:[2]

$$P(L \le x) = \Phi\left(\frac{\sqrt{1-\rho}\,\Phi^{-1}(x) - \Phi^{-1}(p)}{\sqrt{\rho}}\right) \tag{23.2}$$

which we shall call the "Vasicek-distribution." It has two parameters: the default probability p and asset correlation ρ.

Vasicek (2002) suggested using $F_{(V,p,\rho)}$ to fit loan portfolio loss distributions with more complex properties (like heterogeneous default probabilities, positive recoveries, etc.) or those which were obtained differently, e.g., by multifactor models.

For this reason, we frequently utilize moment-matching. The expected loss and the variance of a loan portfolio are calculated analytically or by a small number of simulations.[3] The relations[4]

$$E(L) = p \tag{23.3}$$

$$Var(L) = \Phi_2\,(\Phi^{-1}(p), \Phi^{-1}(p), \rho) - p^2 \tag{23.4}$$

(where Φ_2 denotes the cumulative bivariate Gaussian distribution function) can then be used to calculate the appropriate p and ρ for the best-matching Vasicek distribution.[5]

The α-quantile of this distribution is given by[6]

$$L_\alpha = F_{v,p,\rho}\,(\alpha, 1 - p, 1 - \rho) \tag{23.5}$$

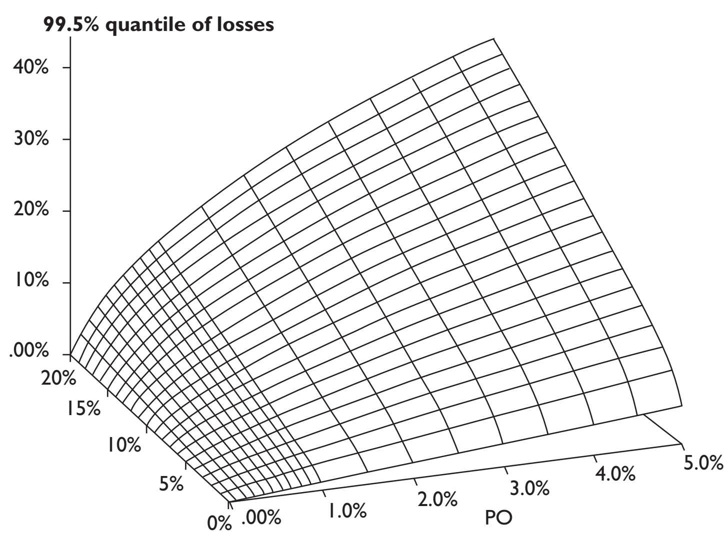
which is important for risk measurement, since credit risk is regularly measured with quantile-related numbers such as value at risk or expected shortfall[7] (see Figure 23.1).

Wehrspohn (2003) showed that Vasicek's model is extendible in a straightforward way to portfolios consisting of several significant subportfolios with different probability of default or asset correlation.

In this chapter we shall focus on the modeling of

- Probabilities of default
- Asset correlations
- Distribution assumptions

Figure 23.1 Loss Quantiles of Formula (23.5) (z axis) as a Function of Probability of Default and ρ



For each of these influencing factors, we shall ask ourselves about the potential impact of possible estimation errors, a realistic range in which the estimates may vary and a way to assess the quality of our estimates or assumptions.

PROBABILITIES OF DEFAULT

The most fundamental input to credit risk models is the probability of default (PD). It is usually estimated using the bank's own loss database combined with publicly available default information, such as rating agencies' reports. By its nature, a debtor's PD is an estimate.

The relative size of confidence sets for default probabilities depends on the number of available observations and the probability itself. As a rule of thumb, smaller PDs are harder to estimate.

Trück and Rachev (2005) examined the influence of the number of observations on PD confidence sets and illustrated how much of the uncertainty surrounding AAA and AA PDs is due to the rareness of such corporates.

The size of confidence sets for PDs might be only a few basis points in "A" grades, but compared to the PD, it is of high impact. Either the relative coefficient of variation $\frac{\sigma}{\mu}$ or the upper confidence bound compared with the estimated PD should be good measures of uncertainty. The higher the ratios, the greater the relative impact of PD underestimation.

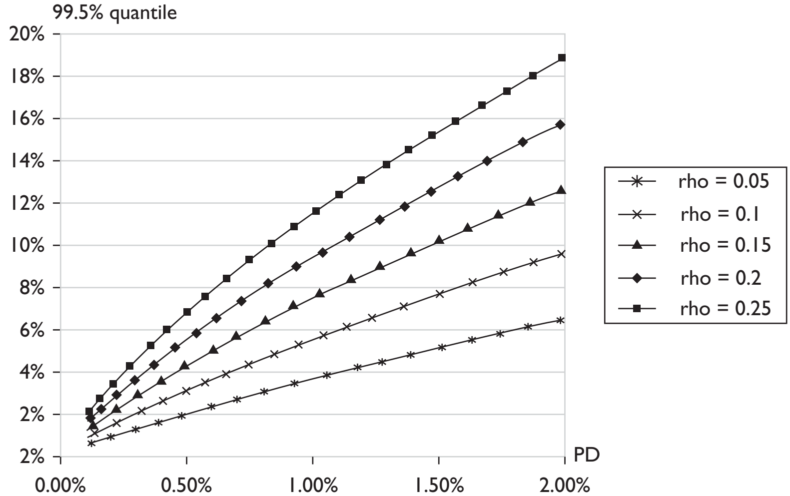
An overview of Trück and Rachev's findings is given in Table 23.1.

The impact of misestimating the PD on the quantiles of Vasicek's distribution is less than proportional in a typical range of PD and ρ (Figure 23.2).

Table 23.1 Uncertainty in Probabilities of Default Estimates According to Trück and Rachev[8]

Rating Grade	PD Mean (bps)	$\frac{\sigma}{\mu}$	Upper Confidence Bound to Mean PD Ratio
Aaa	1	1.6	7.0
Aa	6	1.6	6.7
A	19	0.9	3.0
Baa	49	0.6	1.9
Ba	126	0.4	2.0
B	262	0.2	1.6
Caa	473	0.3	1.3

Figure 23.2 In a Realistic Range of Probabilities of Default, the Impact of PD Changes Is Less Than Proportional



A major driver of model risk in PD estimation is the relation between the upper confidence interval bound and the mean estimate.

Hanson and Schuermann (2005) used four different methods to derive PD confidence intervals from S&P rating data. For investment grades, so-called duration-based confidence intervals obtained by bootstrapping seem to be much tighter than those calculated using other methodologies. Duration-based PD estimates take into account intrayear rating migrations that are neglected by cohort-based ones.[9] Unfortunately, cohort-based estimation is a common practice in calibrating internal rating systems.

For "A" grades, the upper bounds are well below the 3 bps minimum floor introduced by the Basel II rules. Hanson and Schuermann find upper confidence bounds to be roughly twice as high as the PD estimate. However, this ratio can be worse for noninvestment grade ratings. For those grades, other estimation methods may be preferable.

ASSET CORRELATIONS

Cassart, Castro, and Langendries (2007) used both a default-driven and a rating migration–driven technique to derive asset correlations and confidence bounds. The analysis was separated by industry groups and a special focus was placed on intraindustry versus interindustry correlations. For almost half the sectors, correlations were not significantly different from zero, while in other industry sectors they were significant. They also stated that the default-driven approach may not be practicable in good rating grades where defaults are rare. The upper bound for intraindustry asset correlations topped at 0.21 (for farming/agriculture) and mean estimates of 0.13, while in other sectors it was well below this value. This looks like further indication for an uncertainty multiple of 2, with higher values possible if average correlations are lower.

Akhavein, Kocagil, and Neugebauer (2005)[10] compared six different approaches for estimating asset correlations. Interindustry asset correlation was found to average between 0.1444 and 0.2092, depending on the model used. Within the industry group, correlations ranged between 0.1973 and 0.2649. However, a rating transition–based model produced values that were much lower than all other methods. The estimates were 0.0456 (intraindustry) and 0.0785 (intraindustry). This allows for an uncertainty of factor 3 or more when only one model is used.

For the small business sector, Schwarz (2006) used PD variance and bootstrapping techniques to calculate asset correlations for Austrian small- and medium-sized enterprises (Table 23.2). His analysis resulted in correlations around 2 percent, while confidence bounds topped at 6 percent.

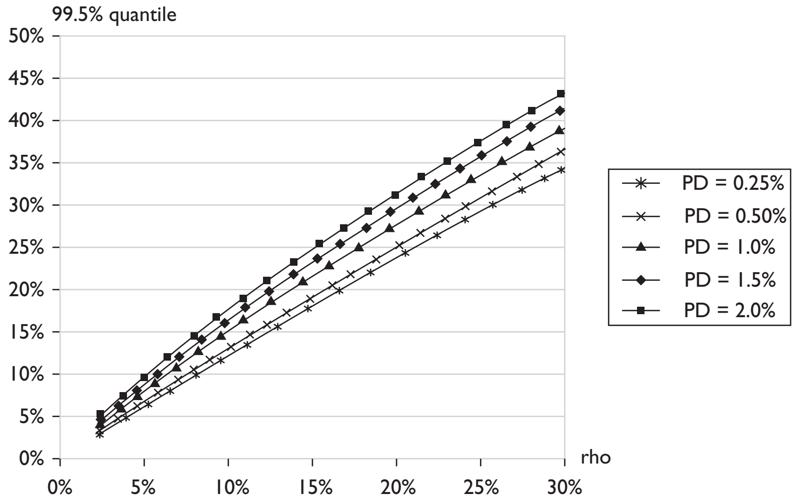
On the one hand, an uncertainty allowing for three times higher correlations can be embarrassing (even though the impact on loss distribution quantiles might be less than factor 3; see figure 23.3). On the other hand, at least it looks like the regulatory assumptions concerning asset correlations look pretty conservative.

It would be desirable to see an analysis on the difference between correlations in normal and economically difficult times to assess "conditional" asset ρ's that are likely to drive loan portfolio risks.

Table 23.2 Small- and Medium-sized Enterprise (SME) Asset Correlations in Austria (Schwarz, 2006)

Size Class	Turnover (Euro)	Mean	95% Confidence Interval	Upper Bound/Mean
Small SME	< €1 m	0.0183	0.0046–0.0414	2.3
Medium SME	< €5 m	0.0259	0.0064–0.0442	1.7
Large SME	< €50 m	0.0216	0.0015–0.0646	3.0
Total		0.0160	0.0094–0.0246	

Figure 23.3 Impact of Asset Correlation Changes Is Less Than Proportional



DISTRIBUTIONAL ASSUMPTIONS

A popular candidate to replace the normal distribution in Equation (23.1) is the t distribution. If X is standard normally distributed and follows a X^2 distribution with ν degrees of freedom, then

$$T = \frac{X}{\sqrt{\frac{C_\nu}{\nu}}} \tag{23.6}$$

is t (or student) distributed with ν degrees of freedom.

For large ν, this distribution looks very much like the normal distribution, but for smaller values of ν, leptokurtic behavior becomes apparent and the tails are fatter.

Generating t-distributed, pseudo-random numbers is as easy or difficult as standard normal-distributed ones.

Bluhm et al.[11] made a simulation study replacing the Gaussian-distributed risk factor in Equation (23.1) with a t-distributed one with $\nu = 10$. In a portfolio of low asset correlation and low PD (0.5%), the quantile was about 3.5-times as high as in the Gaussian world. An even higher impact on the more extreme quantiles was observed. However, the effect is more moderate when examining higher PDs and ρ's.

BACKTESTING

If the Vasicek assumptions hold, default rates should be Vasicek distributed. If the true ρ is unknown, it may be derived using the moment matching approach mentioned above.

The Kolmogorov-Smirnov test can be an approach to test the distribution. We shall give an example using a default rate time series from Moody's corporate ratings.[12] We used two time series from 1983 to 2007 for Baa and Ba letter rating, respectively (Table 23.3).

If one is worried about underestimation of the distribution's tails, however, much more data would be required. Differences between the "real" and assumed distribution behind the 95% quantile, for example, can only be found with several observation in this range. This would imply some multiple of 20 data points.

To get around this, it would be helpful to collect default statistics more often than yearly. Another promising approach would be the use of market-based information.

CONCLUSION

One of the main challenges in credit portfolio model backtesting is the lack of data. Historical time series are short and usually have one data point per year. Due to the fast developments in academic research and practice, there may be many structural breaks in internal time series. This leads to wide confidence sets for key parameters.

To overcome these shortcomings, every opportunity to gain more data per year should be examined. Concerning correlation, the use of equity correlations to estimate asset or default correlations in structural models is a huge

Table 23.3 Statistics for Moody's Default Rates 1983–2007 (Emery, 2008): Empirical Default Rate Distribution Is Compared with Vasicek Distribution

	Baa	Ba
Maximum	1.36%	4.78%
Median	0.00%	0.85%
Average	0.19%	1.16%
Standard Deviation	0.37%	1.18%
Implied Rho	17.75%	11.52%
Kolmogorov's D	0.053	0.116
Critical Value	0.264	0.264

factor (one gets 250 points instead of 1 point). However, this acceleration is a new source of model risk in itself. Market data is often exponentially averaged, so the last business downturn information vanishes over time.

Other possible sources of daily or weekly market data that may be used in future are credit default swaps and bond spreads or even co-movements in rating up- and downgrades.

It would be valuable to have conditional correlation information for weak economic scenarios such as the dotcom or subprime crises. This could offer benchmarking opportunities for the estimated loss distribution tail.

REFERENCES

Akhavein, J., Kocagil, A., and Neugebauer, M. (2005) A Comparative Study of Asset Correlations. *Fitch Ratings Special Report*.

Bluhm, C., Overbeck, L., and Wagner, C. (2003) *An Introduction to Credit Risk Modeling*. Boca Raton, Florida: Chapman & Hall.

Cassart, D., Castro, C., Langendries, R., Alderweireld, T. (2007) Confidence Sets for Asset Correlation. Working paper, www.defaultrisk.com.

Emery, K., Ou, S., and Tennant, J. (2008) Corporate Default and Recovery Rates 1920–2007. *Moody's Special Comment*.

Hanson, S. and Schuermann, T. (2005) Confidence Intervals for Probabilities of Default. Working paper, Federal Reserve Bank, Washington, DC.

Schwarz, R. (2006) Default Correlation: An Empirical Approach. Working Paper, BFI Vienna, Austria.

Smithson, C. (2003) *Credit Portfolio Management*. Hoboken, NJ: Wiley.

Trück, S. and Rachev, S. (2005) Credit Portfolio Risk and PD Confidence Sets through the Business Cycle. *Journal of Credit Risk*, 1(4): 61–88.

Vasicek, O. (1991) Limiting Loan Loss Distribution. Working paper, KMV Corporation. Available at www.moodyskmv.com.

Vasicek, O. (2002) Loan Portfolio Value. *Risk*, December: 160–162.

Wehrspohn, U. (2003) Analytic Loss Distributions of Heterogeneous Portfolios in the Asset Value Credit Risk Model. *RiskNet*, available at www.risknet.de.

NOTES

1. See Bluhm, Overbeck, and Wagner (2003, p. 83f).
2. See Bluhm et al. (2003, p. 84f) or Vasicek (2002).
3. Note that estimating a variance is statistically much easier than estimating extreme quantiles.
4. Vasicek (2002).

5. ρ has to be calculated numerically.
6. Vasicek (2002).
7. Smithson (2003, pp. 161, 266).
8. Trück and Rachev (2005, p. 26).
9. See Hanson and Schuermann (2005, p. 4f).
10. Akhavein et al. (2005, p. 14).
11. Blum, Overbeck, and Wagner (2003, Table 2.6, pp. 110, 105f).
12. Emery, Ou, Tennant (2008, p. 22).

CHAPTER 24

MODEL RISK IN CREDIT PORTFOLIO MODELS

Merton versus CreditRisk$^+$ Models

Anne Kleppe and Christian Oehler

ABSTRACT

Throughout the banking industry two major modeling approaches are widely used to estimate internal economic capital figures in credit portfolio models. These two different modeling approaches are either Merton-based factor models or analytically tractable approaches as provided by CreditRisk$^+$ (CR$^+$) frameworks.

Along with economic capital (EC) at the group portfolio level, reliable contributory economic capital (CEC) figures are essential in any aspect of EC-based bank steering. In order to enter these figures into EC-based capital steering and capital management applications, it is crucial that these figures are robust, accurate, and risk sensitive.

In practice, model-related uncertainties between Merton and CR$^+$ frameworks may lead to inappropriate CEC estimations; ultimately providing incorrect steering signals. One particularly important source of model-related uncertainty concerns the differing approaches to calibration of model-specific risk parameters, especially those parameters which drive default correlations and risk concentrations. This article focuses on the modeling risk related to parameter estimation and appropriate methods to resolving these uncertainties. All analyses presented in this chapter are based on data samples that show the typical characteristics of a universal bank portfolio.

INTRODUCTION

Parameterized Merton Model

The implied parameters of the Merton model are given by

- Counterparty-specific asset correlations R^2
- Factor model with its underlying covariance matrix Σ
- Counterparty-specific factor weights $\omega_{i1}, \ldots, \omega_{im}$, $\Sigma_{k=1}^{m}\omega_{ik} = 1$

The asset correlations are the most crucial parameters of the Merton model; these can be estimated by regression analysis between firms' equity return values and the returns of the model for each factor. For nonpublicly listed companies, a typical approach to estimating the asset correlation is by mapping the company to asset correlations deduced from the public company sector via its size. For retail segments practitioners often take conservative, best-practice values or try to estimate R^2 from historical default rates and rating-migration data. Some banks use the asset correlations given by the Basel II regulatory rules; however, many practitioners of Merton modeling frameworks consider the Basel II asset correlations too small on average and replace them with internally measured figures.

Standard CreditRisk$^+$ Parameterization

If CreditRisk$^+$ (CR$^+$) is implemented in its standard form, the following model-implied parameters must be provided for each counterparty-specific transaction:

- *Sector weights*: $\omega_{i1},\ldots,\omega_{im}$, $\Sigma_{k=1}^{m}\omega_{ik} = 1$. Here m denotes the number of sectors (both specific and systematic) within the model for each transaction i.
- *Volatility of default rates*: The volatilities of individual default rates are inputs for the standard CR$^+$ parameterization; these individual volatilities are subsequently converted into sector default rate volatilities. This conversion applies the following formula (Avesani et al., 2006):

$$\sigma_k = \frac{\sum_{n=1}^{N} \sigma_n \omega_{nk}}{\sum_{n=1}^{N} p_n \omega_{nk}}$$

Here, σ_n, $n = 1,\ldots, N$ denotes the default rate volatility of individual obligors and σ_k, $k = 1,\ldots K$ denotes the default rate volatility of sector k. Individual default probabilities of the obligors are denoted by p_n.

For the intersector correlation, the corresponding intersector covariance $\hat{\sigma}^2$ must be specified as an input parameter. The multisector CR$^+$ model is typically referred to as the compound gamma model. If sector correlations are incorporated into the model the defining quantity of the compound gamma model replacing σ_k is given by β_k, which is defined as the difference between the sector variance σ_k^2 and the constant intersector covariance (see Giese, 2003). In this case, the following formula is applied in order to calculate each β_k: $\beta_k = \sigma_k^2 - \hat{\sigma}^2$.

Parameterization of CreditRisk$^+$ via Merton

The parameterization routine outlined in this chapter differs depending on whether a single sector or a multisector CR$^+$ model is desired ("multisector" in the sense that different obligors can be mapped to different sectors). In this chapter we propose a new approach to calibrate the parameters specific to CR$^+$ via the Merton parameterization with the aim to achieve consistency of the capital figures between both models. Within both the single-sector case and the multisector case the normalized sector weights for the systematic and idiosyncratic sectors are deduced from a given Merton-model parameterization. Along with the sector weights the sector volatilities are also given as an output of the re-parameterization routine. If a multifactor model has nonzero correlation between factors, the constant intersector covariance has to be calibrated from the Merton model factor correlation. The following describes the calibration routines in more detail.

The calibration routine of CR$^+$ specific parameters via the Merton parameterization is an extended and modified version of a parameterization routine described in Gordy (2000). It is based on the idea of "moment matching." It seeks to match both the expected conditional default rates and the respective volatilities of obligors. Given the Merton parameters, the corresponding CR$^+$ parameters are deduced. For this purpose the sectors of the CR$^+$ model are identified with the factors of the Merton model. These sector variables will be denoted by x_k. The following two conditions are imposed on the conditional default probabilities of the Merton model and CR$^+$ model, respectively:

- **Condition (1):** $\mathrm{E}(p_i^{CR^+}(x_k)) = \mathrm{E}(p_i^{Merton}(x_k)) \equiv \bar{p}_i$
- **Condition (2):** $\mathrm{VAR}(p_i^{CR^+}(x_k)) = \mathrm{VAR}(p_i^{Merton}(x_k))$

The first condition is obviously fulfilled by choosing the same obligor specific probability of default (PD) as input data for both models. Therefore, only the second condition requires further investigation. For the conditional PD, we find the following expression within the Merton model:

$$p_i^{Merton}(x_k) = N\left(\frac{N^{-1}(\overline{p}_i) - R_i x_k}{\sqrt{1 - R_i^2}}\right).$$

Assuming that any counterparty is mapped to a single systematic sector only, the following expression holds for the conditional PD within the CR^+ framework:

$$p_i^{CR^+}(x_k) = \overline{p}_i(w_i x_k + (1 - w_i)).$$

Here R_i^2 denotes the asset correlations implied by the Merton model for counterparty i, while w_i denotes the sector weight of counterparty i on the systematic sector x_k.

From these two expressions we find that the variance of the conditional default probabilities for the Merton and CR^+ models are respectively:

$$\mathrm{VAR}\left(p_i^{Merton}(x_k)\right) = N_2\left(N^{-1}(\overline{p}_i), N^{-1}(\overline{p}_i), R_i^2\right) - \overline{p}_i^2$$
$$\mathrm{VAR}\left(p_i^{CR^+}(x_k)\right) = \mathrm{VAR}\left(\overline{p}_i w_i x_k\right) = \overline{p}_i^2 w_i^2 \mathrm{VAR}\left(x_k\right) \equiv \left(\overline{p}_i w_i \sigma_k\right)^2$$

Hence, the following condition is deduced for the CR^+ parameters w_i and σ_k:

$$\text{Calibration equation: } (\overline{p}_i w_i \sigma_k)^2 = N_2(N^{-1}(\overline{p}_i), R_i^2) - \overline{p}_i^2.$$

Now the following problem arises: based on the above condition one could determine the sector weights for every obligor by assuming some value for the sector volatilities σ_k. However, there is no additional information for the choice of these sector volatilities. In order to proceed it is suggested in Gordy (2000) to set all sector volatilities equal to 1, which is also the choice within the original CR^+ implementation. With this choice, some of the sector weights w_i will exceed 1, leading to negative weights on the specific risk sector. Technically, negative weights can be tolerated by CR^+, as long as the number of affected obligors is small and the negative weights are relatively small in magnitude. However, with this choice of sigma, the resulting negative weights will lead to invalid loss distributions within many realistic bank portfolios. As such, the following iterative procedure is applied in order to lessen the extend of negative weights:

Sector volatilities σ_k are set equal to 1 and for each obligor i, the sector weights are then computed according to the following formula:

$$w^2\Big|_{\text{obligor } i} = \frac{1}{\bar{p}_i^2}\left(N_2\left(N^{-1}(\bar{p}_i), N^{-1}(\bar{p}_i), R_i^2\right) - \bar{p}_i^2\right).$$

For all weights $w_i > 1$ the corresponding sector volatilities σ_k are then calculated under the assumption $w_i = 1$, according to the following equation:

$$\sigma_k^2\Big|_{\text{obligor } i} = \frac{1}{\bar{p}_i^2}\left(N_2\left(N^{-1}(\bar{p}_i), N^{-1}(\bar{p}_i), R_i^2\right) - \bar{p}_i^2\right).$$

All other volatilities are left as $\sigma_k = 1$. At this stage different obligors in sector k might have different corresponding sector volatilities σ_k, which is clearly unwelcome. The final sector volatility will therefore be fixed as the arithmetic mean of calculated sector volatilities of all obligors within a sector:

$$\sigma_k^2 = \frac{1}{N}\sum_{i=1}^{N} \sigma_k^2\Big|_{\text{obligor } i}.$$

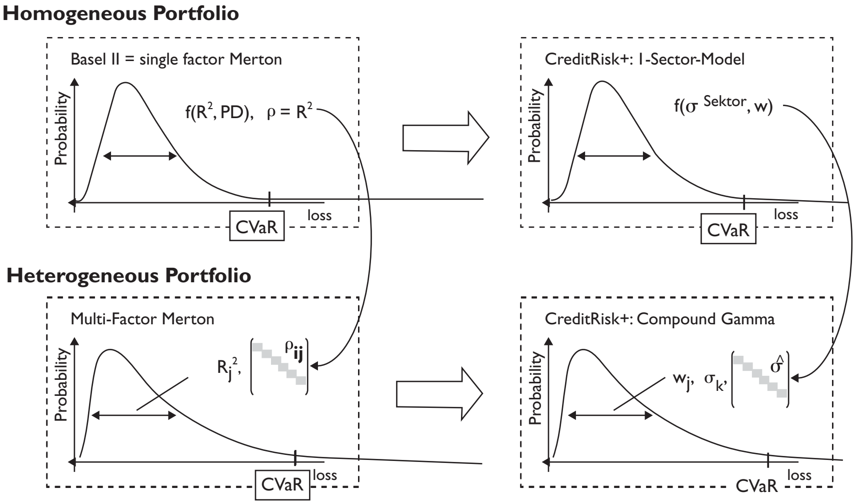
These sector volatilities are then reinserted into the original calibration equation in order to determine the sector weights w_i.

In case of high PDs, the recalibration scheme presented above no longer works properly. High PDs pose a problem within the CR^+ framework, not only in the context of parameterization but in general terms, since the analytical tractability of CR^+ is based on the assumption of a so-called Poisson approximation which is only valid for small PDs. Nonetheless in this case parametrization schemes can be defined that allow the treatment of high PD facilities while still maintaining consistency between Merton and CR^+ contributory economic capital figures.

DEFINITION OF TEST PORTFOLIOS

Test portfolios are defined in order to study the different steps of the parameterization scheme between Basel II, Merton, and CR^+ modeling frameworks. This parameterization scheme is illustrated in Figure 24.1. For homogeneous, granular portfolios, the Basel II formula is the analytic approximation of the single-factor Merton model. Consistency with respect to economic capital figures is therefore expected between Merton and Basel II for portfolios fulfilling these requirements. The outlined parameterization scheme allows for consistent parameter translation between CR^+ and Merton models and leads to capital figures consistent with the Basel II formula in case of homogeneous, granular portfolios.

Figure 24.1 Parameterization Scheme between the Basel II, Merton, and CR$^+$ Modeling Framework



Heterogeneous portfolios show different capital figures with respect to Basel II and Merton models; this is due to Basel II being unaffected by diversification between different industries and countries. Additionally, within Basel II there is no sensitivity with respect to exposure concentrations. As the natural extension of the Basel II formula the multifactor Merton model is considered the benchmark model in case of capital calculations within nongranular and heterogeneous portfolios. As such the Basel II and the multifactor Merton models are still comparable in the sense that model implied differences are easily understood and transparent.

The re-parameterization framework outlined in this chapter makes the CR$^+$ capital figures consistent with both: the Merton-based framework in the case of heterogeneous, nongranular portfolios and the Merton and Basel II model in the case of homogeneous and granular portfolios.

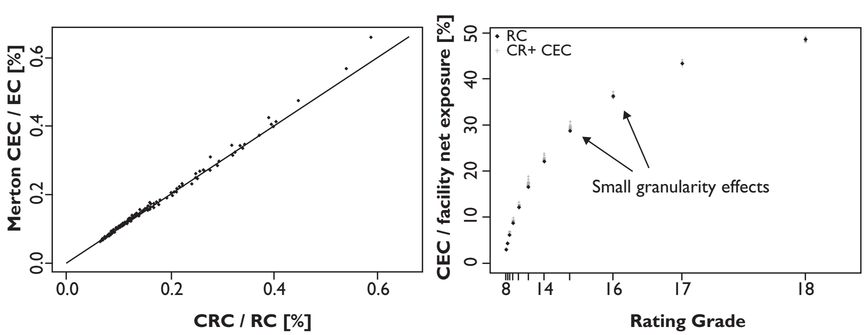
Homogeneous Portfolio

The homogeneous portfolio is a sample of 250,000 single positions (each position belongs to a single counterparty) from a retail portfolio. All clients are domiciled in the same country. This means all clients in the portfolio are affected by the same macroeconomy and no further country or industry diversification can be expected.

In order for the Merton model to reproduce the Basel II capital figures within this portfolio it is necessary that the chosen portfolio is sufficiently granular. For this reason a primary analysis is conducted, where a Monte Carlo–based single-factor Merton model is applied. The contributory economic capital (CEC) figures are then compared with the Basel II contributory regulatory capital (CRC) values. Since the Basel II formula is the analytic solution of a single-factor Merton model, it is expected that the capital figures of the Monte Carlo–based approach and the Basel II formula are the same once the infinite granularity assumption is approximately fulfilled.

Figure 24.2 shows the result of this granularity analysis. The left-hand panel shows a scatter plot between the Monte Carlo–based Merton CEC as a percentage of the total economic capital (EC) versus the CRC as a percentage of the total regulatory capital (RC). For the remainder of this chapter, these ratios are denoted as the internal CR^+/Merton model and the regulatory formula capital allocation key. Each point in the figure corresponds to a single counterparty-specific transaction. Full agreement between the Monte Carlo–based model and the Basel II model would be achieved if each point were located exactly on the diagonal line. As shown in the plot, all points are sufficiently close to the diagonal line. The largest relative differences between the allocation keys CEC/total EC versus CRC/total RC are a few percentages by order of magnitude, suggesting that the granularity assumption is valid. The right-hand panel shows the dependence of allocated capital on the rating grades of the master scale. Here, black diamonds represent the contributory capital figures from the Basel II formula; grey crosses represent the corresponding figures from the Monte Carlo Merton model. The small deviations that can be observed are due to small granularity effects still present in the test portfolio. Nevertheless, granularity effects are very small and play no role for further investigations.

Figure 24.2 Analysis of Granularity Effects Due to the Finite Size of the Test Portfolio: Comparison of Contributory Economic Capital Figures between Monte Carlo–based Single-factor Model and Basel II Formula



Overall, the homogeneous portfolio is large enough to avoid granularity effects. The asset (and hence the default) correlations are driven by a single economic factor, the country of domicile and by one idiosyncratic factor together with the asset correlation R^2. The Basel II regulatory formula as the analytic solution of a single-factor Merton model can therefore be regarded as the benchmark.

Heterogeneous Portfolio

The heterogeneous portfolio is a sample of 25,000 single positions (each position belongs to one single counterparty) of a universal bank portfolio. The counterparties are domiciled in over 10 European countries and the portfolio contains loans from 10 different industries. Overall, this portfolio is reasonably well diversified across countries and industries.

The mean portfolio PD is 2.5 percent while the exposure weighted PD is 1.5 percent, indicating the exposure contribution of the corporate and bank counterparty segment, which have on average a smaller PD. Since the portfolio contains a small proportion of very high net exposures, it displays large granularity effects and is heavily nonhomogeneous. The largest exposure as percentage of the total net exposure is 12 percent, indicating high concentration risk.

In summary, the heterogeneous portfolio contains all sorts of loans, is diversified across different countries and industries and contains large exposures with high concentration risk at the lowest PD ranges. The portfolio has a highly nonhomogeneous, strongly skewed exposure distribution.

The Basel II regulatory capital formula cannot be applied here, since the two basic assumptions leading to the analytic solution are violated:

1. The assumption of a single systematic risk driver
2. The assumption of infinite granularity

Since this portfolio cannot be treated by the Basel II regulatory framework, the more general framework of a multifactor Merton model sets the benchmark. Once the multifactor Merton model is setup, the model risk due to the CR^+ model for this portfolio can be investigated through the re-parameterization according to the Merton model.

Numerical Results: Homogeneous Portfolio

We start the analysis by calculating the capital of the homogeneous portfolio with the Basel II formula and the CR^+ model with basic parameter settings.

The regulatory Basel II capital figures are called regulatory capital (RC) and those calculated by CR^+ are denoted economic capital (EC). Contributory economic capital figures at counterparty level as calculated by the Basel II formula are denoted as contributory regulatory capital (CRC), and those calculated by the internal model CR^+ or Merton are called contributory economic capital (CEC).

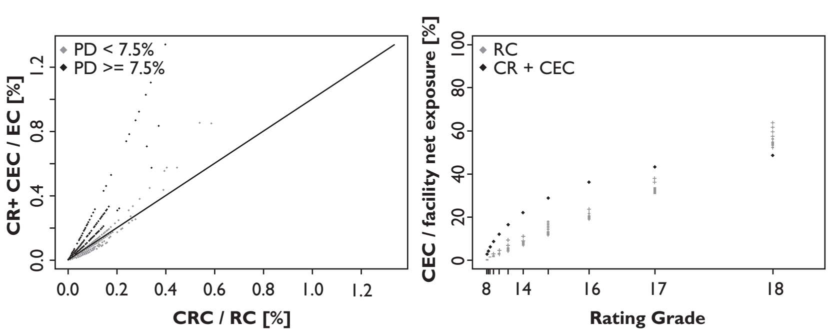
In the standard calibration scheme of CR^+, the PD volatility of counterparties is a crucial modeling parameter. However, within the retail segment, the PD volatility is difficult to estimate via historical default rates. The time series of default rates for the retail segment of most banks are only available over a few years. Even if default rates are available over a longer time horizon, their differentiation according to creditworthiness or rating is mostly not possible for more than five years. For this reason the PD volatility of any counterparty is set to be equal to 50 percent of the PD itself, which is a common value taken by many CR^+ practitioners. The calibration of PD volatilities based on internal default time series usually exhibits much smaller volatilities, this generally leads to significant underestimation of economic capital. By definition, the weight given to the systematic sector is set to 100 percent. A reduction of this weight in favor of the idiosyncratic sector would have the undesirable effect of further reducing the overall capital (see analysis below).

For the calculation of regulatory capital within the Basel II framework, the single asset correlation R^2 is set to 15 percent for all counterparties in the homogeneous retail portfolio. This value corresponds to the Basel II parameter setting for mortgage loans which dominate the exposure in the retail segment. For other retail segments such as credit cards or personal loans, the asset correlation of 15 percent is a rather conservative but reasonable setting. The second parameter used by the regulatory capital formula is the PD which is given by the rating. The estimation of the credit value at risk as well as the capital allocation to counterparty level is done consistently at the 99.95 percent quantile; this is in line with the target rating of many financial institutions. In order to be able to compare the resulting capital allocation of the internal model and the Basel II formula also, the latter is applied at the 99.95 percent quantile. Throughout the whole analysis of this chapter, we consistently use the expected shortfall as the best practice risk measure. The expected shortfall threshold is chosen in order to align the credit value at risk (CVaR) and the expected shortfall at the same numerical value, i.e., ESF(@99.95 percent C.L.) = CVaR(@99.95 percent C.L.).

Given these parameter settings, CR^+ in its basic calibration form severely underestimates the total capital compared with the Basel II regulatory formula: by the means of CR^+ internally estimated portfolio EC amounts to approximately 40 percent of the respective Basel II figure. This is partly a consequence of the conservative setting of the asset correlations; however, the entire difference cannot be attributed to this. Empirical data analyses of default rate data do not provide evidence for choosing a higher PD volatility, which is the only parameter which could cause the increase of the CR^+ portfolio EC. The systematic sector weight within CR^+ is 100 percent and any reduction in favor of the idiosyncratic sector would further decrease the capital.

A similar discrepancy between the regulatory capital and the internal capital occurs if the contributory capital figures at counterparty level are analyzed. The results are plotted in Figure 24.3. The left-hand panel compares the internal CR^+ model and the regulatory formula allocation key. Each dot in the figure corresponds to a single counterparty. The allocation key itself is defined as the allocated capital in percentage of the total capital, calculated by either the CR^+ or the regulatory formula. The diagonal line indicates the position of perfect concordance between the models. It is clear that the allocation key fluctuates heavily around the diagonal, indicating discrepancies between the allocation keys of the two models. Model discrepancies are largest when PDs are highest. The plot differentiates between PD $<$ 7.5 percent (gray triangles) and PD $>$ 7.5 percent (black diamonds). From a purely methodological perspective, the differences for high PDs are due to the fact that CR^+ itself is only applicable for small PDs. The largest relative differences between the allocation key are in the order of several hundred percent. This is due to the fact that the analytic tractability of CR^+ assumes a Poisson approximation, which is only valid if PDs are very small. When counterparties have PDs of 10 percent or higher, this assumption is violated, causing additional model-implied error for capital calculation and allocation.

Figure 24.3 Results of Capital Allocation for Homogeneous Portfolio with Basic Parameter Settings



The right-hand panel shows the dependency of the allocated capital of the counterparty net exposure on the PD denoted by the rating grade. The CRC figures show a characteristic dependence on the counterparty PD, whereas the CEC figures have a constant slope. Considering the economics of the situation, the Basel II capital formula shows the correct dependence, since the relative concentration risk is highest for the counterparties with small PD, i.e., the EC/EL ratio is expected to increase with decreasing PD. Since the CR^+ model is not sensitive to this effect, the model in its initial, basic parameterization is not capable of reflecting risk sensitivity with respect to concentration risk. It should be mentioned here that if choosing some higher value for the PD volatility, the capital on portfolio level can normally be matched satisfactorily even when using the standard CR^+ parameterization. However, on counterparty level results will still deviate drastically and the qualitative behavior of the dependence of the CEC figures on the rating grade as displayed within the right hand side of Figure 24.3 remains unchanged. For this reason, capital allocation fails in case of the standard CR^+ parameterization.

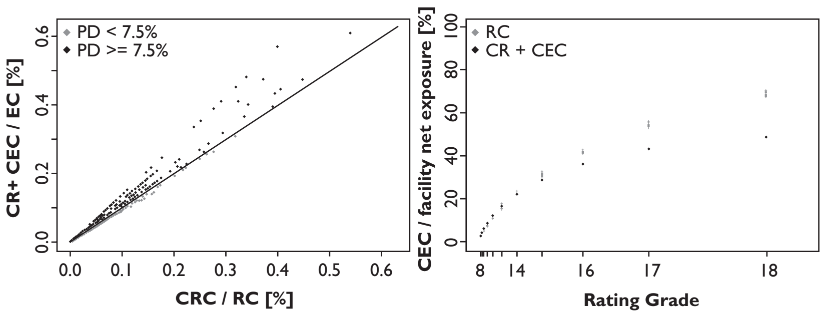
The second part of the analysis of the homogeneous retail portfolio is done via the application of the re-parameterization scheme between the Merton and CR^+ frameworks. Here the Basel II formula is just a single factor Merton model parameterized by the customer PD and the constant value of $R^2 = 15$ percent. The re-parameterization scheme calculates the weights of the systematic and idiosyncratic sectors as well as the PD and sector volatility of the CR^+ model.

The re-parameterization scheme works successfully and fully aligns the regulatory formula and the CR^+ internal capital model. The portfolio EC calculated by CR^+ is just 0.7 percent greater than the regulatory portfolio EC. Similarly good results are achieved for the allocation key, and the CEC dependency on the counterparty rating as shown in Figure 24.4.

The scattering of the allocation key around the diagonal line in the left-hand panel is reduced compared with the basic parameter settings. The largest deviations are still observed for the counterparties with high PDs. The relative differences between the allocation key for high PDs is 25 percent at most. This is a significant improvement compared with the basic parameterization where the allocation key values differed by up to several hundred percent. Considering the capital allocation with respect to the counterparty PD (right-hand panel) the re-parameterization scheme also leads to a similar slope (and therefore similar risk sensitivity) with respect to concentration risk when compared with the Merton framework.

In summary, it can be concluded that within this homogeneous and granular retail portfolio the re-parameterization scheme via the Merton

Figure 24.4 Results of Capital Allocation for Homogeneous Portfolio with CR⁺ Parameter Settings as Calculated by Re-parameterization Scheme from Merton to CR⁺



model delivers comparable results between the Merton/Basel II model and the CR^+ model. Although the Merton and CR^+ models are conceptually different approaches to credit portfolio modeling, the model implied risk due to parameter calibration approaches can be avoided by applying the parameterization rules proposed in this article.

Numerical Results: Heterogeneous Portfolio

The CVaR calculations of the heterogeneous portfolio are performed by applying the advanced re-parameterization scheme via Merton to the CR^+ model. In the case of the homogeneous test portfolio, the single-sector CR^+ model with basic parameterization failed to achieve results consistent with the Merton model as well as realistic risk sensitivity for nondiversified portfolios. It is therefore expected that the heterogeneous case will display similarly poor correspondence. Besides, internal bank analysis of default and rating migration data published by ratings agencies show PD volatilities for the corporate sector, which are too small to get meaningful capital figures.

We begin the parameterization by assigning asset correlations to the counterparties of the Merton model. To do this we take market best practice values of R^2 according to the different business sectors contained in the heterogeneous portfolio. The correlation model underlying the multifactor Merton model contains 10 industry factors and 10 European country factors. The factors itself are represented by MSCI indices and the correlations between the factors are estimated on a weekly return basis over a period of three years between January 1, 2005 and January 1, 2008.

For reasons of brevity the values of the asset and factor correlations are not displayed in detail. At a summarized level asset correlations are 15 percent for the retail sector, between 20 and 40 percent for the corporate sector and 30 and 60 percent for banking business, depending on the size and the industry sector of the companies. The factor correlations are approximately 70 percent on average, where correlations between industries and countries range from approximately 55 to 95 percent. All parameter settings are taken from real portfolio parameterizations which are currently running in the banking industry.

After having parameterized the Merton framework, the CR^+ parameters are deduced by the re-parameterization scheme. As mentioned previously, the regulatory analytic approximation of a single-factor Basel II model is not applicable to a heterogeneous well-diversified portfolio. The Monte Carlo–based multifactor Merton model—the natural extension of the Basel II capital formula—is used to set the benchmark results for the CR^+ model.

Having applied the re-parameterization scheme, the difference between the CVaR capital figures of the CR^+ and the Merton model amounts to only 2 percent at portfolio level. It can be concluded therefore that both models are in very good agreement at portfolio level once the re-parameterization scheme is applied. The regulatory capital calculated by the Basel II formula at the regulatory, lower-confidence level of 99.9 percent yields capital figures which are 10 percent below the internal Merton capital and 12 percent below the CR^+ capital. The diversification effects given in the internal models are therefore not sufficient to compensate the CVaR at the higher, internal 99.95 percent quantile compared with Basel II at 99.9 percent. At first glance this is counterintuitive since most EC capital modelers expect the internal capital figures to be below the regulatory capital. The high internal capital is explained by the use of high asset and factor correlations (when calibrated over a recent period of three years, namely, between January 1, 2005, and January 1, 2008, with the current financial crisis explicitly excluded), as well as by significant nongranularity effects caused by large exposure concentrations to counterparties with very small PD. These high parameter values driving the default correlations are based on current calibrations of actual market data over the past few years. Asset correlations and factor correlations have experienced an increase over the past 15 years; this can be attributed to the increasing integration of the European financial markets and the growing interdependencies of import and export due to the globalization of the markets.

The results of capital allocation corresponding to the regulatory Basel II formula and the Merton model can be deduced from Figures 24.5 and 24.6.

As expected, the allocation key between the diversified, multifactor Merton model and the regulatory capital model are in total disagreement,

Figure 24.5 Capital Allocation Key Between Relative Contributory Economic Capital of Multifactor Merton Model and Relative Contributory Regulatory Capital

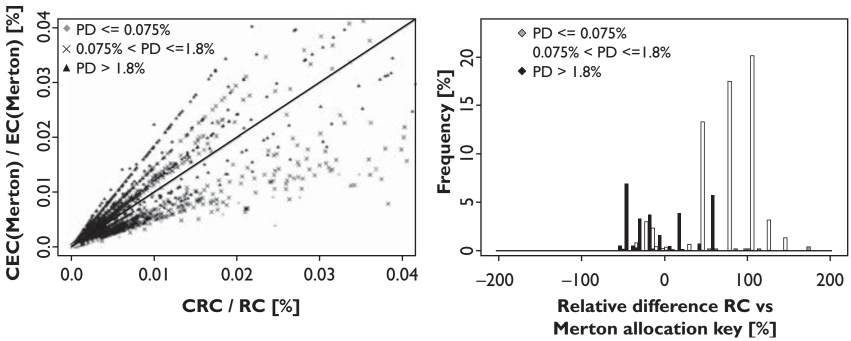
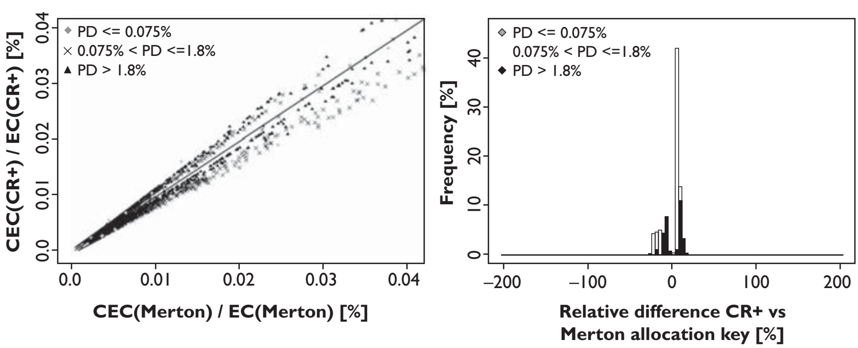


Figure 24.6 Capital Allocation Key Between Relative Contributory Economic Capital of Multifactor Merton Model and Relative CEC of CR$^+$ Model



due primarily to the nonsensitivity of the Basel II framework to country and industry diversification. Secondarily, the disagreement is a consequence of Basel II's lack of sensitivity to concentration risk. The largest model deviations are therefore seen for counterparties with very small PDs (PD $<$ 70 bps), where distortion of the allocation key is highest. The frequency distribution of the relative differences between the two allocation keys exhibits deviations of up to 200 percent. We conclude that the Basel II framework is not suitable for capital allocation.

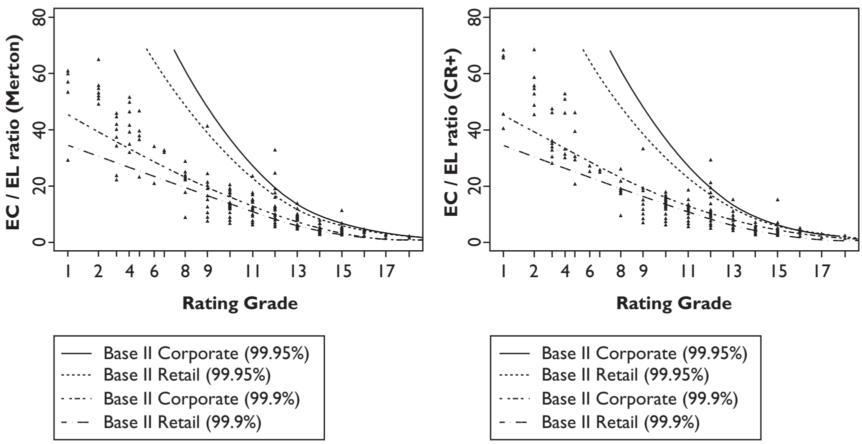
If the re-parameterization scheme is applied and the allocation key between the CR$^+$ and the Merton model is compared, a different situation emerges (Figure 24.6). The scattering of the allocation key around the diagonal line in Figure 24.6 is much smaller compared with that in Figure 24.5.

The close concordance between the models is shown once more on the right-hand panel of Figure 24.6. The relative differences in allocation key between CR^+ and Merton are well centered and the maximum disagreement is 20 percent. Considering the strong heterogeneity of the underlying portfolio and the fact that CR^+ does not resolve the full correlation model, the agreement is highly satisfactory.

In the following the concentration risk effects with respect to the counterparty creditworthiness is discussed and the outcome of the two models is compared. An intuitive way of considering concentration risk effects is to take the ratio between contributory capital and expected loss as a risk measure, i.e., the EC/EL ratio. In contrast to the total capital consumption of a transaction or the contribution of counterparty specific EC to total EC, the EC/EL ratio has no dependence on the counterparty exposure; it is a measure indicating high dependence on credit events in the portfolio and is an illustrative alternative to study effects of default correlations and risk concentrations. For example, a transaction or counterparty may have very low total or relative EC consumption; nevertheless, it can be highly risky in the sense that its default is highly correlated to other clients. This would be indicated by a high EC/EL ratio. Vice versa, high capital consumption can be less risky if the counterparty default is not correlated to other clients in the portfolio. In this case the risk would be indicated by EL and the EC/EL ratio would be low. From an economic viewpoint, concentration risk and therefore the EC/EL ratio must be a function of PD. The reason is that obligors with very small PDs are either large corporates or financial institutions. These counterparties are highly affected by cyclical effects of the economy. Looking at spread movements of such companies, the relative movements of the spread in economic downturns compared with the average value in economically calm environments can jump dramatically; on the other hand, spread movements for counterparties with lower creditworthiness are much smaller. As such, within a sound credit portfolio model which is risk sensitive with respect to concentration risk, the EC/EL ratio should increase with decreasing PD.

The results of this analysis are shown in Figure 24.7. The left-hand panel shows the EC/EL ratio for any counterparty as a function of PD indicated by the rating grades (logarithmic scale) of the master scale as calculated by the internal, multifactor Merton model. In addition, the regulatory EC/EL ratio for retail and corporate clients are displayed at different confidence levels of 99.9 and 99.95 percent. The increasing curve of EC/EL with decreasing PD is nicely reproduced. The diversification effects of the internal Merton model with respect to Basel II are indicated by lower EC/EL ratios on average. Nongranularity effects can be easily observed,

Figure 24.7 The EC/EL Ratio as Function of PD Indicated by Rating Grades of the Master Scale



since for each rating grade the EC/EL ratio is spread over a large interval. Facilities with high relative exposure and low granularity receive a higher capital charge and those with high granularity get capital relief. After having applied the re-parameterization scheme from Merton to CR^+, the same analysis is conducted by taking the contributory capital figures of CR^+ into account (as illustrated on the right-hand side of 24.7). It is apparent that the CR^+ capital allocation shows the same slope and the same granularity effects as the Merton model. Again, the re-parameterization scheme results in model correspondence. The final plots shown here can be considered the most strenuous comparison tests of credit portfolio models. The particular shape of the slopes as displayed within Figure 24.7 is due to diversification effects with respect to Basel II and to the huge impact of concentration risk due to nongranularity effects.

CONCLUSION

This chapter analyses model risks implied by the inconsistent parameterization of differing credit portfolio models. Two major modeling approaches, namely CR^+ and Merton models, are shown to contain significant model-implied risk for portfolio-wide capital figures and the concentration-risk-vital allocation keys.

The aim of this chapter is to analyze the model-implied risk due to different methods of calibration of the risk parameters underlying the CR^+ and Merton models.

A novel approach to overcome these parameterization-implied discrepancies is outlined, which is based on a translation of Merton model parameters to CR^+ parameters. As a result, enhanced model consistency can be achieved. It is shown that when homogeneous, nondiversified, granular portfolios are considered full consistency between CR^+, Merton, and Basel II model capital figures can be achieved. Parity between internal models and Basel II is not achievable for diversified, heterogeneous portfolios exhibiting strong nongranularity effects and exposure concentrations; this is due to the basic nature of Basel II models which do not take into account industry diversification and correlation effects and assume full granularity. The re-parameterization approach provides consistency between the CR^+ and multi-factor Merton models, the latter of which is considered the natural extension of the single-factor Basel II formula.

REFERENCES

Avesani, G.R., Liu, K., Mirestean, A., and Salvati, J. (2006) Review and implementation of credit risk models of the financial sector assessment program. IMF Working paper WP/06/134, International Monetary Fund.

Giese, G. (2003) Enhancing CreditRisk$^+$. *Risk*, 16(4): 73–77.

Gordy, M.B. (2000) A comparative anatomy of credit risk models. *Journal of Banking & Finance*, 24: 119–149.

This page intentionally left blank

CHAPTER 25

MODEL RISK FOR MARKET RISK MODELING

Jason C. Hsu, Vitali Kalesnik, and
Shane D. Shepherd

ABSTRACT

Any model is a simplified version of reality, and thus always contains the possibility that the simplifying assumptions do not well match, or properly account for, the true governing processes of the world. We refer to model risk as the risk of not accurately estimating the probability of future losses due to a failure of a risk model. Sources of model risk in risk measurement models include differences between the assumed and actual distribution and errors in the logical framework of the model. Thus, successfully estimating the risk to one's portfolio requires not only accurate estimates for the inputs to the model, but also a detailed understanding of the distributions from which the risky outcomes arise. Understanding how a model might fail is the first step toward rectifying the problem of model risk. We demonstrate a measurement of this model risk associated with the problem of parameter uncertainty in a value-at-risk model. Oftentimes the parameters of a distribution may be estimated with low precision, or there may be disagreement about the governing processes on which to base a risk model. Existing standard risk models do not adequately handle this parameter uncertainty. We show how these traditional methods for handling parameter uncertainty often fail, and provide a technique for quantifying risk more accurately.

INTRODUCTION

Model risk comes primarily from two possible shortcomings in the model development process. These shortcomings come out in phrases such as "models are only as good as the assumptions they are based on"; and "models are only as good as the data behind them." Commonly, these risks express themselves as misspecified distributions and parameter uncertainty. Consider the well-known Black-Scholes options pricing model, one of the most successful pricing models in finance. Even so, it falls prey to both of these problems. First it assumes a normal distribution on asset returns. Thus in a world filled with fat tails and volatility smiles, a naïve Black-Scholes user will underpay for all options, misprice at-the-money compared with out-of the-money options, and quickly go broke. Second, the price generated from a Black-Scholes equation is only as valid as the estimate used for the asset volatility. This crucial parameter must be estimated from historical volatility or forecasted from a volatility model. The uncertainty behind this volatility estimate can explode on those with too much faith in their model outputs.

Furthermore, tertiary risks abound. These can be as basic as implementation risk (say, a trader uses Black-Scholes to value an American put) to regime shifts (volatility smiles suddenly appearing) to unknown or unanticipated risks (such as worsening liquidity placing stress upon heavily leveraged arbitrageurs such as Long Term Capital Management).

In this chapter we focus on the model risk contained in value-at-risk (VaR) models, and primarily on the risk of parameter uncertainty. We build upon the ideas discussed in Hsu and Kalesnik (2009) and demonstrate that properly accounting for parameter uncertainty can result in posterior distributions that do a superior job of capturing return characteristics without resorting to exotic distributions that are difficult to work with and may not properly describe asset return behavior in any case.

The issues of misspecified distributions have received tremendous attention in the last decade since the collapse of Long Term Capital Management. In particular, the distributions used in standard VaR analysis do not adequately capture the frequency of extreme shocks to asset prices (kurtosis) or the size of those shocks (negative skew). This topic even went mainstream with the publication of Nassim Taleb's 2007 book, *The Black Swan: The Impact of the Highly Improbable*, which discussed in detail the manifestations of fat tails. The difficulty of addressing these problems in risk management has received further attention from authors such as Derman (1996), Hendricks (1996), and Nocera (2009).

More advanced practitioners of VaR methods have begun to employ more esoteric distributional assumptions around asset returns, modeling them with Levy, Cauchy, or other fat-tailed stable Paretian distributions in an effort to capture this kurtosis and a negative skew. Lucas (2000) is but one of the practitioners who have published journal articles discussing these techniques. However, when VaR estimates are produced using more esoteric distributional assumptions, the process loses much of its intuitive appeal. The parameter choice for these distributions can be complicated and difficult to agree upon. It becomes increasingly difficult to express committee views in the parameter choice. Additionally, these models fall prey to a bias towards certainty inherent in the modeling process: as the level of complexity of the model grows, almost always our estimation of the remaining uncertainty shrinks more quickly than the uncertainty itself. This is particularly true when the output of the model is so incredibly simple—a single dollar figure, expressing the maximum likely loss.

UNCERTAINTY AROUND THE MEAN ESTIMATE

Our approach can be illustrated with an example of disagreement around something as simple as mean return. Suppose a five-member investment committee is attempting to measure VaR on their equities index portfolio at the beginning of 2009. Given the tremendous volatility through the bear market of 2008, members of the committee could hold widely differing views on possible returns for 2009. Three members of our committee believe that the unprecedented monetary stimulus and likely fiscal stimulus package will, over the course of the year, right the economy and rescue the United States from a prolonged recession. They believe that, given the sharply discounted valuation ratios, expected returns for 2009 will come in at 20 percent. On the other hand, two members of the committee believe that the worst is yet to come. They see the economy entering into a multi-year recession with no improvement on the horizon, and see room for valuation levels to fall even further. The bearish group believes that equities will continue downwards with a substantially negative return of −20 percent. To illustrate the point, both sides will agree that volatility will average 15 percent (later we will look at volatility uncertainty).

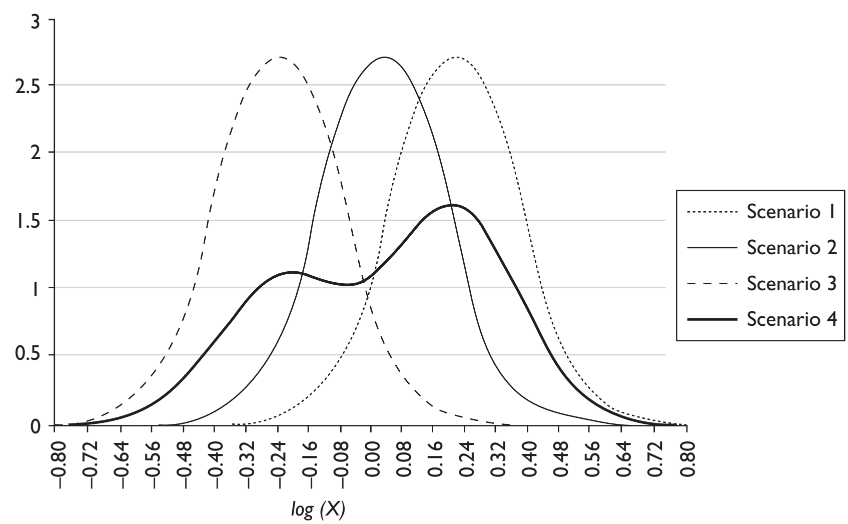
Given this agreement, how should the committee model the market risk faced by their portfolio? We will consider four different methods to quantify the disparate views amongst the committee members. We will then utilize these assumptions in a log-normal distribution to determine the VaR for the portfolio. In the following notation, $N(\mu, \sigma)$ is a normal distribution with mean and standard deviation of (μ, σ).

1. We could use the majority opinion number of +20 percent as our mean estimate. Our distribution is: $\ln r_1 \sim N(20\%, 15\%)$.
2. We could use the mean estimate of the committee members for our expected return. This gives a distribution of: $\ln r_2 \sim N(4\%, 15\%)$.
3. To protect against a worst-case scenario, the committee could decide to assume an expected return of −20% and use $\ln r_3 \sim N(-20\%, 15\%)$.
4. Finally, the committee could explicitly model the uncertainty in the expected mean [Insert eq. here] return. This gives a return distribution of:

$$\text{In } r_4 \sim \begin{Bmatrix} prob = \frac{2}{5}, N(-20\%, 15\% \\ prob = \frac{3}{5}, N(-20\%, 15\% \end{Bmatrix}$$

We plot the four ex ante distribution functions in Figure 25.1. Note how explicitly modeling the mean uncertainty captures the bimodal views of the group while assigning a far lower probability to the mean outcome (which no individual expects to occur) than using the second scenario. From the

Figure 25.1 Probability Density Function for Equity Returns with Uncertainty of the Mean



distribution we can also see that scenario 4 results in a distribution with large negative kurtosis, even though the individual distributions had kurtosis of zero before blending them together. Thus the mean uncertainty results in thinner tails rather than the fat tails modeled by Cauchy or Levy distributions. However, we can see that the parameter uncertainty leads to decidedly non–log-normal distributions, even when the starting distributions are themselves log-normal.

The risk statistics presented in Table 25.1 give further insight into these distributions. The first four columns characterize the ex ante distributions. The fifth column shows the VaR at the 5 percent confidence level for the equity portfolio. The sixth column shows the expected loss conditional on observations in the lowest 5 percent tail of the distribution. And the seventh column shows the results of an investment decision rule: if we wish to cap our expected losses at −25 percent with a 5 percent probability, this column shows the maximum percentage of our assets to be invested in equities, with the remainder in a zero-return cash fund.

The differences in the first three distributions all come as a result of parameter selection. The first is clearly suboptimal because it ignores information; the majority rule approach fails to take into account the range of possible outcomes and will underestimate the VaR and often be overinvested. Scenario 2 shows that using the mean estimate delivers an improved risk assessment—the 5 percent VaR level and expected shortfall both increase dramatically. However, given this setup, the investment allocation decision does not change. Scenario 2 is still a naïve approach, and underestimates both the VaR and the expected shortfall. Note that both scenario 2 and scenario 4 have the same expected mean, but in scenario 4 the resulting standard deviation becomes considerably larger by modeling the mean uncertainty.

Although the worst-case assumption shown in scenario 3 shows the most similar risk characteristics to the mean uncertainty model, it is still suboptimal

Table 25.1 Risk Statistics for Equity Returns with Uncertainty of the Mean

	Mean	Volatility	Skewness	Kurtosis	5% VaR	Expected Shortfall	Max % Invested (5% chance of 25% loss)
Scenario 1	20.00%	15.00%	0.00%	0.00%	4.59%	10.42%	100.00%
Scenario 2	4.00%	15.00%	0.00%	0.00%	18.70%	23.70%	100.00%
Scenario 3	−20.00%	15.00%	0.00%	0.00%	36.05%	39.95%	69.36%
Scenario 4	4.00%	24.68%	−20.44%	−72.89%	31.13%	36.11%	80.30%

because it results in a significant under allocation to equities. This 11 percent under allocation costs the portfolio 44 basis points in annual expected return.

The table also shows numerically what was strikingly depicted in the figure: the mean uncertainty approach leads to negative skewness and large negative kurtosis. However, we can also see that the negative kurtosis is dominated by the increase in the volatility from 15 to 24 percent. This suggests that exotic explicit modeling of fat tails may not be as important as properly modeling parameter uncertainty.

UNCERTAINTY AROUND THE VARIANCE ESTIMATE

Perhaps we have a different situation, where a group of investors may agree on the mean return but differ as to expected future volatility. In this section we illustrate how modeling uncertainty in our standard deviation estimates can improve VaR forecasts. We take the same approach as before, with a five-member investment committee. This time they all agree on an expected annual return of 10 percent, but three members expect a relatively quiet market with volatility of 12 percent while two members forecast a much higher volatility of 25 percent. Again we compare four approaches to modeling VaR in this situation. In the first, the majority rule decision would model volatility at 12 percent; in the second, we average the individual votes and utilize a volatility of 17.25 percent; in the third, we apply the worst-case scenario of 25 percent; and the fourth directly models the uncertainty in the variance parameter estimate in the same manner as that used in the previous section for the mean return.

Figure 25.2 shows the ex ante probability density functions for each of the four scenarios, and Table 25.2 shows the risk characteristics. As expected, scenario 1 and scenario 3 provide the least and most, respectively, conservative risk assessments. The interesting comparison is between the use of the average forecast, scenario 2, and directly modeling the uncertainty in scenario 4. We see that scenario 4 provides a higher standard deviation than that in scenario 2. However, the difference between these two is nowhere near as large as the earlier case where we modeled uncertainty around the mean expected return (an increase in volatility from 15 to 24.6 percent, compared with a rather mild increase from 17.2 to 18.3 percent). Also, uncertainty in the variance estimate produces high positive kurtosis instead of the negative kurtosis produced by uncertainty in the mean estimate. This transforms the starting log-normal distributions into a fat-tailed distribution, and it is this increase in kurtosis that drives the major differences in risk assessment between scenario 2 and scenario 4.

Figure 25.2 Probability Density Function for Equity Returns with Uncertainty of the Mean

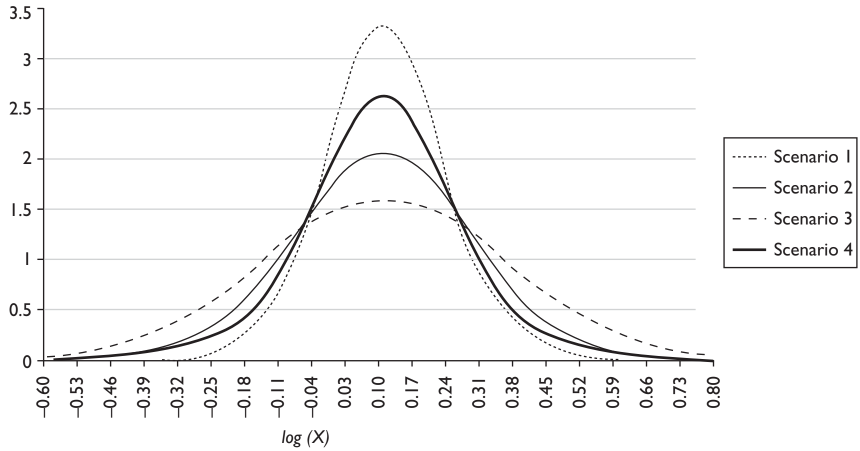


Table 25.2 Risk Statistics for Equity Returns with Uncertainty of the Mean

	Mean	Volatility	Skewness	Kurtosis	5% VaR	Expected Shortfall	Max % Invested (5% chance of 25% loss)
Scenario 1	10.00%	12.00%	0.00%	0.00%	9.31%	13.93%	100.00%
Scenario 2	10.00%	17.20%	0.00%	0.00%	16.74%	22.40%	100.00%
Scenario 3	10.00%	25.00%	0.00%	0.00%	26.77%	33.87%	93.38%
Scenario 4	10.00%	18.34%	0.00%	147.20%	18.09%	26.68%	100.00%

HISTORICAL TESTING OF THE MEAN UNCERTAINTY MODEL IN VAR

We next conduct a historical test of how explicitly modeling uncertainty in the mean return would have impacted VaR estimates. Our test is based on a setup similar to our example of mean uncertainty. We follow an investment committee of five members who estimate VaR for their investment in the S&P 500 index from 1987 through June 2009. The committee is divided into two camps. Three members are believers in mean reversion; they estimate the future return for their equity investment based on a comparison of the past 36-month returns to a long-run average equity premium. The second group of two members believe in trend following, and base their forecasted return on the average past three-month return.[1] At times these expectations will be very similar, and at times they will differ dramatically.

This gives us a chance to demonstrate how the amount of disagreement affects the quality of VaR calculations. For simplicity, both groups agree to estimate the expected volatility by using the historical three-year average. We should note that neither of these two strategies are being proposed as a "best-implementation" type of strategy. Rather, they have been chosen as contrasting forecast models and their ability to demonstrate how the level of disagreement in mean return estimates can influence the performance of VaR models.

For each month in our time series, we calculate four estimates of 5 percent VaR: an estimate based on the mean reversion forecast, and estimate based on the trend following forecast, an estimate based on the weighted average forecast, and an estimate based on our technique of explicitly modeling the uncertainty in the expected return. We then judge the quality of the forecasts based on the frequency of violations.

In Figure 25.3, we see the history of the four VaR estimates based on actual market returns. Although the forecasts vary quite a bit through time, each of the models delivers a very similar average forecast of between −6.25 and −6.62 percent. The trending forecast, being based on a much shorter amount of data, produces much more volatile estimates. Yet all four measures are highly correlated and rather slow moving through time.

Table 25.3 displays these model performance numbers. Across all months in our sample, we see that both the mean reversion forecast and the trending forecast have historically underestimated VaR. Their resulting violations of

Figure 25.3 Historical VaR Estimates Based on S&P 500 Returns

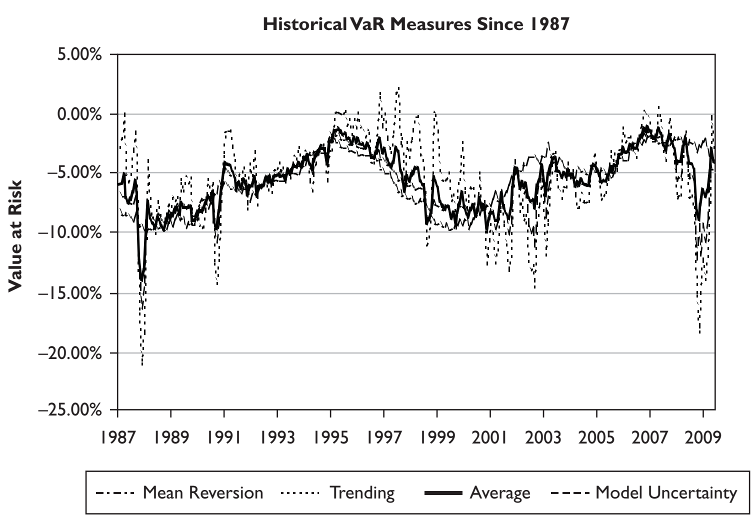


Table 25.3 Frequencies of 5% VaR Forecast Violations, 1987–2009

	Mean Reversion	Trending	Average	Mean Uncertainty	Expected Shortfall
All Months	7.78%	8.89%	6.67%	5.56%	−8.33%
High Disagreement	11.19%	11.19%	8.21%	5.97%	−9.15%
High Agreement	4.41%	6.62%	5.15%	5.15%	−7.52%

7.78 and 8.89 percent significantly exceed the 5 percent target. However, the interesting comparison is between the third estimate, taking the average forecast, and the fourth, which involves directly modeling the mean uncertainty. Both of these measures still underestimate the 5 percent VaR bound, but perform considerably better than either of the individual forecasts. While the average estimate produces a 6.67 percent violation frequency, the mean uncertainty model results in a 5.56 percent violation—still exceeding, but much closer to, the 5 percent target and a full percentage point better than the average estimate.

The second part of Table 25.3 demonstrates how the strength of disagreement in the forecasts influences the quality of the VaR estimates. We rank each month based on the absolute difference between the mean reversion estimate and the trending estimate. We then split the sample into two, with those in the half with the greatest difference considered "High Disagreement" periods and those in the second half labeled "High Agreement" periods.

When our mean reversion and trending models have relatively similar forecasts, little can be gained from modeling this disagreement. We see that the mean reversion estimate slightly underestimates, and the trending forecast overestimates, the VaR figures. And the average forecast and mean uncertainty forecast are equally good (and quite good in this historical test) at 5.15 percent. However, the months where there is a high disagreement in the forecasts show a different result. The two naïve forecasts both perform poorly at 11.19 percent violations, and the average forecast produces an 8.21 percent violation rate. However, the mean uncertainty model still scores quite well with a 5.97 percent violation rate—very good, considering the alternative.

Note that the periods of agreement tend to be quiet market periods of lower volatility and higher returns. In these circumstances, the penalty for a poor VaR estimate is generally relatively light. It is in the periods of high disagreement, characterized by higher volatility, greater negative returns, and turning points in the market, that the penalties for underestimating VaR are at their greatest. These periods include the market crash in October 1987, the entire 12 months from June 2008 to May 2009, and much of the crash in the technology bubble of 2000 to 2001. These are the times when we would most like to have reliable estimates, and these are exactly the times when the relative performance of the mean uncertainty model is at its best.

The final column of Table 25.3 shows the expected shortfalls estimated by the mean uncertainty model. It predicts an average shortfall of −8.33 percent across all months, with a significantly greater shortfall of −9.15 percent in the High Disagreement months compared to an expected shortfall of −7.52 percent in the High Agreement months. This also shows the greater value of precise VaR estimates during times of higher market uncertainty.

CONCLUSION

We have presented a way to directly model uncertainty in key parameters commonly used in value-at-risk models. Such disagreements are commonplace in financial markets at large as well as within organizations. We show that by directly incorporating diverse views into our risk models, the resulting VaR estimates exhibit superior performance characteristics to other commonly used techniques. Not only do the resulting ex ante distributions better capture the assumptions underlying the investment, they also result in more accurate VaR predictions in our test case for the S&P 500 returns. Particularly when the level of disagreement is high, leading to high levels of parameter uncertainty, we see that directly modeling this parameter uncertainty performs exceptionally well.

Standard risk management approaches fail to consider parameter uncertainty, which has led to improper risk management. Blind faith in parameter estimates has too often led to blind faith in the resulting VaR outputs, and when these estimates are too often exceeded, the proposed solution is commonly to fatten up the tails by using exotic distributions. We show, however, that directly modeling the uncertainty in mean and variance returns using standard log-normal distributions can result in posterior distributions with high degrees of skewness and kurtosis. If we accept a simple world of time-varying expected returns and variances, the resulting uncertainty around these constantly shifting parameters places us squarely in this world of interesting and effective posterior distributions.

REFERENCES

Derman, E. (1996) Model Risk. *Quantitative Strategies and Research Notes.* Goldman, Sachs, New York.

Hendricks, D. (1996) Evaluation of Value at Risk Models Using Historical Data. Federal Reserve Bank of New York, *Economic Policy Review*, 2(1): 49–69.

Hsu, J. and Kalesnik, V. (2009) Risk Managing the Uncertainty in VaR Model Parameters. In G. Gregoriou (ed.), *The VaR Implementation Handbook*. New York: McGraw-Hill.

Lucas, A. (2000) A Note on Optimal Estimation from a Risk-Management Perspective Under Possibly Misspecified Tail Behavior. *Journal of Business and Economic Statistics*, 18(1): 31–39.

Nocera, J. (2009) Risk Management. *The New York Times*, January 4, 2009.

Taleb, M. (2007) *The Black Swan: The Impact of the Highly Improbably*. New York: Random House.

NOTES

1. The amount of expected mean reversion is parameterized so that both groups have, on average, the same expected return for their equity investment.

This page intentionally left blank

CHAPTER 26

EVALUATING THE ADEQUACY OF MARKET RISK MODELS

Carsten S. Wehn[1]

ABSTRACT

The recent financial crisis has shown that models often have their weaknesses, which is also valid against risk models. Thus, the credibility of a risk model's results has to be defended more intensely than ever. In this chapter, it's argued that constructing and implementing a model to measure market risk is still necessary and unavoidable, because having only rudimentary information regarding market risks is better than having nothing quantitative at all. More than ever, it is crucial to identify and communicate the modeling assumptions, and it is indispensable to work continuously to improve the quality of market risk models with a view towards their adequacy for risk measurement and management purposes.

The chapter presents in a systematic and concise manner the building blocks of a market risk model and describes in detail two approaches for assessing the model's adequacy: first by assessing the individual building blocks and by subsequently backtesting the model as a whole. Embedded in a regular validation process, this can help improve the model continuously. Regulatory issues are also considered.

MARKET RISK MODELING AND MODEL RISK

Building Blocks of Market Risk Models

The impact of market risk factors on the gains and losses associated with a portfolio of investments is referred to as market risk. Theses risk factors might be, e.g., interest rates, equity prices, currencies, or implied volatilities. The starting position for applying adequate statistical means to model and measure market risk is relatively good as market risk factors are generally readily observable. The measurement of market risk is preceded by a prediction of the distribution of the future (and therefore unknown) gains and losses. These gains and losses of a portfolio are denoted as a random variable[2] G_t. This is completely analogous to the situation encountered when pricing a financial asset as the risk modeling of portfolio gains and losses takes all relevant information into account. This information usually is given by observed risk factor prices or risk factor returns. Thus, the relevant (univariate) predictive distribution is $F_t := F_t\,(G_t \,|\, R_{t-1}, R_{t-2}, \ldots)$, where $R_{t-1}, R_{t-2}, \ldots$ denotes (multivariate) past risk factor returns.

Several building blocks mark the derivation of a predictive distribution for portfolio gains and losses: first, an identification of relevant risk factors (such as interest rates for different maturities and rating classes, equity indices or equity prices, exchange rates, implied volatilities and so on) is required.

Second, one must choose a relevant multivariate distribution $F_t(R_t \,|\, R_{t-1}, R_{t-2}, \ldots)$ for the risk factor returns $R_t \,|\, R_{t-1}, R_{t-2}, \ldots$. Here, we should note that we are dealing with a conditional return distribution (i.e., conditional on previous observations). Widely used models include generalized autoregressive conditional heteroscedasticity (GARCH) and autoregressive conditional heteroscedasticity (ARCH) models for returns of time series.[3] Time series of risk factors experience so-called stylized facts. In the model setup above, this is reflected in the conditional distribution of the risk factor returns. The stylized facts comprise autocorrelation of risk factor values and serially almost uncorrelated respective returns. Often volatility clusters are observable for the returns and the (unconditional) distribution of the risk factor returns is leptokurtic. The daily return distribution has a mean around zero and is almost symmetrical.

Third, risk models usually make assumptions concerning the relationship (mapping) between risk factors (or risk factor returns) and the portfolio gain and loss function (e.g., sensitivities and "greeks," full valuation, present value grids, etc.).

Most risk models show parameters of the conditional distribution F_t such as a quantile for the value at risk or conditional means for the expected shortfall. Value at risk for example is defined as the α-quantile[4] of the conditional distribution of a portfolio's gains and losses: $q_t^\alpha = -F_t^{-1}(1-\alpha)$. Nevertheless, the following methods focus mainly on the entire conditional distribution and in some parts also rely on the respective quantiles. For a daily risk measurement process, this mapping between risk factors and portfolio gains and losses changes due to new or matured trades. Moreover, new information that occurs has to be taken into account for the derivation of the conditional distribution of the risk factors' returns $F_t(R_t \mid R_{t-1}, R_{t-2}, \ldots)$, for a current discussion of calculating the relevant gains and loss figures, see Finger (2005). As a financial institution commonly reserves economic capital for risks such as market risk, it is concerned with not reserving too much (which would be inefficient) or too little capital (thus underestimating the risks). Therefore, the adequacy of the predictive distribution is of high importance for the acceptability of a risk model in the bank to generate trading impulses for the correct management measurements.

Typical Modeling Errors

The above-mentioned setup to model market risk in a risk model helps us to identify typical errors due to modeling. Some of them stem directly from the introduced building blocks. When selecting risk factors for the model, we face the problem of parsimonious modeling. This means that not all risk factors can be taken into account within the model. Some of this stems from the "nature" of the "real" risk factor: the yield curve, for example, is continuous, implying that key rates have to be identified as risk factors. We want to include into the model only risk factors with liquid quotations. Finally, if we introduce too many risk factors, we face statistical (and potentially numerical) issues, because the noise associated with estimating a distribution for the risk factors increases and the explanation of the main risk drivers in the portfolio becomes more and more complicated.

Many common distributions for $F_t(R_t \mid R_{t-1}, R_{t-2}, \ldots)$ comprise the normal distribution. This may only be consistent with the stylized facts if the time series properties like heteroscedasticity are also taken into account. Nevertheless, many distributions used in the risk modeling context are not able to reflect rare events with high impact sufficiently. The estimation of the distribution comes along with a respective estimation error.

Concerning the mapping between risk factors and the portfolio's gains and losses, one has to admit that, in practice, approximations have to be incorporated as valuation models might be very complex on a product level and valuation under a certain scenario might be time consuming or involve many simulations, thus giving rise to numerical issues. The potential recalibration of valuation models is another issue to be considered. Thus, common approximations include linear or quadratic approximation, present value grids, or analytical functions instead of numerical simulations for valuation functions.

From a regulatory point of view, see Basel (1996), typical errors include the basic integrity of the model in situations where a bank's risk system is not capturing the risk, incorrect calculation/algorithm or implementation, the general necessity to improve a model's accuracy, insufficient precision in assessing the risk, inferential or statistical errors, and so on.

Having these toeholds in mind, we will later provide insights into how adequate means for improving the market risk model can be identified.

Results from Model Risk in a Market Risk Model

Next, we briefly discuss the impact of model risk on the risk management process. Model risk caused by the market risk model can impact the risk management process at all steps, which has to be taken into account. During risk analysis and identification, a poor model may not be able to reflect risk adequately or even completely fail to take into account an important source of risk. This can lead to missing impulses for risk management decisions, e.g., when the risk measure does not reflect the real risk by the resulting figures.

Potential under- or overestimation of risk on different levels might undermine the model's credibility and lead to an inefficient use of risk capital. Portfolio effects might be misleading and result in wrong decisions, resulting from either a false sense of security or insecurity.

The potential impact of model risk in a market risk model makes it obvious that a tough regular validation and backtesting process is a necessary requirement for the institution. This will be examined in a later section in this chapter.

A PRIORI VALIDATION

Risk Factor Selection

As discussed above, the implementation of the risk model faces the problem of parsimonious modeling that makes a reduction to key risk factors and risk factor groups necessary. This raises questions such as which key rates should be taken out of the yield curve or whether we are able to model different parts of a volatility surface, etc.

The selection of the relevant risk factors depends to a large extent on the strategy of the portfolio, as the following consideration shows: from a pure stochastic modeling perspective more than 95 percent of the yield curve moves can be explained by parallel shifts, twists, and butterflies, i.e., the first three eigenvectors. This enables us to reduce the number of risk factors dramatically. But, if on the other hand, our portfolio follows a steepening strategy between five and six years, we have the necessity to introduce these risk factors into the model even if they are highly correlated, since even small differences will show up in gains or losses.

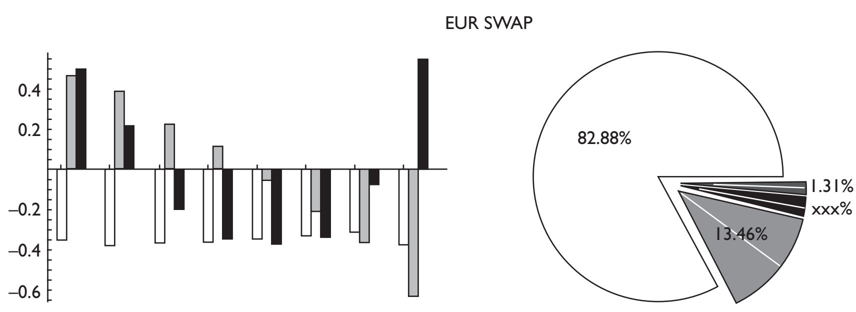
Hence, typical means to select risk factors rely on common statistical tools such as regression analysis for calculating the respective contributions of the individual risk factors to the portfolio's gains and losses, comparison of different sets of risk factors, accompanied by tests on homoscedasticity and autocorrelation.

A selection criterion is the respective impact of the key risk factors on the portfolio's gains and losses. Here, many well-known methods are ready available, but on the other hand, we make other assumptions such as a constant portfolio composition. Furthermore, a nonlinear regression might get highly complex.

A principle component analysis describes the relevance of different risk factors by eigenvalue decomposition and can test the significance of eigenvectors and eigenvalues. The selection is derived by the impact of the key risk factors to the risk factors' total variation.

The application of a principle component analysis to key interest rates, as in Figure 26.1, shows that the first eigenvector which is a parallel shift already contributes 83 percent to the total variation of the yield curve, the second eigenvector, a twist, contributes 13 percent, and the third eigenvector, a butterfly, contributes 2 percent. Thus, more than 98 percent of

Figure 26.1 Application of a Principle Component Analysis to the Interest Yield Curve



Data set from beginning 2008 to mid 2009. Euro Swap key interest rates 2, 3, 4, 5, 7, 10, 15, and 30 years. First three eigenvectors (left) and explanation of the principle component (right).

the total variation of the yield curve can be explained by three simple movements.

The principle component analysis is easy to implement and is a well-known procedure, but, as already mentioned in the example provided above, neglects the portfolio's decomposition.

As an extension, one could also think of a parametric approximation like a Nelson-Siegel parameter set in a yield-curve context or local volatility models. This is a well-established procedure in pricing context but the interpretation of the parameters and their distribution is not always intuitive and may give rise to numerical instabilities.

Risk Factor Distribution

Focusing on the modeling of the risk factors' distribution, we have to recall the model above. Keeping in mind that we are modeling a conditional distribution, we distinguish between the time series model and the distribution of the noise term.

Time series models used for market risk should incorporate the stylized facts. Some examples comprise white noise and (geometric) Brownian motion, ARCH models with stochastic volatility and variance $h_t = \omega + \alpha\varepsilon_{t-1}^2$, where ε_{t-1} is the residual term from the last time step, GARCH models with $h_t\ \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}$ and the special case of the exponential weighting scheme used by RiskMetrics (see Mina and Xiao, 2001).

The relation between risk factors and risk factor returns is modeled by either logarithmic returns (e.g., for equities), relative returns (e.g., for interest rates), or absolute returns (e.g., for spreads on interest rates). The respective selection depends on the risk factor and the observation that the noise term should consist of independent, identically distributed, random variables.

Common distributional assumptions for the noise term in the time series model are the normal distribution or student-t distributions. As tools for the validation of the composite distribution, ordinary statistical tests in combination with graphical means like *pp* or *qq* plots seem useful.

Transformation

The third part of the building block is the mapping between the risk factor (returns) and the portfolio's gains and losses. This mapping is usually done on an instrument level. Here, if the model (like historical simulation) allows a complete evaluation, no further approximations (besides potentially some parameters) are to be made. If due to complex valuation functions or numerical complexity, one has to establish an approximation, this can be done by a Taylor series expansion (thus by the "greeks"). The derivation is often done numerically by applying a small shift to the risk factor (differential quotient), thus requiring that an appropriate shift size be determined. Often, larger shifts seem on an individual basis plausible as the approximated values are located in the tails of the distribution. Other practices include a quadratic function by a three-point approximation, interpolation, use of present value grids, etc. The goodness-of-fit can be judged by a measure such as the L_2 metric (or L_1) between the true valuation function and the approximation.

A POSTERIORI BACKTESTING

By backtesting, we complement the validation of the main building blocks of the market risk model by a retrospective view. We compare the results of a portfolio's risk estimation with the respective gains and losses.

Theoretical Foundation for Backtesting Methods

Common statistical tests most often require observations stemming from independent and identically distributed (i.i.d.) random variables. The above described setup in market risk measurement emphasis that this is clearly not the case within the backtesting framework. This gives rise to the question of how to transform the given data to enable a further application of statistical methods and an in-depth analysis of how to improve the risk model.

Diebold, Gunter, and Tay (1998) present a basis to validate the predictions when observing realized values and consequently to heal the missing i.i.d. property. This approach dates back to an idea by Rosenblatt (1952) and uses the so-called Rosenblatt transformation:

Theorem 1: Let $\tilde{f}_{t|X_{t-1}, X_{t-2}, \ldots}$ be the conditional densities generating the time $(X_t)_{t \in \mathrm{N}}$ series and let f_t be the respective predictive densities. These predictive densities are assumed to be continuous and $f_t(x) > 0$ for all x. For the probability integral transform $U_t := F_t(X_t) = \int_{-\infty}^{X_t} f_t(u)du$ it holds that

$$P\left(F_t(X_t) \le \varsigma\right) = P\left(U_t \le \varsigma\right) = \int_0^{\varsigma} \frac{\tilde{f}_{t|X_{t-1},X_{t-2},\ldots}\left(F_t^{-1}(x)\right)}{f_t\left(F_t^{-1}(x)\right)} dx.$$

If further on $\tilde{f}_{t|X_{t-1}, X_{t-2}, \ldots} f_t$ for all t, then $F_t(X_t) = U_t \sim \mathrm{U}_{[0,1]}$.

Proof of Theorem 1: Because of the non-negativity and the continuity of the predictive density, it holds that $F_t(x) = \int_{-\infty}^{x} f_t(u)\,du$ is strictly increasing and continuous and hence, invertible. For $\varsigma \in [0,1]$ it follows that

$$P\left(U_t \le \varsigma\right) = P\left(F_t(X_t) \le \varsigma\right) = P\left(X_t \le F_t^{-1}(\varsigma)\right) = \int_{-\infty}^{F_t^{-1}(\varsigma)} \tilde{f}_{t|X_{t-1},X_{t-2},\ldots}(u)du$$

$$= \int_0^{\varsigma} \tilde{f}_{t|X_{t-1},X_{t-2},\ldots}(F_t^{-1}(x)) \cdot \left(F_t^{-1}\right)'(x)dx = \int_0^{\varsigma} \frac{\tilde{f}_{t|X_{t-1},X_{t-2},\ldots}(F_t^{-1}(x))}{f_t(F_t^{-1}(x))} dx$$

The second property follows directly from $P(U_t \le \varsigma) = \int_0^{\varsigma} 1dx = \varsigma$.

Now, we are in a position to draw conclusions regarding the time series' serial dependence. This is given by the next theorem.

Theorem 2: *With the same notation and assumptions as in*

Theorem 1, it holds that

$$P\left(U_t \le \varsigma_t, \ldots, U_1 \le \varsigma_1\right) = 0 \int_{[0,\varsigma_t]\times\ldots\times[0,\varsigma_1]} \frac{\tilde{f}_{t|X_{t-1},X_{t-2},\ldots}(F_t^{-1}(x_t))}{f_t(F_t^{-1}(x_t))} \cdot\ldots\cdot \frac{\tilde{f}_{1|X_0,X_{-1},\ldots}(F_1^{-1}(x_1))}{f_1(F_1^{-1}(x_1))} d\lambda_t .$$

I further on $\tilde{f}_{t|X_{t-1},X_{t-2},\ldots}$ *f_t for all t, then* $F_t(X_t) = U_t \overset{\text{iid}}{\sim} U_{[0,1]}$.

Proof of Theorem 2:[5] For $(\varsigma_t, \ldots, \varsigma_1)' \in [0,1]^t$ it follows that

$$P\left(U_t \le \varsigma_t, \ldots, U_1 \le \varsigma_1\right) = P\left(F_t(X_t) \le \varsigma_t, \ldots, F_1(X_1) \le \varsigma_1\right)$$

$$= P\left(X_t \le F_t^{-1}(\varsigma_t), \ldots, X_1 \le F_1^{-1}(\varsigma_1)\right)$$

$$\overset{(*)}{=} \int_{(-\infty,F_t^{-1}(\varsigma_t)]\times\ldots\times(-\infty,F_1^{-1}(\varsigma_1)]} \tilde{f}(u_t, \ldots, u_1) d\lambda_t$$

$$= \int_{[0,\varsigma_t]\times\ldots\times[0,\varsigma_1]} \tilde{f}(F_t^{-1}(x_t), \ldots, F_1^{-1}(x_1)) \cdot \begin{vmatrix} \frac{\partial F_1^{-1}}{\partial x_1} & \frac{\partial F_1^{-1}}{\partial x_t} \\ \frac{\partial F_t^{-1}}{\partial x_1} & \frac{\partial F_t^{-1}}{\partial x_1} \end{vmatrix} d\lambda_t$$

$$\overset{(**)}{=} \int_{[0,\varsigma_t]\times\ldots\times[0,\varsigma_1]} \tilde{f}_{t|X_{t-1},X_{t-2},\ldots}(F_t^{-1}(x_t)) \cdot\ldots\cdot \tilde{f}_{1|X_0,X_{-1},\ldots}(F_1^{-1}(x_1)) \cdot \frac{\partial F_1^{-1}}{\partial x_1} \cdot\ldots\cdot \frac{\partial F_t^{-1}}{\partial x_t} d\lambda_t$$

$$\overset{(***)}{=} \int_{[0,\varsigma_t]\times\ldots\times[0,\varsigma_1]} \frac{\tilde{f}_{t|X_{t-1},X_{t-2},\ldots}(F_t^{-1}(x_t))}{f_t(F_t^{-1}(x_t))} \cdot\ldots\cdot \frac{\tilde{f}_{1|X_0,X_{-1},\ldots}(F_1^{-1}(x_1))}{f_1(F_1^{-1}(x_1))} d\lambda_t$$

Equation (*) is derived by t the dimensional transformation theorem for integrals, equation (**) stems from the decomposition of random variables, and (as equation (***)) from the property that each predictive density f_i is measurable by the σ-field of $X_{t-1}, X_{t-2},\ldots$ and therefore, for $i < j$, it follows that

$$\frac{\partial F_i^{-1}}{\partial x_j} = 0,$$

whereas for $i = j$ it follows that

$$\frac{\partial F_i^{-1}}{\partial x_j} = \frac{1}{F_i'(F_i^{-1})}.$$

The second property follows directly from

$$P\left(U_t \leq \varsigma_t, \ldots, U_1 \leq \varsigma_1\right) = \int_{[0,\varsigma_t]\times\ldots\times[0,\varsigma_1]} 1\cdot\ldots\cdot 1 d\lambda_t = \varsigma_t \cdot \ldots \cdot \varsigma_1$$

We can derive a corollary that is helpful in the further context:

Corollary: *Let* $[a_t, b_t]$ *be an interval of the domain of the predictive distribution with* $F_t(b_t) - F_t(a_t) = \alpha$ *for all points in time t. If* $\tilde{F}_{t|X_{t-1}, X_{t-2}, \ldots} = F_t$ *for all t, it holds that* $1_{[a_t, b_t]}(X_t) \overset{\text{iid}}{\sim} \text{Ber}_{(\alpha)}$.

These results allow a treatment of the daily pairs of predicted distributions and realized gains $(F_t, g_t)_{t=1}^{T}$ over a period of T trading days and yields the possibility to step into mathematical and statistical backtesting methods for the validation of the risk model. The distribution $F_t(\cdot)$ is the abovementioned conditional predictive distribution, whereas $\tilde{F}_t$ is the nonobservable 'true' distribution of the portfolio's gains and losses.

The Rosenblatt transformation forces us to focus on the standardized gains $U_t := F_t(G_t)$ or if value at risk is the chosen risk measure on the exceedances $O_{t,\alpha} := 1_{(-\infty, -q_t^{\alpha})}(G_t)$. In an ideal risk measurement framework, due to the Rosenblatt transformation, the standardized gains U_t should be i.i.d. uniform distributed, i.e., $U_t \overset{\text{iid}}{\sim} U_{[0,1]}$, and the exceedances Bernoulli distributed according to the corollary, i.e., $O_{t,\alpha} \overset{\text{iid}}{\sim} \text{Ber}_{(1,\alpha)}$.

Useful Statistical Tests for Backtesting

This section introduces some statistical tests that can be used to backtest the risk model. For a deeper description of the tests and the derivation of a decision tree that can be used in a practical context, we refer to Wehn (2005, 2008).

As a direct consequence of the above-introduced Rosenblatt transformation, we will first examine statistical tests based on exceedances and then statistical tests based on standardized gains. Every statistical test is characterized by its null hypothesis, critical values, and a brief discussion of advantages and disadvantages.[6]

Tests Based on Exceedances

Probably the best-known statistical test for backtesting is the so-called traffic light approach (TLA) τ_{TLA}. This test was introduced in the context of first allowing internal market risk models for regulatory purposes (see Basel, 1996). The TLA has the null hypothesis of H_0: $(1-\tilde{\alpha}) \in [0,1-\alpha] \Leftrightarrow \tilde{\alpha} \geq \alpha$, meaning that it is judging whether the observed relative frequency of exceedances $\tilde{\alpha}$ is significantly lower than the required level of α. As the individual exceedances are to be Bernoulli distributed, their sum is binomial with $\sum_{t=1}^{T} O_{t,\alpha} \sim \text{Bin}_{(T,1-\alpha)}$ leading to a critical zone of $K_{T,\beta} = \{x \in \{0,1\}^T \mid \sum_{t=1}^{T} x_i \geq k_{T,B}\}$ with respective critical values of $k_{T,\beta} = \min\{m \mid 1 - F_{\text{Bin}_{(T,1-\alpha)}}(m-1) \geq \beta\}$.

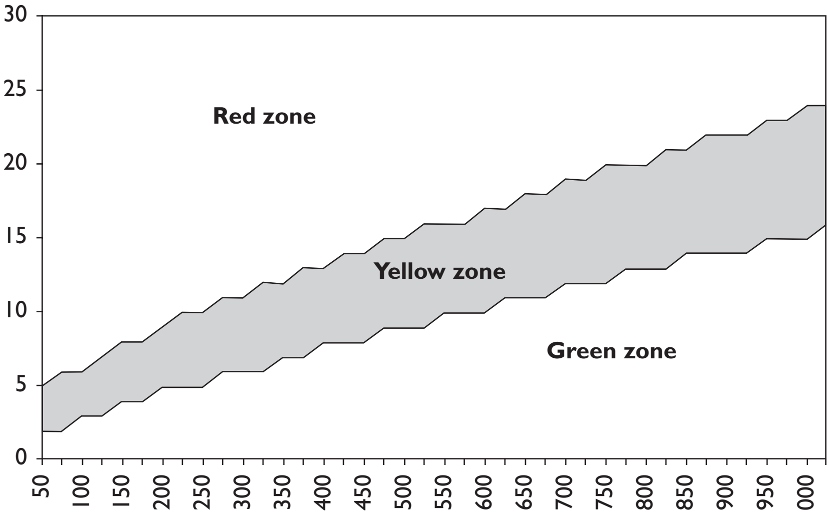
The Basel Committee on Banking Supervision in its 1996 landmark paper laid down two different values for β, namely $\beta_y = 0.05$ and $\beta_r = 0.0001$ for the first kind error. The interpretation then is as follows: if the null has to be rejected by β_y, then a yellow light is shown, if it has to be rejected by β_r, then a red light is shown. Entering the yellow or red zone is directly linked to higher regulatory capital requirements (see section "Regulatory Concerns") and can even lead to a refusal of the use of the internal risk model for regulatory purposes.

Figure 26.2 shows the critical values for different lengths T of the time series of exceedances for the two values β_r and β_y. For regulatory purposes, the length of $T = 250$ trading days is relevant.

A simple and understandable concept as well as setting incentives toward a conservative modeling make for the often discussed advantages of the TLA, whereas the last point could also be considered to be a disadvantage as there is no penalty imposed for overestimating the risks (a conservative risk model is not necessarily an adequate one). Exceedances are relatively rare events (especially on the regulatory level of $\alpha = 99\%$) leading to a low quantity of observations, whereas the values for β_r and β_y are rather high.

A whole class of tests is based on the likelihood ratio. This allows the statistical treatment of the exceedances as well as the whole distribution, as we will see below. Likelihood ratio tests (LRTs) are best uniform selective tests. Kupiec (1995) proposes a two-sided extension of the TLA null by H_0: $\tilde{\alpha} = \alpha$. He first concludes that according to theorems 1 and 2

Figure 26.2 Different Critical Values under the Traffic Light Approach



Values for different lengths of the observable time series for level 0.05 and 0.0001, including classification into green, yellow, and red zones.

(including the corollary), the statistic $t_1 := \min\{t \mid O_{t,\alpha} = 1\}$ is geometrically distributed, meaning that $P(t = t_1) = (1 - \tilde{\alpha}) \cdot \tilde{\alpha}^{t_1 - 1}$. The respective LRT for the "time until first failure" (TUFF) is given by

$$\lambda_{TUFF} = -2\ln\left(\frac{(1-\alpha)\cdot\alpha^{t_1-1}}{\frac{1}{t_1}\cdot\left(1-\frac{1}{t_1}\right)^{t_1-1}}\right).$$

It holds by Wilks' (1938) theorem asymptotically that $\lambda_{TUFF} \overset{asympt}{\sim} \chi_1^2$. Thus, if $\lambda_{TUFF} > F_{\chi_1^2}^{-1}(1 - \beta)$ for a given β, the null has to be rejected. The TUFF test is characterized by a relatively low discriminatory power and the need for large observation samples for small values of β and large values of α (for example, considerably more than 600 when using $\beta = 0.01$ and) $\alpha = 0.99$).[7]

Kupiec (1995) also proposes a second LRT with the same null. The test statistic for the "proportion of failure" test (POF) is given by

$$\lambda_{POF} = -2\ln\left(\frac{(1-\alpha)^{\tau}\cdot\alpha^{T-\tau}}{\left(\frac{\tau}{T}\right)^{\tau}\cdot\left(1-\frac{\tau}{T}\right)^{T-\tau}}\right),$$

where $t := \sum_{t=1}^{T} O_{t,\alpha}$ again is assumed to be binomially distributed. It holds again that $\lambda_{POF} \overset{\text{asympt}}{\sim} \chi_1^2$. This LRT corresponds to a two-sided binomial test, hence serving as an extension of the above-mentioned TLA. Like for TLA, the second kind error for the POF test can also be reduced significantly by using larger observation periods. Due to the two-sided property, the errors of the second kind are clearly higher than in the TLA, but the POF test consequently removes the disadvantage of the one-sided TLA.

The first LRT focusing on the serial dependence of exceedances based on the corollary is a statistic proposed by Christoffersen (1998). It takes into account a transition matrix of the kind

$$\begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}, \text{ where } p_{ij} := P(O_{t,\alpha} = j \wedge O_{t-1,\alpha} = i).$$

Independence of sequential exceedances results in $P_{ij} := (P_{0i} + P_{1i}) \cdot (P_{0j} + P_{1j}) = P(O_{t,\alpha} = j)P(O_{t-1,\alpha} = i)$. Accordingly, the null hypothesis is given by $(P_{01}, P_{11})' \in \{(x, x)' | x \in (0,1)\} = \Theta_0 \subset \Theta = (0,1)^2$ and with the definition of $t_{ij} := \Sigma_{t=1}^{T} 1_j (O_{t,\alpha}) 1_i (O_{t-1,\alpha})$ for $i, j \in \{0,1\}$ the number of observations capturing first state i and then state j, the respective statistic is given by

$$\lambda_{ind} = -2\ln\left(\frac{\left(1-\frac{t_{01}+t_{11}}{T-1}\right)^{(t_{00}+t_{10})}\left(\frac{t_{01}+t_{11}}{T-1}\right)^{(t_{01}+t_{11})}}{\left(1-\hat{p}_{01}\right)^{t_{00}}\hat{p}_{01}^{\;t_{01}}\left(1-\hat{p}_{11}\right)^{t_{10}}\hat{p}_{11}^{\;t_{11}}}\right), \text{with the estimator } \hat{p}_{i1} := \frac{t_{i1}}{t_{i1}+t_{i0}}.$$

For the respective LRT statistic, it follows that $\lambda_{ind} \overset{asympt}{\sim} \chi_1^2$. A very similar test is the χ^2 statistic with

$$\tau_{\chi^2(ind)} = \sum_{i=0}^{1}\sum_{j=0}^{1}\frac{\left(t_{ij}-\frac{(t_{i0}+t_{i1})(t_{0j}+t_{1j})}{T-1}\right)^2}{\frac{(t_{i0}+t_{i1})(t_{0j}+t_{1j})}{T-1}} \text{ and } \tau_{\chi^2(ind)} \overset{asympt}{\sim} \chi_1^2.$$

Christoffersen (1998) combines the POF test with the depicted LRT test on independence which results in a null hypotheses of $(\alpha, P_{01}, P_{11})' \in \{(\alpha, x, x)' | x \in (0,1)\} = \Theta_0 \subset \Theta = (0,1)^3$ and derives a LRT test

$$\lambda_{combined} = -2\ln\left(\frac{\alpha^{(t_{00}+t_{10})}\left(1-\alpha\right)^{(t_{01}+t_{11})}}{\left(1-\hat{p}_{01}\right)^{t_{00}}\hat{p}_{01}^{\;t_{01}}\left(1-\hat{p}_{11}\right)^{t_{10}}\hat{p}_{11}^{\;t_{11}}}\right), \text{ with } \lambda_{combined} \overset{asympt}{\sim} \chi_2^2.$$

It follows that $\lambda_{combined} = \lambda_{ind} + \lambda_{POF}$ (see Christoffersen, 1998). This test, by construction has a lower discriminatory power than other tests because of the joint null hypotheses. The test yields good results to differentiate between on average good predictions (the POF test would not reject this) and timely dynamic (i.e., heteroscedastic processes) predictions.

At this point, we can derive a natural extension to the binomial test by approximating the binomial distribution with a Poisson distribution with the estimator $\hat{\nu} := \sum_{t=1}^{T} O_{t,\alpha}$ for the intensity and its expected value $\nu = (1-\alpha) \cdot T$. The respective LRT statistics follows from

$$\lambda_{Poisson} = -2\ln\left(\frac{\frac{\nu^k}{k!}\exp(-\nu)}{\frac{\hat{\nu}^k}{k!}\exp(-\hat{\nu})}\right) = -2\ln\left(\frac{\left((1-\alpha)\cdot T\right)^{\sum_{t=1}^{T} O_{t,\alpha}}\exp\left(-(1-\alpha)\cdot T\right)}{\left(\sum_{t=1}^{T} O_{t,\alpha}\right)^{\sum_{t=1}^{T} O_{t,\alpha}}\exp\left(-\sum_{t=1}^{T} O_{t,\alpha}\right)}\right)$$

The behavior of the Poisson LRT is very similar to the POF test, stemming from the fact that the POF test assumes the binomial distribution and the Poisson LRT assumes the Poisson distribution.

Tests Based on the Entire Predictive Distribution

Tests based on exceedances can only give a first impression of the adequacy of the risk modeling due to the fact that they focus on the main parameter of the distribution, i.e., the respective quantile, the value at risk. As mentioned above, the aim is not only to predict a certain (conditional) parameter but rather to predict a whole conditional distribution of a portfolio's gains and losses. Hence, tests focusing on the whole distribution relying on theorem 1 and 2 come to mind. These tests comprise goodness-of-fit tests, LRTs as well as others.

Starting with the fact that $F_t(G_t) = U_t \overset{\text{iid}}{\sim} U_{[0,1]}$ when the right distribution $F_t(\cdot)$ is predicted, one can easily apply goodness-of-fit tests like the test by Kolmogorov-Smirnov and the statistic

$$\tau_{KS} = \sup_{0 \le x \le 1}\left|\frac{\sum_{t=1}^{T} 1_{[0,x]}(U_t)}{T} - x\right| = \max_{i=1,\dots,T}\left|\frac{i}{T} - U_{i:T}\right|$$

for the case of a $U_{[0,1]}$ distribution. Crnkovic and Drachman (1996) applied a certain version of the Kolmogorov-Smirnov test dating back to the Kuiper statistic with

$$\tau_{CD} = \max_{i=1,\dots,T}\left(\frac{i}{T} - U_{i:T}\right) + \max_{i=1,\dots,T}\left(U_{i:T} - \frac{i}{T}\right).$$

Whereas the Kolmogorov-Smirnov test is very sensitive around the median of the distribution, the test by Crnkovic and Drachman weighted the entire distribution almost equally. Crnkovic and Drachman also proposed a weighting scheme (such as) $w(x) = -\frac{1}{2}\ln(\cdot(x(1-x)))$ to set more weights to the tails of the distribution with a statistic

$$\max_{i=1,\dots,T}\left(w\left(U_{i:T}\right)\cdot\left(\frac{i}{T}-U_{i:T}\right)\right)+\max_{i=1,\dots,T}\left(w\left(U_{i:T}\right)\cdot\left(U_{i:T}-\frac{i}{T}\right)\right).$$

The distribution of the Kolmogorov-Smirnov test statistic as well as that of the Kuiper test statistic can be derived analytically. For the case of considering a weighting scheme the respective distribution will be derived by a numerical simulation. The disadvantage of the test by Crnkovic and Drachman is that it has to take into account a large number of realizations ($\geq$1,000).

A further well-known goodness-of-fit test is the χ^2 goodness-of-fit test that is given for the preceding case of a $U_{[0,1]}$ distribution by

$$\tau_{\chi^2} = \sum_{i=1}^{m}\frac{\left(\sum_{t=1}^{T}1_{\Delta_i}\left(U_t\right)-\frac{T}{m}\right)^2}{\frac{T}{m}}, \text{ where}$$

$$\Delta_i := \left[\frac{i-1}{m},\frac{i}{m}\right) \text{ for } i,\dots,m-1 \text{ and } \Delta_m := \left[\frac{m-1}{m},1\right]$$

for a number of m classes and $P(\Delta_t) = \frac{1}{m}$. This statistics is asymptotically χ^2 distributed with $\tau_{\chi^2} \overset{asympt}{\sim} \chi^2_{m=1}$.

Berkowitz (2001) proposes a modified test based on a transformation by the normal distribution and testing for autoregressive (AR(1)) properties of the transformed time series $Z_t := \Phi^{-1}_{(0,1)}(F_t(G_t))$. The likelihood function for a Gaussian noise in a AR(1) series follows by

$$L_{(Z_1,\dots,Z_T)}(\mu,\sigma,\rho)=\varphi_{\left(\frac{\mu}{1-\rho},\frac{\sigma^2}{1-\rho^2}\right)}\left(Z_1\right)\cdot\prod_{t=2}^{T}\varphi_{\left(\mu+\rho Z_{t-1},\sigma^2\right)}\left(Z_1\right),$$

and thus the LRT by

$$\lambda_{NT} = -2\ln\left(\frac{L_{(Z_1,\dots,Z_T)}(0,1,0)}{L_{(Z_1,\dots,Z_T)}(\hat{\mu},\hat{\sigma},\hat{\rho})}\right),$$

with the respective estimators $\hat{\mu},\hat{\sigma}$ and $\hat{\rho}$.[8] This test by Berkowitz with property $\lambda_{NT} \overset{asympt}{\sim} \chi^2_3$ only rejects deviations from the mean and variance of a distribution; it does not identify other distributions with a mean of 0 and a variance of 1. Also, it cannot recognize heteroscadesticity. Thus, several authors (including Dowd, 2004) recommend applying a test on the hypotheses of a normal distribution in addition to the test by Berkowitz.

To conclude, we provide a test for the serial independence property of the standardized returns U_t. Defining Δ_i as above, we define $\delta_{ij} := 1_{\Delta_i}(U_{t-1})1_{\Delta_i}(U_t)$ for $t = 2,\ldots,T$ as the realization of a standardized return first in class i and then in class j. This consideration yields, in conjunction with $\delta_{i\cdot} := \Sigma_{j=1}^{m}$ and $\delta_{\cdot j} := \Sigma_{i=1}^{m} \delta_{ij}$, the following $\chi^2_{(m-1)^2}$ distributed test statistics:

$$\tau_{\chi^2(ind2)} = \sum_{i=1}^{m}\sum_{j=1}^{m} \frac{\left(\delta_{ij} - \frac{\delta_{\cdot j}\delta_{i\cdot}}{T-1}\right)^2}{\frac{\delta_{\cdot j}\delta_{i\cdot}}{T-1}}$$

Obviously, the test $\tau_{\chi^2(ind)}$ is a special case of $\tau_{\chi^2(ind\ 2)}$ for only two classes $\Delta_0 = [0,\alpha)$ and $\Delta_1 = [\alpha,1]$.

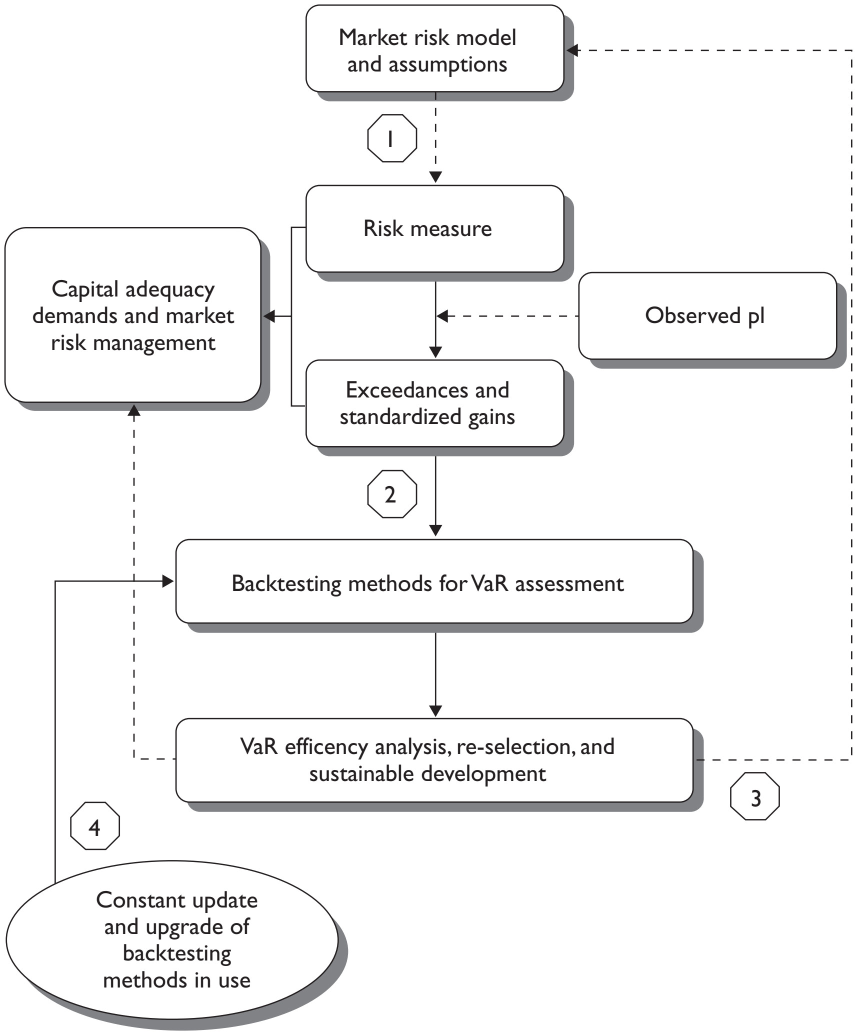
EMBEDDING VALIDATION AND BACKTESTING PROCEDURES

Regular Validation and Backtesting Process

During the validation and backtesting process, the bank regularly examines the adequacy and validity of the predicted risk values and to a certain extent the predicted (conditional) distribution of the portfolio's gains and losses. If any doubts about the adequacy exist, one can modify the model to better cope with the risks by re-examining the construction of the model, by either introducing new risk factors, changing distributional assumptions about the risk factor returns, or by a better mapping between returns and portfolio gains and losses.

A systematically formulated backtesting and validation process can be sketched by several steps (Figure 26.3): the definition of the model and the respective approach used to calculate the risk (step 1 in the figure), including the assumptions such as risk factor selection, distribution, and transformation. This setup is regularly updated through the backtesting and validation of results. The second step (2) is the procedure of backtesting itself. The risk model must be analyzed using a set of backtesting techniques. An ex post analysis takes place in the third step (3), leading to potential modifications in the risk model and the potential impact to capital adequacy is judged. Also, new backtesting methods have to be considered from time to time (step 4).

Figure 26.3 Embedding the Results of Backtesting in a Regular Validation and Backtesting Process (Adapted from Lehikoinen, 2007)



In practice, both parts, a validation part and a backtesting part should be proceeded on a regular basis. By the validation procedures as in section "A Priori Validation," we can improve the different building blocks individually, whereas in a backtesting context, as in section "A Posteriori Validation," the impact of the test results can become more complicated. Wehn (2008) derives a decision tree to improve the model gradually. Briefly, by some of the tests, we observe a misestimation (under- or overestimation) of risk (e.g., the TLA, the λ_{POF} or τ_{KS}, τ_{CD}, or τ_{χ^2}), which might be healed by inspecting the risk factor distribution. Other tests focus on the serial information (e.g., λ_{ind}, $\tau_{\chi^2(ind)}$, or $\tau_{\chi^2(ind\ 2)}$). This might be caused by missing information due to missing risk factors, a coarse modeling of time pattern, or a crude mapping between risk factor returns and the portfolio gains and losses.

Regulatory Concerns

Basel II (2009) formulates, besides several qualitative standards, the current requirements for a regular validation process for approved market risk models which requires, beyond the TLA tests, a demonstration that the assumptions made are appropriate (e.g., distribution, pricing models, etc.). The backtests should use hypothetical changes in portfolio value with end-of-day positions to remain unchanged using different confidence intervals tests on a subportfolio level. Among others, regulators require the use of hypothetical portfolios to assess structural properties like insufficient data histories where a mapping to proxies has to be done. A conduction of the procedures should be done on a periodical basis but especially when there have been any significant structural changes.

Market risk models approved by the regulatory authorities have to meet capital requirements, see Basel II (2009), that is determined by a qualitative and a quantitative factor. The quantitative factor is directly linked to the backtesting results from TLA. Thus, the regulatory capital requirements depend highly on the adequacy of the model, as Figure 26.3 already suggests.

CONCLUSION

This chapter illustrated and systemized the main modeling assumptions associated with a market risk model. The process for assessing the quality of the model is now twofold: on the one hand, the building blocks' assumptions can be validated individually and, on the other hand, the model's output can be judged by means of backtesting. Embedded in a regular process, the results can help to communicate the market risk model's strengthens and weaknesses with a view to steadily improving the adequacy of the model. This is a crucial step in every risk measurement and management cycle.

REFERENCES

Basel Committee on Banking Supervision. (1996) Supervisory Framework for the Use of "Backtesting" in Conjunction with the Internal Models

Approach to Market Risk Capital Requirements. Available at http://www.bis.org/publ/bcbs22.htm.

Basel Committee on Banking Supervision. (2009) Revisions to the Basel II Market Risk Framework. Consultative Document. Available at www.bis.org/publ/bcbs148.pdf.

Berkowitz, J. (2001) Testing Density Forecasts, with Applications to Risk Management. *Journal of Business & Economic Statistics*, 19(4): 465–474.

Campbell, S. (2005) A Review of Backtesting and Backtesting Procedures. Working paper, Federal Reserve Board, Washington, DC.

Christoffersen, P. (1998) Evaluating Interval Forecasts. *International Economic Review*, 39(4): 841–862.

Crnkovic, C. and Drachman, J. (1996) Quality Control. *Risk Magazine*, 9(9): 138–143.

Diebold, F.X., Gunter, T.A., and Tay, A.S. (1998) Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review*, 39(4): 863–883.

Dowd, K. (2004) A Modified Berkowitz Back-test. *Risk Magazine*, 17(4): 86ff.

Finger, Ch. (2005) Back to Backtesting. *RiskMetrics Monthly Research*, May.

Haas, M. (2001) New Methods in Backtesting. Working paper.

Kupiec, P.H. (1995) Techniques for Certifying the Accuracy of Risk Measurement Models. *Journal of Derivatives*, 2: 73–84.

Lehikoinen, K. (2007) Development of Systematic Backtesting Processes of Value-at-risk. Master's thesis, Helsinki University of Technology, Finland.

Mina, J. and Xiao, J.Y. (2001) Return to RiskMetrics: The Evolution of a Standard. Technical Document. Available at http://www.wu.ac.at/banking/sbwl/lvs_ss/vk4/rrmfinal.pdf.

Reiss, R.-D. and Thomas, M. (2007) *Statistical Analysis of Extreme Values, with Applications to Insurance, Finance, Hydrology and Other Fields*, 3rd ed. Basel, Switzerland; Boston, MA: Birkhäuser/Springer.

Rosenblatt, M. (1952) Remarks on a Multivariate Transformation. *Annals of Mathematical Statistics*, 23: 470–472.

Stahl, G., Wehn, C.S., and Zapp, A. (2006) Backtesting within the Trading Book. *Journal of Risk*, 8(2): 1–16.

Wehn, C.S. (2005) *Ansätze zur Validierung von Marktrisikomodellen: Systematisierung, Anwendungsmöglichkeiten und Grenzen der Verfahren*. Aachen, Germany: Shaker Verlag.

Wehn, C.S. (2008) Looking Forward to Back Testing. *Risk Magazine*, May.

Wilks, S.S. (1938) The Large-sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *Annals of Mathematical Statistics*, 9: 60–62.

NOTES

1. Dr. Carsten S. Wehn is head of market risk control at DekaBank, Frankfurt, Germany. Market risk control is responsible for the measurement of market and liquidity risk of the bank and the development of risk methods and models as well as the validation of the adequacy of the respective risk models. All opinions expressed herein are the author's and should not be cited as being those of his affiliated institutions. None of the methods described herein are claimed to be in actual use at DekaBank.
2. We only consider the discrete case of $t \in \{1,2,3,\ldots\}$ with a time step of one day.
3. For example in the popular RiskMetrics model (a special univariate GARCH (1,1) case with exponential weighting scheme), it holds that $R_t = h_t \varepsilon_t$ with a stochastic volatility of $h_t^2 = \lambda h_{t-1}^2 + (1-\lambda) R_{t-1}^2$ (cf. Mina and Xiao, 2001).
4. Usually, α is set to a high level of 0.95 or 0.99.
5. Reiss and Thomas (2007, p. 235) give a very short and concise proof for the Rosenblatt transformation based on the Fubini theorem. The interested reader is encouraged to have a look at this smart proof, too.
6. A comprehensive complementary discussion of backtesting techniques is given in Campbell (2005).
7. A potential extension of the TUFF test is given by Haas (2001), where supplementary to the time until the first exceedance $t_1 := \min\{t \mid O_{t,\alpha} = 1\}$ also the intervals between the individual exceedances $t_1 := \min\{t \mid \Sigma_{j=1}^{t} O_{j,\alpha} = i\} = t_{i=1}$ for $i := 2, \ldots \Sigma_{j=1}^{T} O_{j,\alpha}$ are taken into account. Under independence, it holds, that $P(t = t_i) = (1-\alpha_i) \cdot \alpha_i^{t_i - 1}$ and the respective LRT follows by

$$\lambda_{TUFF+} = -2 \sum_{i=1}^{\Sigma_{j=1}^{T} O_{j,\alpha}} \ln \left(\frac{(1-\alpha) \cdot \alpha^{t_i - 1}}{\frac{1}{t_i} \cdot \left(1 - \frac{1}{t_i}\right)^{t_i - 1}} \right) \text{ with } \lambda_{TUFF+} \overset{asympt}{\sim} \chi^2_{\Sigma_{j=1}^{T} O_{j,\alpha}}.$$

   Further on, this LRT can be combined with the POF test leading to $\lambda_{mix} := \lambda_{TUFF+} + \lambda_{POF} \overset{asympt}{\sim} \chi^2_{\Sigma_{j=1}^{t} O_{j,\alpha}+1}$. As theses statistical tests do not lead to significant further or complementary insights, they are not mentioned in the rest of the chapter.
8. The test by Berkowitz can easily be extended to a test on autocorrelation by the following statistics:

$$\lambda_{NT(uncorr)} = -2 \ln \left(\frac{L_{(Z_1,\ldots,Z_T)}(\hat{\mu}, \hat{\sigma}, 0)}{L_{(Z_1,\ldots,Z_T)}(\hat{\mu}, \hat{\sigma}, \hat{\rho})} \right)$$

PART VII

ECONOMIC CAPITAL AND ASSET ALLOCATION

This page intentionally left blank

CHAPTER 27

MODEL RISK EVALUATION BY INDEPENDENT REVIEWS

Katja Pluto[1]

ABSTRACT

While bank internal validation of rating systems or market risk value-at-risk models are an established practice in many banks, and has to a significant degree been driven by regulatory requirements, few banks have rigorously validated their economic capital systems. The chapter describes how validation of economic capital models can reduce and manage model risk in practice. It examines the process, pitfalls, and benefits of internal validation, and its embedding into the overall risk strategy of a large, globally diversified, banking organization.

INTRODUCTION

Model risk does not only express itself in quantifiable confidence levels around parameters or risk capital figures. It is probably as much, if not more, driven by model assumptions, data quality, IT implementation, and correct use of the model.

This chapter evaluates how qualitative reviews of risk models can help to mitigate model risk. We discuss and assess the determinants that make the process successful in mitigating model risk, which prerequisites need to be taken, and how reviews should be embedded into the overall governance structure of a bank. Many of our findings will sound only too familiar to model owners and developers; however, we find that they are consistently underrepresented in the academic literature on validation and model risk.

Our findings relate in principle to all risk models of a bank, be they "back-office"/risk controlling or "front-office"/pricing models, and be they supervisory recognized value-at-risk (VaR) or internal ratings based approach models under Pillar 1 of Basel II or internal models for economic capital modeling. Model reviews will differ in detail with regard to supervisory requirements and/or constraints; the general principle of mitigating model risk by rigorous reviews and validation is however independent of supervisory recognition.

The remainder of the chapter is structured as follows: The second section explains the regulatory frame; and the third section explains internal model governance and how it can mitigate model risk. This chapter's fourth section discusses the organizational setup and scope of model reviews and model validations, and requirements with regard to independence. The fifth section gives an overview of validation techniques; and the final section concludes with lessons learnt about the most important determinants in managing and mitigating model risk via independent model reviews.

REGULATORY REQUIREMENTS AND "FIT FOR PURPOSE" CHARACTERISTICS OF MODELS

Model reviews and validation are formally required for all bank internal models that determine minimum capital requirements under Pillar 1 of the new Basel II Accord. These are VaR models and incremental default risk charge models for market risk, internal rating systems for credit risk, and advanced measurement approaches for operational risk. Other bank internal risk models, such as pricing models or Pillar 2 economic capital models, are not required to be validated.

From a bank internal perspective, the distinction appears however artificial: risk management units should ensure that all models used for commercial decisions and/or risk and capital management are fit for purpose, that they are used for their intended purpose only and that they fit within the risk landscape and business strategy as a whole.

We therefore do not distinguish between "regulated" and "unregulated" models in the following, and do not refer to any specific supervisory requirements. The benchmark for mitigating model risk via review and validation is the model's "fit for purpose" assessment.

Review and validation are closely linked to the intended use of the model, and might vary according to this use. Valuation models for the trading book, as an example, require a high degree of accuracy—traders cannot afford to err on either the conservative or aggressive side, without pricing themselves out of the market and/or bearing losses in the long run. Economic capital models, on the other hand, try to assess a very conservative scenario (≥99.9 percent quantile), which is very likely never to occur in reality. Economic capital is seen as a buffer against such adverse events. Erring a bit on the conservative side is thus much less dangerous and will probably not lead to major findings in a validation or review.

MODEL GOVERNANCE AND MANAGEMENT OVERSIGHT

The biggest model risk, as the current market crisis has shown, does not lie in statistical uncertainty in the sense of confidence intervals around estimators or similar. It lies in

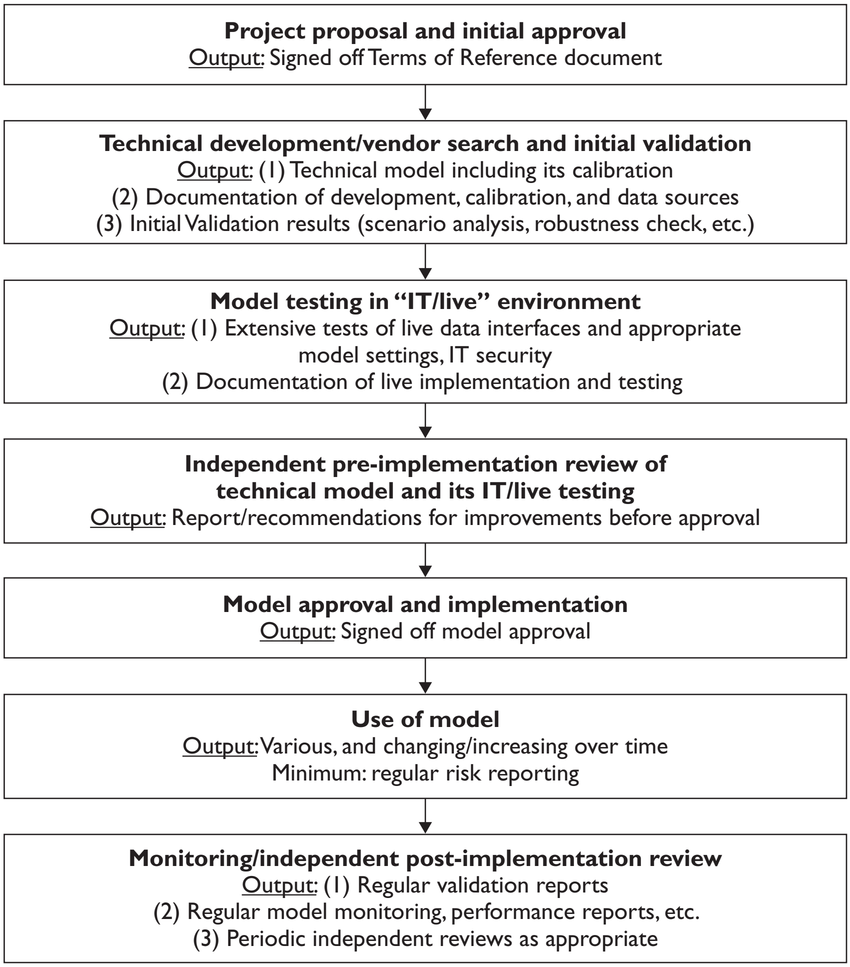
- Poor governance processes that open the door for model misuse, flawed model inputs and/or a missing understanding of the assumptions and limitations of a model; or
- Major risks not identified and not modeled at all

A comprehensive governance process can reduce these model risks significantly, even though it might not mitigate them completely. The process illustrated in Figure 27.1 has proven useful in practice and ensures an appropriate management oversight over risk models and will thereby minimize model risk.

Terms of Reference and Initial Model Approval

The terms of reference for each model should include a short description of the purpose of the model, its scope and relation to other risk models, the proposed methodology, and its envisaged use in risk management. The approving risk committee thereby receives an understanding of how the model fits into the overall risk landscape of the bank and can assess whether all significant risks of the business are covered. From a management oversight perspective, it is this coverage of all significant risks by some potentially rather simple models that provides a better assurance against model risk than sophisticated methodologies for selected risks only.

Figure 27.1 Model Governance Process



Technical Model Development

Data

The data used for the model development should be relevant, complete, and current. Model risk sets in if the data available fall short on any of these characteristics. Especially the first property might face severe limitations in practice that model developers have to live with. Examples include:

- The use of traded assets data to calibrate correlations for nontraded assets within credit risk models, such as the standard Merton model. Equity prices from listed companies certainly do not describe the dependency structure of nontraded borrowers completely, but within a Merton model context, trading data is needed as an input, such that the approximation of nontraded borrowers by some "similar" traded companies seems the only way to go. It is a shortcoming that model developers and users should be aware of.

- The use of credit default swap (CDS) spreads to calibrate credit risk includes liquidity premiums into the credit risk charges, which might or might not be appropriate. Again, model developers need to assess whether CDS spreads are "relevant enough" for the question they like to answer, or whether there are other, more relevant data available.

The use of data that do not exactly match the model is almost always driven by the fact that better data are not available. It is the model developer's responsibility to assess whether the deviation is justifiable under model risk aspects. At the same time, requirements on the relevance of data should be seen in perspective: each model is a simplification of reality, and no data will ever fit perfectly. The right balance between accuracy and availability needs be struck and explained to model users.

Model Assumptions

Each model comes with a set of implicit and explicit assumptions. These assumptions might or might not hold in practice; they are more often than not necessary to make the model work. They are not to be seen as a disadvantage per se—models are always simplifications of the reality.

Model assumptions should be clearly stated and assessed with regard to their impact on the model results. Special care should be given to implicit assumptions, which all too often get forgotten too easily. The following list provides a sample of common implicit assumptions:

- *Use of correlations:* correlations are linear and average measures of dependency. They cannot capture nonlinear dependencies or sudden changes in data history. Moreover, they tend to critically depend on the calibration data period and can be highly unstable over time. Ignoring these characteristics of correlations, and using them as the one and only dependence measure in complex models, can lead to severe misspecification.
- *Selection of relevant data set:* the decision on the time period from which calibration data are drawn is as important as the methodology choice itself. By setting out for a certain period, one implicitly assumes that the pattern of this time series will likely to be repeated within the forecasting horizon of the model. Data availability aside,

the choice is nothing else than a judgmental call on the relevant time period, which should be seen as such and be justified within the model development documentation.

Model assumptions and model limitations are the two sides of the same coin. Looking back at the market turmoil of 2008, and the subsequent criticism of quantitative finance and model-based risk management, much of it came from a lack of understanding of these assumptions and limitations on model users' and critics' side—and a lack of explanation and transparency of them on the model developers' (the "quants") side. Model developers do have the responsibility to be transparent about their explicit and implicit model assumptions and their judgmental decisions in developing the model.

Where possible, model developers should undertake scenario tests in the sense of using different sets of assumptions in the development process. Choices might be quite limited in reality—more often than not are they driven by data availability and/or methodological or computational restrictions. The assumption of normal distributions for market risk models, for example, has its origins as much in the well-behaving stochastics of the distribution as in any other reason.

Model Specification

The calibration data will be used for determining the relevant determinants and risk drivers for the model, and to calibrate the parameters. Model risk can be quantified within this step: all parameter estimations should be accompanied by confidence intervals around them—a standard statistical technique that is surprisingly seldom used in finance and for risk models. Confidence intervals around estimations would give an indication of how reliable the estimates are.

Moreover, they could be used for sensitivity and robustness analysis (i.e., evaluation of model results if confidence bounds instead of the estimated parameters were used). Such analysis gives a comprehensive sense of model risk in the statistical sense. If the "stressed" model results (in the meaning of results obtained by using the confidence bounds) are close to the initial model results, statistical model risk is small. Large deviations indicate a significant model risk. They could be driven by either large confidence intervals around the estimations (due to small data sample, wide dispersion, or general lack of model fit) or by a high significance of the parameter for the overall model. In either case, the model risk needs to be taken seriously, and alternative methodologies, different model specifications, and/or better and richer data might be sought.

Implementation

IT Specification

The IT implementation of each model poses several challenges to each risk model, and contributes probably more significantly to model risk than statistical noises in data or estimation methodologies. Things that can go wrong are numerous and very often driven by misunderstandings between IT and model developers. Examples include:

- Incorrect data feeds/interfaces, e.g., local currency versus accounting currency for the banking group
- Inconsistent data feeds from different source systems, e.g., zero-bond yields versus "normal" bond yields
- Incorrectly implemented methodologies

A good IT specification from the model developer is vital; a close coordination between IT and model developer during the implementation phase is desirable. In any case the IT implementation should be extensively tested by model developers and model users. Tests include replication of test cases as well as user friendliness and robustness to prevent follow-up issues with unintended incorrect model use.

Input Data

The model developers need furthermore to ensure that the model can be regularly run, i.e., that the required input data are available. The requirement sounds simple, however, getting calibration data in a one-off exercise and assuring regular and automated input data feeds into the model can be very different things in practice.

Documentation

The entire model development and use process should be adequately documented, including detailed user manuals. Documentation should be at a level such that a third person experienced in the specific model category would be able to understand and if necessary replicate or improve the model. The user manual should be at a level that allows business users to understand the major model characteristics, its application and limitations, and the required inputs, and to guide them through the output analysis.

Missing documentation is a material source of model risk: it leads to inadequate applications of models and in its worst case to models being employed after the developer has left, with nobody understanding the whereabouts of the model, which effectively becomes a black box.

Approval

No model should be used without explicit approval by the designated risk committee. The approval procedure is, similarly to the initial approval for model development, a qualitative measure to ensure that the model fits into the overall model landscape of the bank, that its development and results are consistent with the intentions at the start of the development and that the model owners and model users are confident with the usage in practice.

Approval should only be given after the model has been validated and independently reviewed.

Monitoring

Once the model is in use, the results should be regularly monitored. Standard practices are, e.g., backtesting and outlier analysis for market risk and overwrites for internal rating systems. The monitoring is not to be seen as a regulatory impetus, especially for backtesting, but rather as a way to detect shifts in model performance as early as possible. These shifts can come from various sources: a changing economic environment could render the old calibration invalid; the portfolio composition might have changed and requires a recalibration or redevelopment, or other.

Whether a bank sets predetermined thresholds that trigger action, or whether analysis and potential recalibration or redevelopment are decided on a case-by-case basis, seems of secondary importance. The monitoring and the extensive discussion of the monitoring results provide the process for dealing with model risk under portfolio dynamics.

VALIDATION VERSUS INDEPENDENT REVIEW

Throughout this chapter, we distinguish between the notion of an independent review and a model validation. Validation is used for a more narrow, highly model-based, and probably statistically/quantitatively dominated exercise. It will include out-of-sample tests of the model in question and a number of additional techniques.

Model validation should be performed by the model developer, i.e., it is formally not independent from the model development. A first initial validation should be due directly after the model specification. Additional validations, or rather result replications/testing, is required after the IT implementation of the model. Finally, each model must be monitored periodically in order to ensure its ongoing fit-for-purpose property. Standard tools are

- Number and direction of model overrides—how often do model users believe the model to deliver incorrect results
- Comparison of model estimations and risk realizations[2]

Independent review, on the other hand, comprises the review of the entire model development process. Reviewers will assess:

1. The formal setup (terms of reference, assigned responsibilities)
2. Model calibration data quality and relevance for the model
3. Model design, parameter estimations, and ultimate model setup ("makes sense" test)
4. Model validation

Where applicable, reviews should extend into model implementation, in particular:

5. Whether the model is indeed used for the purposes and asset classes it has been build for
6. Whether the IT implementation matches the intentions of the model developers, and that it had been tested accordingly

The independent review, as the name indicates, should be performed by a unit independent of the model owners and developers. Independence can be achieved in different ways, whose pros and cons are explained in Table 27.1.

There is no single preferred arrangement for independent reviews; banks must assess the strengths and weaknesses of each arrangement in the light of their specific circumstances.

Each model should be independently reviewed before it can be internally approved and used for its designated purpose. The independent review will give model owners, approvers, users, and senior management the assurance that model risk—both in the statistical sense as well as the more profane implementation risk—has been mitigated to the maximum possible extent.

INDEPENDENT REVIEW AND VALIDATION TECHNIQUES

(Statistical) validation techniques are manifold and depend on risk category, available data, model type, and many other factors. There is extensive literature and supervisory requirements, which are referred to in the literature

Table 27.1 Independent Review Arrangements

	Advantages	Challenges
Independent review teams	• Independence assured • Good availability if review needs arise	• "Losing touch with reality"—unrealistic recommendations and/or missing important model risks • Attractive to highly qualified staff?
Internal peer review	• Highly qualified staff • Realistic recommendations and good knowledge of model risks • Internal knowledge transfer	• Independence potentially scrutinized by cross-review arrangements • Potentially restrained resource availability if review needs arise
External reviewers	• Availability • Independence assured • Potentially wide industry experience	• External knowledge transfer • Independence assured How are follow-up/cross assignments with the external agency managed? • Costs
Audit	• Independence assured	• Detailed quantitative background could be a potential issue

list at the end of this article. As we intend to assess model risk and its mitigation from a more comprehensive, qualitative and process driven perspective, we do not go into detail.

Rather, the section focuses on independent review techniques that take model risk sources "around" the core model methodology into account. Table 27.2 provides an extensive, albeit nonconclusive list of independent review techniques, and their advantages and challenges.

It is up to the model reviewer to decide which techniques are appropriate for each model in question. We do not recommend complete suites of tests that are to be applied without assessment of their applicability; such suites provide a spurious sense of objectivity. Each test in itself comes with its baggage of implicit and explicit assumptions, which might or might not be appropriate for the specific model. Model reviewers should have the experience and capability, as well as the freedom, to decide what is appropriate in each case. The decision should, as everything related to model governance, be transparent and well documented.

Table 27.2 Validation and Review Techniques

Technique	Advantages	Challenges	Application
Critical assessment of model assumptions	• High impact of model results and model risk	• Alternative assumptions often computationally unfeasible	• All risk models
Qualitative review of methodology	• Challenge of methodology	• No quantitative assessment of model risk	• All risk models
IT system validation including data feed	• High impact on model risk	• Time consuming	• Selected parts of all risk models
Use test	• The only true reality check	• More valuable to external than internal addressees	• All risk models
Input data validation	• High Impact on model risk	• Time consuming	• Selected parts of all risk models
Parameter validation and sensitivity testing	• Core methodology validation	• Difficulties with vendor models	• All risk models
P/L attribution	• Easy	• Indication about directional accuracy only	• Risk categories with clearly attributable P/L
Benchmarking	• Comfortable feeling of not being the outlier	• Shows differences between model results only • No clear statement about which model is the correct one • Compare like and like? • Data availability	• Very selectively only
Model replication	• Full methodology, data, and calibration validation	• Time and resource consuming	• Very selected pieces of EC models only
Backtesting	• Only truly statistical validation	• Data availability	• Market risk only

CONCLUSION

Model risk can be mitigated through strong model development, validation, and independent review governance. It goes beyond pure statistical measures of risk and takes processes around the model into account. Particular attention should be given to all implicit and explicit assumptions that go into each risk model, including the calibration data sample, whose selection in itself forms an assumption about historic periods deemed relevant to model future patterns. Other significant sources of model risk are the implementation and use of model. The embedding of all risk models into appropriate senior management approval processes, whether supervisory prescribed or not, mitigates model risks by ensuring adequate embedding into the overall risk landscape of a bank.

Resources to mitigate these different sources of model risk need to be allocated on a case-by-case basis. The core model methodology, while important in itself, forms only one part of the overall equation for model risk and should not be overstated. Experience from bank internal governance processes indicates that model risk is more likely to be caused by the incorrect implementation, data, and model use, rather than methodology.

REFERENCES

Basel Committee on the Global Financial System. (2000) *Stress Testing by Large Financial Institutions: Current Practice and Aggregation Issues*. Basel, Switzerland: Bank for International Settlements. Available at: http://www.bis.org/publ/cgfs14.pdf?noframes=1.

Blochwitz, S., Hamerle, A., Hohl, S., Rauhmeier, R., and Rösch, D. (2005) Myth and Reality of Discriminatory Power for Rating Systems. *Wilmott Magazine*, January: 2–6.

Blochwitz, S. and Hohl, S. (2006) Validation of Banks' Internal Rating Systems: A Supervisory Perspective. In: Engelmann, B. and Rauhmeier, R. (eds.), *The Basel II Risk Parameters: Estimation, Validation and Stress Testing*. New York: Springer; 243–262.

Blochwitz, S., Martin, M.R.W., and Wehn, C.S. (2006) Statistical Approaches to PD Validation. In: Engelmann, B. and Rauhmeier, R. (eds.), *The Basel II Risk Parameters: Estimation, Validation and Stress Testing*. New York: Springer; 289–306.

Burnham, K.P. and Anderson, D.R. (2002) *Model Selection and Multi-model Inference: A Practical Information-Theory Approach*. New York: Springer.

Engelmann, B. (2006) Measures of a Rating's Discriminative Power: Applications and Limitations. In: Engelmann, B. and Rauhmeier, R. (eds.), *The Basel II Risk Parameters: Estimation, Validation and Stress Testing*. New York: Springer; 263–288.

Greene, W.H. (2003) *Econometric Analysis*. London: Pearson.

Hamerle, A., Rauhmeier, R., and Rösch, D. (2003) Uses and Misused of Measures for Credit Rating Accuracy. Working paper, University of Regensburg, Germany.

McNeil, A., Frey, R., and Embrechts, P. (2005) *Quantitative Risk Management*. Princeton, NJ: Princeton University Press.

Peter, C. (2006) Estimating Loss Given Default: Experiences from Banking Practice. In: Engelmann, B. and Rauhmeier, R. (eds.), *The Basel II Risk Parameters: Estimation, Validation and Stress Testing*. New York: Springer; 143–176.

Rauhmeier, R. (2006) PD-Validation: Experience from Banking Practice. In: Engelmann, B. and Rauhmeier, R. (eds.), *The Basel II Risk Parameters: Estimation, Validation and Stress Testing*. New York: Springer; 307–346.

Siegel, S. and Castellan, N.J. (1998) *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.

Tasche, D. (2005) Rating and Probability of Default Validation. In: Basel Committee on Banking Supervision (BCBS), *Studies on the Validation of Internal Rating Systems*. BCBS Working paper 14, 28–59.

NOTES

1. Katja Pluto heads the Risk Methodology unit at HSBC Holdings plc. The opinions expressed in this chapter are her own and do not necessarily represent the views of HSBC.
2. We do not use the notion of "backtesting" here, as it is linked to a statistically viable hypothesis testing. The length of the required time series is usually only available for market risk, where daily data are available. For other risk categories, annual or at most monthly data prevail, and the required time series would never be reached.

CHAPTER 28

ASSET ALLOCATION UNDER MODEL RISK

Pauline M. Barrieu and Sandrine Tobelem

ABSTRACT

In the present chapter, we propose a robust portfolio allocation methodology when there is some ambiguity concerning the dynamics leading asset prices. The decision maker considers several prior models for the asset price dynamics and displays an ambiguity aversion against those priors. We have developed a two-step ambiguity robust methodology to compute the portfolio optimal weights that offers the advantage to be more tractable and easier to implement than the various approaches proposed in the literature. This methodology decomposes the ambiguity aversion into a model-specific absolute ambiguity aversion as well as relative ambiguity aversion across the set of different priors. The weights inferred by each prior are transformed through a generic Absolute Ambiguity Robust Adjustment (AARA) function ψ. Then, the optimal transformed weights are mixed through a Relative Ambiguity Robust Adjustment (RARA) function π that reflects the relative ambiguity aversion of the investor for the different priors considered.

INTRODUCTION

In this chapter, we propose a new approach to account in a robust way for model ambiguity aversion in the asset allocation problem: the Absolute Ambiguity Adjustment Allocator. Our motivation to propose a simple methodology to account for ambiguity aversion is essentially due to the complexity to solve the asset allocation problem under ambiguity proposed so far in the literature (see for instance the smooth penalty optimization problem proposed by Klibanoff, Marinacci, and Munkerji, 2005).

Our methodology offers a trade-off between robustness and optimization. Our objective is not to find the optimal decision for the decision maker (for a given choice criterion), since the complexity of most practical frameworks makes this task almost impossible. We rather focus on an approach that allows the decision maker to combine different priors in a practical and tractable way to take the best decision in a robust sense. The question is really about finding a robust solution that encompasses all the different pieces of information given by the different priors but also the ambiguity the decision maker is facing regarding the set of priors she considers.

The chapter is organized as follows: first, we give a general background for an investor facing a portfolio optimization problem under model ambiguity, we then present the absolute ambiguity robust adjustment (AARA) function ψ, which transforms the optimal weights computed under each prior considered, according to the idiosyncratic ambiguity aversion the investor displays for each prior. In the third section, we introduce the relative ambiguity robust adjustment (RARA) mixture measure π that accounts for the systematic ambiguity aversion of the different priors considered. The function π allows us to mix the individual optimal weights obtained in the precedent phase through the function ψ. We finally present a more complex theoretical example that illustrates our AARA methodology and exhibits how the portfolio allocation is affected by ambiguity aversion.

SETTINGS

We consider an investor with a given initial wealth. She wants to allocate her wealth among the $N + 1$ different assets available in the market. X^{ϕ} represents the value of her portfolio at a future time horizon and the control variable ϕ represents her strategy, i.e, how she allocates her wealth among the assets. More precisely, ϕ is a vector of weights where each component corresponds to the proportion of wealth the investor is allocating to a given asset, a negative value translating the fact that this particular asset is sold. Each element of ϕ belongs to $[-1{:}1]$. Note that there is an investment constraint: $\sum_{i=0}^{N} \phi^i = 1$ to express the idea that 100 percent of the initial wealth is invested.

We assume the investor considers several different models, or priors, to represent the dynamic of X^{ϕ}. The investor is ambiguous against those models.

We propose a new methodology to solve the allocation problem under ambiguity. More precisely, we proceed typically in two steps as we distinguish two forms of ambiguity:

- *Absolute ambiguity:* This refers to the ambiguity aversion the agent has for a given prior. It is taken into account through the transformation function ψ, denoted thereafter as the AARA function. The function ψ is applied to the different optimal weights computed per prior model Q. The function ψ reflects the specific ambiguity aversion of the investor toward each prior Q. It solely modifies the optimal weights computed previously as if Q were the true model. Therefore, this function ψ transforms the different optimal weights of each model independently. More precisely, we compute the optimal asset allocation ϕ^{Q} associated with each model Q as if Q was the true model. Then, we compute the adjusted weights as the expected value of the weights transformed by a robust ambiguity aversion transformation ψ, parameterized by a prior-specific ambiguity parameter γ^{Q}.
- *Relative ambiguity:* This refers to the relative ambiguity aversion the agent has for the set of priors Q. It is taken into account through the distribution π defined on the set of priors, that we denote thereafter as the RARA function. The function π represents the relative ambiguity aversion of the investor among the priors considered. We call it relative, because π takes into account the relative ambiguity aversion of the investor toward the different priors. The function π also reflects the overall ambiguity aversion of the investor toward the set of priors, i.e., how much the investor believes that the set of priors contain priors close to the true model.

  The final Ambiguity Robust Adjustment Allocation ϕ^{ARA} is then defined as: $\phi^{ARA} \equiv \Sigma_Q \Psi(\phi^Q, \gamma^Q)\, \pi(Q)$. Our methodology offers several advantages.
  - *Tractability:* We can make the distinction between the absolute ambiguity toward a given model, and a relative ambiguity among all the models selected by the investor.
  - *Simplicity:* The portfolio robust weights under ambiguity can be easily computed, whereas other portfolio optimization methodologies are too complex to allow for closed-form solutions and require numerical solutions.
  - *Flexibility:* The different models selected by the investor do not have to belong to any particular parameterized family. We free ourselves from the Gaussian settings, mostly used in empirical papers (where it is more or less always assumed that the stock returns are log-normal).

In the next two sections, we present more precisely the AARA and RARA functions.

ABSOLUTE AMBIGUITY ROBUST ADJUSTMENT

In order to characterize the AARA function ψ, we first present some axiomatic characterization of the AARA function. Second, we focus on the parameterization of the absolute ambiguity parameter γ^{Q}.

Axiomatic Characterization of the AARA Function ψ

The aversion of the investor toward ambiguity is taken into account in our model by a concave transformation function ψ of the optimal weights obtained for each model Q. This ψ adjustment is model specific and rescales the optimal weights inferred by each model Q. More precisely, ψ deals with the specific ambiquity aversion of the investor toward each prior.

Let us first recall some of our notations:

- *Set of priors Q:* We denote by **Q** the set of priors, or models. We suppose that **Q** is a finite set, for the sake of simplicity in our argument.[1]
- *Prior Q:* We denote by Q an element of **Q**.
- *Prior optimal weight* $\phi^{\mathbf{Q}}$**:** We denote by ϕ^{Q} the vector of optimal weights obtained under the assumption that Q is the true probability measure.

The function ψ must satisfy some key properties to be consistent with the rationality of the agent:

- *Universality of the function ψ:* ψ is the same for all the priors. The investor treats the absolute ambiguity aversion with the same type of transformation across all the different priors. What distinguishes the absolute ambiguity aversion transformation across the priors is the specific ambiguity aversion parameter the investor attributes to each prior. ψ is parameterized with the positive absolute ambiguity aversion parameter γ^{Q}, that is model specific.
- *Maximum weight:* The ambiguity aversion parameter γ^{Q} defines the maximum weight obtained after the ψ transformation of the optimal weights defined under the prior Q. As the absolute optimal weights ϕ^{Q} are bounded by 1, $\psi(1, \gamma^{Q})$ represents the maximum weight the investor will assign to an asset after the AARA transformation.
- *Monotonicity:* One of the key characteristics of ψ is its monotonicity property. ψ preserves the relative order of the optimal weights ϕ^{Q}

deduced by a given prior Q, so that the relative preference of the investor toward the different assets given a prior Q is preserved through the transformation ψ.

- *Convexity:* The function ψ is concave on [0;1] and convex on [−1;0], so that the function ψ reduces more the absolute biggest weights given by the optimized portfolios under each prior considered. The convexity scale is parameterized through an aversion coefficient γ^Q: the bigger the aversion coefficient, the more averse the investor is to large weights inferred by Q. Also the convexity of ψ will depend upon the investor ambiguity aversion toward the model Q considered.

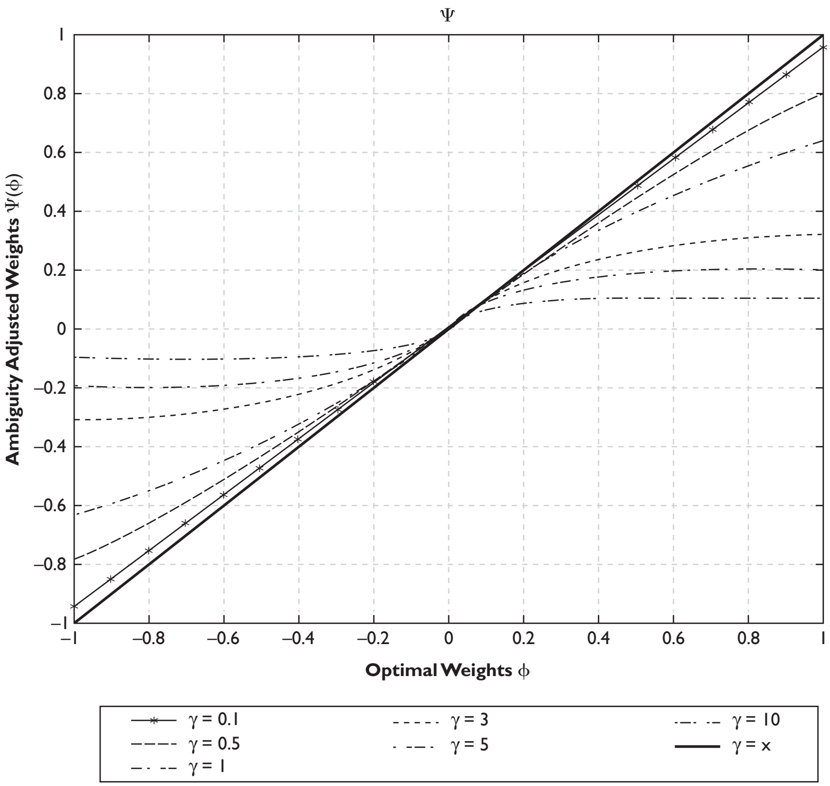
In the following are some additional properties of the function ψ:

- *Symmetry:* The function ψ is symmetric around zero. The investor has the same aversion against positive or negative weights of the same absolute value. In a context where short selling is possible, there is no reason to differentiate the long or short weights of the same magnitude in terms of ambiguity aversion.
- *Invariant point:* There is no ambiguity aversion for a zero weight. This property boils down to the fact that the investor shows no ambiguity aversion toward a zero weight: if the model Q assigns no weight on a given asset, the transformation ψ should not modify the "neutrality" of the model Q toward this asset.
- *Limit behavior:* When the investor is infinitely averse to ambiguity, it will prevent her from trading as she trusts none of her priors, and therefore all the portfolio weights should be defaulted to zero. $\Psi(\phi,\gamma) \xrightarrow[\gamma\to 0]{} \phi$.
  On the contrary, if the investor is neutral to ambiguity, the function ψ should leave the prior-dependent weights invariant, $\Psi(\phi,\gamma) \to_{\gamma\to\infty} 0$.
- *Weight shrinking effect:* The absolute ambiguity adjusted weights are smaller than the optimal weights computed under a given prior Q in absolute terms.
  As an illustration, we plot in Figure 28.1 the following example for the function ψ for different values of γ:

$$\psi(\phi,\gamma) = \frac{1-\exp^{-\gamma\phi}}{\gamma}, 0 \le \phi \le 1,\ \psi(\phi,\gamma) = \frac{1-\exp^{-\gamma\phi}}{\gamma}, -1 \le \phi \le 0.$$

In the following subsections, we discuss in more details the ambiguity aversion parametrization of the AARA function and the specific role played by the risk-free asset.

Figure 28.1 An Example of the Function ψ



The Particular Role of the Risk-Free Asset

The risk-free asset plays a specific role in the ambiguity-adjusted optimal asset allocation. Such an asset has a certain future value and there is no model risk associated with it. It therefore plays a specific role in the ARA asset allocation problem.

Indeed, the risk-free asset can be assimilated to a refuge value in the following sense: the more the investor is averse to ambiguity, the more she will invest in the risk-free asset. The ambiguity aversion leads the investor to invest less in ambiguous assets (as the function ψ reduces the optimal weights obtained for a prior Q).

Therefore the “disinvested” risky investment value, which is the difference between the value invested in risky assets under no ambiguity and the value invested in case the investor is ambiguity averse, is transferred into the risk-free asset.

For the risk-free asset, the ARA weight is defined as: $\phi_0^{ARA} = 1 - \sum_{i=1}^{N} \phi_i^{ARA}$.

In our model, the risk-free asset plays the role of a refuge value. Indeed, the sum of the transformed weights is not equal to one. Therefore the residual proportion of wealth not allocated in the risky assets is put into the risk-free asset. The transformed weight of the risk-free asset represents the amount of money the investor is reluctant to invest on risky assets because of her aversion to ambiguity.

The ARA asset allocation is bounded by the subjective expected utility asset allocation.

Now that we have presented the absolute ambiguity robust adjustment function, we focus in the next section on the relative ambiguity robust adjustment function.

RELATIVE AMBIGUITY ROBUST ADJUSTMENT

Once we have transformed the weights through the AARA function, we need to aggregate those solutions across the different models considered. We propose a transformation π to account for the mixing of the different priors ambiguity-adjusted optimal weights.

Once the weights have been computed for each prior Q and have been independently adjusted for ambiguity aversion through the AARA function, we need to combine them across all priors. We propose a RARA function through a mixture measure π, where $\pi(Q)$ represents the likelihood according to the agent anticipation that the model Q occurs, as well as the relative ambiguity aversion the investor displays toward Q among the set of priors.

Characterization of π

The RARA function π allows the investor to give relative weights to each of the priors she considers, taking into account:

- The relative ambiguity aversion among the priors (i.e., taking into account the correlation of the different priors)
- The overall ambiguity aversion of the investor toward the set of priors

The investor aversion to ambiguity is dynamic, in the sense that depending on the period considered, she will be more or less confident about her priors and the overall set of priors she considers. Therefore, we allow the function π to adapt dynamically and expend or contract the total investment size whether the total ambiguity aversion decreases or increases overtime.

As pointed out by Epstein and Schneider (2007), the ambiguity aversion of an investor is not monotonically decreasing over time, as a Bayesian updating system suggests. Our RARA function allows the investor to adjust her portfolio weights dynamically, depending on her overall belief of how much her priors can explain the true distribution.

The measure π is a RARA measure if:

- At best, a model Q perfectly fits reality and is none ambiguous: $\pi(Q) = 1$. At worse, it is completely ambiguous: $\pi(Q) = 0$. We have: $0 \leq \pi(Q) \leq 1$.
- The set of models can at best perfectly represent reality, so we have: $\sum_Q \pi(Q) \leq 1$.

Note that the RARA measure is not necessarily a probability measure. In this respect, π differs from the distribution measure μ on the set of priors commonly used so far (as in Klibanoff, Marinacci, and Munkerji, 2005). Indeed, most of the time we will have: $\sum_Q \pi(Q) < 1$, which translates the fact that the investor does not believe that she has a full understanding of the true model leading asset returns dynamic.

Estimating π

Many methods could be used in order to calibrate the value of $\pi(Q)$. We propose a simple empirical methodology which takes into account the relative historical performance of the different priors. First, we compute the time series of performance measures on the different priors considered, evaluated over a given time window. The measure π can then be computed as a weighted average of the performance measures. We illustrate the computation of π in an empirical example found in Barrieu and Tobelem (2009a).

THEORETICAL EXAMPLE

In Barrieu and Tobelem (2009b), we show that our ARA methodology gives comparable results to the Klibanoff, Marinacci, and Munkerji methodology. In this section, we provide a theoretical example for our ARA methodology, with a more complex setting than the one presented in Klibanoff, Marinacci, and Munkerji (2005). More precisely, we show that our ARA methodology allows us to solve cases, where the set of priors is continuous and the distribution for the risky and ambiguous asset are also continuous, which would not be possible in a setting such as the one proposed by Klibanoff, Marinacci, and Munkerji.

Settings

As in Klibanoff, Marinacci, and Munkerji (2005), we assume there exists only three different assets in the financial market:

- A risk-free, nonambiguous asset S^0
- A risky nonambiguous asset S^1, i.e., there is an uncertainty about the future value of S^1, however, there is no uncertainty about the distribution that leads this value
- An ambiguous nonrisky asset S^2, i.e., there is uncertainty about the model that leads the future value of this asset, although under the assumption that we know this model Q, the value of S^2 is deterministic

For the sake of simplicity, we assume all assets have an initial value of 1 and we consider only one period from time 0 to time T. The set of priors $\{Q^q\}_{0 \leq q \leq d}$ is a continuous set. We define by π the distribution on the different priors and we assume that all the priors are equipotent for the investor (all priors have the same likelihood): $\forall q \subset [0, d]$, $\pi(Q^q) = \frac{1}{d}$.

Also, we assume that under the prior Q^q, the ambiguous asset S^2 is equal to q. Therefore the ambiguous asset follows a uniform law on the prior distribution π.

We assume that the value of the different assets at horizon time T is given as follows:

- $S_T^0 = r$, the deterministic risk free return is r. The asset S_T^0 displays a constant return r whatever the prior considered: it is nonrisky and nonambiguous.
- S_T^1 follows a normal distribution with mean d (where $r < d$), and standard deviation σ. The risky asset follows the same normal distribution under any prior Q^q: it is a nonambiguous, risky asset.
- $S_T^2 = q$, the ambiguous asset follows a uniform distribution on the priors distribution. It displays a constant return q depending on the prior Q^q considered. Under a given prior Q^q, it is a risk-free asset: it is an ambiguous, nonrisky asset.

In addition, the investor utility function is defined as: $u(x,\lambda) = -\exp^{-\lambda x}$, where λ stands for the investor risk aversion. The decision maker wants to form a portfolio which maximizes the future expected wealth utility, where the future wealth X_T^ϕ is defined as: $X_T^\phi = \sum_i \phi^i S_T^i$, where the ϕ^i denote the weight of the asset S^i in the investor portfolio. Note that we have: $X_0^\phi = \sum_i \phi^i S_0^i = 1$.

In the following subsection, we compute the ARA transformed weights.

The ARA Transformation

Let us now compute the weights obtained through an ARA transformation. First we compute the optimal weights under each prior Q^q. We consider a classical setting where the investor wants to maximize the expected utility of her future portfolio value. Under a given prior Q^q, the investor wants therefore to solve the following program: $\max_\phi E[u(X_T^\phi)]$.

We can distinguish two cases:

- *Case where* $0 \leq q \leq r$: The ambiguous asset has always a return lower than the risk-free, ambiguous-free asset. Therefore, under a prior Q^q, the investor will only consider investments in the risky asset and the most profitable risk-free asset, i.e., in the present case S^0. In this case we have $\phi^2 = 0$, and to simplify we can denote $\phi^1 = \phi$ and $\phi^0 = 1 - \phi$. We deduce that the optimal solution in this case is[2]:

	$\mathbf{0 \leq q \leq r}$
$\phi^{0,q}$	$1-\dfrac{d-r}{\lambda\sigma^2}$
$\phi^{1,q}$	$\dfrac{d-r}{\lambda\sigma^2}$
$\phi^{2,q}$	0

- *Case where* $r < q \leq d$: The ambiguous asset has always a return greater than the risk-free, ambiguous-free asset. As previously argued, under a prior Q^q, the investor will only consider investments in the risky asset and the most profitable risk-free asset, i.e., in the present case S^2. In this case we have $\phi^0 = 0$, and to simplify we can denote $\phi^1 = \phi$ and $\phi^2 = 1 - \phi$.

Under a given prior Q^q, the risk-free return of the ambiguous asset is q. By similar calculus as previously done, we find that the optimal solution in this case is:

	$\mathbf{r < q \leq d}$
$\phi^{0,q}$	0
$\phi^{1,q}$	$\dfrac{S^1-q}{\lambda\sigma^2}$
$\phi^{2,q}$	$1-\dfrac{S^1-q}{\lambda\sigma^2}$

We now need to apply the AARA transformation to the optimal weights obtained for each prior Q^q. To simplify the case example, we assume that the parameter γ remains the same for all the priors considered (this goes along with the fact that the investor affects a homogeneous weight to all the priors: a priori, the investor considers all the priors equally ambiguous). We also apply the RARA transformation across all the priors considered.

The final ARA weights are therefore defined as: $\phi^{i,ARA} = \int \psi(\phi^{i,q},\gamma)\pi(q)dq, i \subset \{1,2\}$ and $\phi^{0,ARA} = 1 - \phi^{1,ARA} - \phi^{2,ARA}$.

We will assume that the investor risk aversion λ is such that $\lambda < \frac{d}{\sigma^2}$, so that all the optimal weights under all the priors Q^q are defined on the interval$[0,1]$. The calculus when $\lambda \geq \frac{d}{\sigma^2}$ would be similar.

The weights under the ARA transformation are defined respectively as:

$\phi^{0,\text{ARA}}$	$1-\phi^{1,\text{ARA}}-\phi^{2,\text{ARA}}$
$\phi^{1,\text{ARA}}$	$\frac{d-r}{d\gamma}+\left(\frac{r}{d}-\frac{\lambda\sigma^2}{d\gamma}\right)\frac{1-\exp^{-\gamma\frac{d-r}{\lambda\sigma^2}}}{\gamma}$
$\phi^{2,\text{ARA}}$	$\frac{d-r}{d\gamma}+\exp^{-\gamma}\left(\frac{\lambda\sigma^2}{d\gamma}\frac{1-\exp^{\gamma\frac{d-r}{\lambda\sigma^2}}}{\gamma}\right)$

Description of the ARA Weights

It is interesting to study the optimal weights at the limit values of the parameter γ^3.

When the ambiguity aversion goes to infinity, the investor invests all her wealth in the risk-free, nonambiguous asset:

$\lim_{\gamma\to\infty}\{\phi^{0,\text{ARA}}(\gamma)\}$	1
$\lim_{\gamma\to\infty}\{\phi^{1,\text{ARA}}(\gamma)\}$	0
$\lim_{\gamma\to\infty}\{\phi^{2,\text{ARA}}(\gamma)\}$	0

It is also interesting to consider the weights when the investor has no aversion to ambiguity ($\gamma = 0$):

$\phi^{0,\text{ARA}}(0)$	$\frac{r}{d}\left(1-\frac{d-r}{\lambda\sigma^2}\right)\frac{r}{d}\left(1-\frac{d-r}{\lambda\sigma^2}\right)$
$\phi^{1,\text{ARA}}(0)$	$\frac{r}{d}\left(\frac{d-r}{\lambda\sigma^2}\right)+\left(\frac{d-r}{r}\right)\frac{d-\frac{d+r}{2}}{\lambda\sigma^2}$
$\phi^{2,\text{ARA}}(0)$	$\frac{d-r}{d}\left(1-\frac{d-\frac{d+r}{2}}{\lambda\sigma^2}\right)$

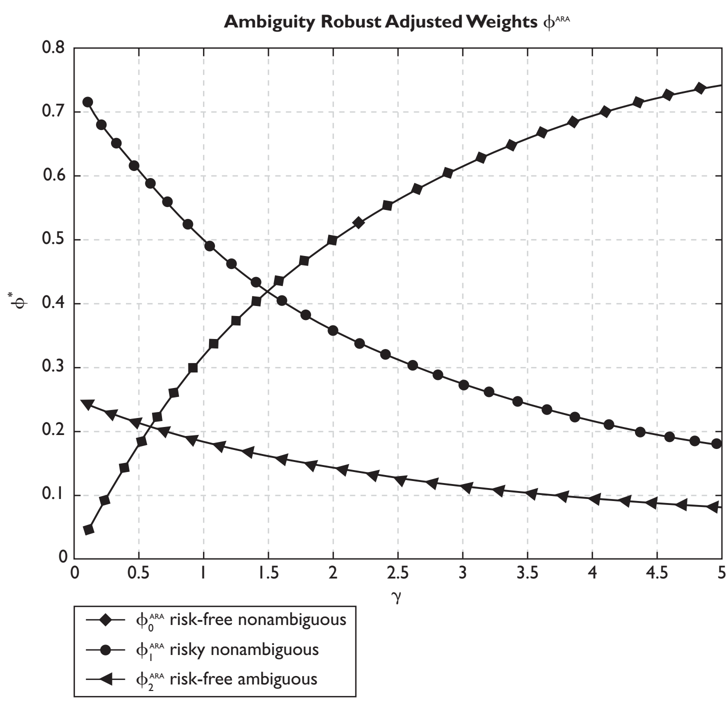
When the aversion to ambiguity is null, we are in the case of the subjective expected utility, as in Savage (1954): the weights are equal to the expected value of the different priors Q_q conditional weights, weighted by the measure π.

In Figure 28.2, we have plotted the optimal weights with a parameter γ ranging from 0.1 to 5. The aversion to ambiguity affects the risky asset as well as the ambiguous asset. The weights of the risky asset and the ambiguous asset decrease with respect to an increase in the aversion parameter γ, whereas the allocation of the risk-free, nonambiguous asset increases.

CONCLUSION

In this chapter, we proposed an easy to implement robust ambiguity methodology that allows the investor to adapt her portfolio to her level of ambiguity aversion. The parameterization for the absolute ambiguity aversion and the relative ambiguity aversion have been studied in Barrieu and Tobelem (2009a), where we test empirically a linear form for the function π and where we propose an ad hoc methodology to set the value of the

Figure 28.2 ϕ^{ARA}



parameter γ, depending on the past performance of the different priors Q_q considered. Further studies should focus on a nonlinear form of the RARA function π, that allows the investor to overweight in a nonlinear way the assets for which the priors agree (to represent the fact that the more the priors agree on a given asset weight, the less ambiguous the asset is). Indeed, the question of mixing the optimal weights obtained through heterogeneous priors remains a challenging one in many scientific fields.

REFERENCES

Barrieu, P. and Tobelem, S. (2009a) Robust Asset Allocation Under Model Risk. *Risk Magazine*, February: 91–93.

Barrieu, P. and Tobelem, S. (2009b) Robust Decision Under Ambiguity. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1318579.

Epstein and Schneider (2007) Learning Under Ambiguity, Review of Economic Studies, 74(4), 1275–1303.

Klibanoff, P., Marinacci, M., and Mukerji, S. (2005) A Smooth Model of Decision Making Under Ambiguity. *Econometrica*, 73: 1849–1892.

Savage, L.J. (1954) *The Foundations of Statistics*. New York: Wiley.

Tobelem, S. (2009) Asset Allocation Under Ambiguity. Ph.D. thesis, Statistics Department, London School of Economics.

NOTES

1. The results can be extended to a countable set, and even an uncountable parameterized family; see the theoretical example we present below.
2. For a proof, see Tobelem (2009).
3. For a detail of the calculus, see Tobelem (2009).

This page intentionally left blank

INDEX