Financial Econometrics Modeling

Derivatives Pricing, Hedge Funds and Term Structure Models

Edited by Greg N. Gregoriou and Razvan Pascalau

Financial Econometrics Modeling: Derivatives Pricing, Hedge Funds and Term Structure Models

Also by Greg N. Gregoriou and Razvan Pascalau

FINANCIAL ECONOMETRICS MODELING: Market Microstructure, Factor Models and Financial Risk Measures

NONLINEAR FINANCIAL ECONOMETRICS: Markov Switching Models, Persistence and Nonlinear Cointegration

NONLINEAR FINANCIAL ECONOMETRICS: Forecasting Models, Computational and Bayesian Models

Financial Econometrics Modeling: Derivatives Pricing, Hedge Funds and Term Structure Models

Edited by

Greg N. Gregoriou
Professor of Finance, State University of New York (Plattsburgh)
Research Associate EDHEC Business School, Nice, France

and

Razvan Pascalau
Assistant Professor of Economics, State University of New York (Plattsburgh)

palgrave
macmillan

First published in 2011 by
PALGRAVE MACMILLAN

Palgrave Macmillan in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan in the US is a division of St Martin's Press LLC, 175 Fifth Avenue, New York, NY 10010.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

Palgrave® and Macmillan® are registered trademarks in the United States, the United Kingdom, Europe and other countries.

ISBN: 978–0–230–28363–3 hardback

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

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

A catalog record for this book is available from the Library of Congress.

10 9 8 7 6 5 4 3 2 1
20 19 18 17 16 15 14 13 12 11

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

Contents

Tables

Figures

Acknowledgments

We thank Lisa von Fircks and Renee Takken at Palgrave and the team at Newgen Imaging Systems. We also thank a handful of anonymous referees for the selection of papers for this book. Neither the contributors nor the publisher is held responsible for the contents of each chapter. The contents of each chapter remain the sole responsibility of each author.

About the Editors

Greg N. Gregoriou has published 40 books, over 55 refereed publications in peer-reviewed journals, and 20 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/CRC Press, Palgrave-Macmillan, and Risk/Euromoney books. His articles have appeared in the *Journal of Portfolio Management, Journal of Futures Markets, European Journal of Operational Research, Annals of Operations Research, Computers and Operations Research*, etc. Professor Gregoriou is co-editor and editorial board member for the *Journal of Derivatives and Hedge Funds*, as well as editorial board member for the *Journal of Wealth Management*, the *Journal of Risk Management in Financial Institutions*, the *Journal of Multinational Energy and Value* and the *Brazilian Business Review*. A native of Montreal, Professor Gregoriou obtained his joint Ph.D. in Finance at the University of Quebec at Montreal which merges the resources of Montreal's four major universities (University of Quebec at Montreal, 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). He is also Research Associate at the EDHEC Business School in Nice, France.

Razvan Pascalau joined the School of Business and Economics at SUNY Plattsburgh in fall 2008. He graduated with a Ph.D. in Economics and an M.Sc. in Finance from the University of Alabama. He also holds an M.Sc. in Financial and Foreign Exchange Markets from the Doctoral School of Finance and Banking in Bucharest, Romania. In 2004, he worked full time for the Ministry of Finance in Romania as a Counselor of European Integration. His primary field of interest is (Applied) Time Series Econometrics with an emphasis on modeling nonlinear structures in macro and financial data. His research interests also include topics related to financial risk management, international finance, and managerial finance/economics. He has published in *Applied Economic Letters* and the *Journal of Wealth Management*.

Notes on Contributors

Tom Arnold is Associate Professor at the Robins School of Business at the University of Richmond. He is the current holder of the F. Carlyle Tiller Chair in Business. He is a CFA and has a Ph.D. from the University of Georgia. He has published in the *Journal of Finance*, the *Journal of Business*, the *Financial Analysts Journal*, the *Journal of Futures Markets*, the *Journal of Financial Research*, the *Engineering Economist*, the *Journal of Financial Education*, and other journals. His research interests include derivative securities, real option analysis, market microstructure, corporate finance, investments, and finance pedagogy.

Mohamed El-Hedi Arouri is currently an Associate Professor of Finance at the University of Orleans, France, and a Researcher at EDHEC Business School. He holds a Master's degree in Economics and a Ph.D. in Finance from the University of Paris X, Nanterre. His research focuses on the cost of capital, stock-market integration, and international portfolio choice. He published articles in refereed journals such as the *International Journal of Business and Finance Research*, *Frontiers of Finance and Economics*, *Annals of Economics and Statistics*, *Finance*, and *Economics Bulletin*.

Muddun Bhuruth is Professor of Computational Mathematics in the Department of Mathematics at the University of Mauritius. He holds an M.Sc. in Applied Mathematics from the University of Bordeaux I, France, and a Ph.D. in Mathematics from the University of Mauritius. His current research interests include numerical algorithms for pricing financial options and Krylov subspace methods for large-scale eigenvalue problems.

Ravindra Boojhawon is Senior Lecturer in the Department of Mathematics at the University of Mauritius. He holds a B.Sc. and a Ph.D. in Mathematics from the University of Mauritius. His current research interests include numerics for finance and Krylov subspace methods for linear systems.

Timothy Falcon Crack holds the Chair in Finance at Otago University in New Zealand. He graduated with a Ph.D. in Finance from the Massachusetts Institute of Technology and has published in the *Journal of*

Finance, the *Journal of Business*, the *Financial Analysts Journal*, the *Journal of Futures Markets*, the *Journal of Financial Education* and other journals. He is the author of two books: *Heard on the Street: Quantitative Questions from Wall Street Job Interviews* and *Basic Black-Scholes: Option Pricing and Trading*. His research interests include empirical capital markets, derivatives, econometrics, quantitative active equity trading strategies, fixed income theory, and market microstructure.

Raphaële Chappe practiced as an attorney for eight years in the financial services industry. In her last position, she worked as a Vice President with Goldman Sachs in the Tax Department, where her principal responsibilities consisted in providing tax and legal advice with respect to the firm's proprietary and balance sheet investments. She has extensive teaching experience in business, economics and finance, both at the graduate and undergraduate level, in a variety of universities, including NYU Polytechnic University, Long Island University, Parsons The New School for Design and John Jay College of Criminal Justice. Her current research interests include hedge funds and Ponzi schemes, the economics of regulation and antitrust, law and economics, risk management, and white collar crime and the role played by ethics in the business world. Raphaële holds a Master of Laws (LLM) in International Tax from New York University (School of Law) and a Master's in Comparative Business Law from the University of Pantheon-Sorbonne in Paris. She is currently pursuing her Ph.D. in Economics at the New School for Social Research in New York City.

Carolyn V. Currie is a member of the Association of Certified Practicing Accountants, the Chartered Secretaries Association, and a Fellow of Finsia, a merger of the Australian Institute of Banking and Finance and the Securities Institute. Her experience represents almost four decades in the public and private sector, as a merchant banker, regulator, internal auditor, and financial trainer. For the past fifteen years she has been a Senior Lecturer in Financial Services at the University of Technology Sydney (UTS) as well as managing director of her own consulting company and several private investment companies.

Ashvin Gopaul is an Associate Professor of Mathematics in the Department of Mathematics at the University of Mauritius. He holds a B.Sc. degree and a Ph.D. in Mathematics from the University of Mauritius. His current research interests include numerics of finance and numerical partial differential equations.

Sam Hakim is an Adjunct Professor of Finance at Pepperdine University in Malibu, California. He is concurrently a Vice President of Risk Management at Energetix LLP, an energy company in Los Angeles, California. Previously, Dr. Hakim was Director of Risk Control at Williams, an oil and gas company in Houston. Dr. Hakim was also financial economist at Federal Home Loan Bank in Washington, DC. Between 1989 and 1998, Dr. Hakim was an Associate Professor of Finance and Banking at the University of Nebraska at Omaha. Dr. Hakim is the author of more than 40 articles and publications. He is an Ayres Fellow with the American Bankers Association in Washington, DC, and a fellow with the Economic Research Forum. He holds a Ph.D. in Economics from the University of Southern California.

Andrew Hughes Hallett is Professor of Economics and Public Policy in the School of Public Policy at George Mason University. From 2001 to 2006, he was Professor of Economics at Vanderbilt University (Nashville, Tenn.) and before then at the University of Strathclyde in Scotland. He is a graduate of the University of Warwick, holds a doctorate from Oxford University, and was Visiting Professor in the Economics Department at Princeton University (1992–1994), Bundesbank Professor at the Free University of Berlin (2005), and has been a Visiting Professor at the Universities of Frankfurt, Rome, and University of Paris X. He is currently a Research Fellow in the Centre for Economic Policy Research in London, a Fellow of the Royal Society of Edinburgh and former chair of their Economics Committee, Joint Editor of the *Scottish Journal of Political Economy*, and an official correspondent of the American Mathematical Society. In 2007 he was appointed to the Council of Economic Advisers to the Government of Scotland. His research interests are in the fields of international economic policy, the political economy of monetary integration, and the theory of economic policy. In these areas he has published over 260 papers in leading academic journals and is ranked in the top 1 percent of economists worldwide by publications and citations. He has also acted as consultant to the World Bank, the International Monetary Fund, the Federal Reserve, the European Commission, and various parliamentary select committees.

Chih-Ying Hsiao completed a B.Sc. in Mathematics at National Taiwan University and a Diploma in Mathematics at Bielefeld University in Germany. Then she worked at the Center for Empirical Macroeconomics and as instructor at Bielefeld University. She holds a Ph.D. in Economics from the University of Bielefeld. Currently she works as a

Research Associate on the project "Assessing and Estimating Credit Risk" at University of Technology, Sydney. Her research includes publications in estimation of continuous-time interest-rate models, estimation using filtering methods, and optimal investment strategies.

Fredj Jawadi is currently an Associate Professor at the University of Evry Val d'Essonne and researcher at Amiens School of Management and Economics in France. He holds a Master in Econometrics and a PhD in Financial Econometrics from the University of Paris X Nanterre (France). His research topics cover modeling asset price dynamics, non-linear econometrics, international finance and financial integration in developed and emerging countries. He has published in international refereed journals such as *Journal of Risk and Insurance, Applied Financial Economics, Finance, Economics Bulletin*, four books and several book chapters.

Matteo Modena is in his final year of Ph.D. in Economics at the University of Glasgow. His postgraduate studies have been supervised by Professor Ronnie MacDonald. Modena holds a degree in Economics (University of Milan–Bicocca) and an M.Sc. in Economics and Finance (University of Edinburgh). He is interested in macroeconomics, monetary policy, and applied econometrics. Modena attended some editions of the Summer School of Econometrics (CIDE Bertinoro, Italy). He has teaching experience in the field of financial markets, monetary policy, and bond pricing. Modena is currently furthering his professional experience as a risk manager in the hedge-fund industry.

Simon Neaime is Professor and Chair of the Department of Economics, American University of Beirut, Lebanon. He has numerous academic journal articles published in top ranking economic and finance journals and has also published several graduate textbooks in financial, monetary and international economics. His most recent articles appeared in the *Review of Middle East Economics and Financ*e, the *Journal of Economic Integration*, the *North American Journal of Economics and Finance*, and the *International Economic Journal*.

Christian Richter is Principal Lecturer and Field Leader for Accounting, Finance and Economics at University of East London (UK). Previously Christian was Lecturer in Economics at the Universities of Cardiff, Loughborough and Kingston. Christian is a graduate in Economics of the Universities of Mainz and Glasgow and holds a PhD (Economics) of the University of Strathclyde. Christian has published in international economics journals such as *International Journal of Finance and Economics, Computational Economics, International Economics and Economic Policy*, and *The World Economy*. Christian recently became associate editor of *International Economics and Economic Policy*. He is also the chair of the International Network for Economic Research (INFER, www.infer-research.net <http://www.infer-research.net/>). His research interests are business cycle econometrics, financial econometrics, time-frequency analysis.

Adam Schwartz is an Associate Professor at the Williams School of Commerce, Economics, and Politics at Washington and Lee University. He has a Ph.D. from the University of Georgia and has published in the *Journal of Financial and Quantitative Analysis*, *Journal of Futures Markets*, *Journal of Financial Research*, *Journal of Financial Engineering, Journal of Derivatives*, and other journals. His research interests include derivative securities, real option analysis, investments, and finance pedagogy.

Willi Semmler is Professor at the Department of Economics at the New School, New York. He is author or co-author of more than 85 refereed articles in international journals and is author or co-author of 11 books. He is member of the New York Academy of Sciences and has been a visitor of Columbia University and Stanford University and the Cepremap in Paris. He was Fortis Bank Visiting Professor of the University of Antwerp and Visiting Professor at the University of Marseilles–Aix-en-Provence and has taught financial economics for the European Quantitative Economics Doctorate Program at universities in Italy, Spain, Portugal, and Germany. He taught at the Bielefeld University and serves on the board of directors for the Center for Empirical Macroeconomics, Bielefeld University, and has served to evaluate research projects for the National Science Foundations of Austria, Germany, Belgium, and the UK. He has also lectured at the UNAM in Mexico City, the University of Orleans, France, and Chuo University, Tokyo. He evaluates research projects for the European Union and serves as a consultant for the World Bank and the ILO. He is on the scientific committee of the Society for Nonlinear Dynamics and Econometrics, the Society for Computation in Economics and Finance, and the International Working Group on Computational and Financial Econometrics. Recently, the second edition of his book *Asset Prices, Booms and Recessions* (2006) was published. He is also co-editor of the book series at Springer on "Dynamic Modeling and Econometrics and Finance." Another recently co-edited book is *Foundations of Credit Risk Analysis* (2007).

Yannick Desire Tangman is a Ph.D. student in Mathematics at the University of Mauritius and teaches courses in financial engineering at the University of Technology, Mauritius. He holds a B.Sc. in Mathematics with Computer Science from the University of Mauritius and is currently carrying out his work under a University of Mauritius postgraduate scholarship. His research work focuses on fast algorithms for pricing American and Asian options.

Christian Thomann is a Senior Research Fellow at the Center for Risk and Insurance at the Leibniz University in Hanover, Germany. His research interests are in applied financial econometrics and insurance economics. He obtained his doctorate at the University of Hanover. His dissertation focuses on the regulation of terrorism insurance in Germany. He has edited three volumes and written two monographs on insurance economics. He has also worked as an adviser in health insurance to the German development bank and to the World Health Organization.

Abstracts

1 The Operation of Hedge Funds: Econometric Evidence, Dynamic Modeling, and Regulatory Perspectives

Willi Semmler and Raphaële Chappe

This chapter presents a stochastic dynamic model that can be used to describe situations in asset management where the use of leverage to generate above-average promised returns can result in some hedge funds finding themselves embroiled in a Ponzi financing scheme. Greater transparency is necessary to reduce such opportunities as well as to understand systemic risk implications of hedge-fund operations. This chapter also assesses the new regulatory frameworks proposed by the Obama Administration and by the European Commission to that effect.

2 Inferring Risk-Averse Probability Distributions from Option Prices Using Implied Binomial Trees

Tom Arnold, Timothy Falcon Crack, and Adam Schwartz

We generalize the Rubinstein (1994) risk-neutral implied binomial tree (R-IBT) model by introducing a risk premium. Our new risk-averse implied binomial tree (RA-IBT) model has both probabilistic and pricing applications. We use the RA-IBT model to estimate the pricing kernel (i.e., marginal rate of substitution) and implied relative risk aversion for a representative agent; we are the first to use wholly implied methods for this. We also discuss applications of the RA-IBT to stochastic volatility option pricing models.

3 Pricing Toxic Assets

Carolyn V. Currie

The current global financial crisis has been triggered by a huge increase in credit default swaps and securitization of bundled mortgages into collateralized debt obligations (CDOs) which have layers of equity and debt with the latter then bundled into entities known as CDO squared. The pricing of both has been faulty, resulting in misplaced credit ratings of AAA or above, resulting in severely degraded assets known as "toxic." This article discusses a new pricing policy for toxic assets which does not rely on traditional finance theory but uses the pricing associated with simple floor plans developed in Australia during the twentieth century for gold and other commodity and agricultural products. Binomial theory can be used to price the associated collar that relies on put and call options. This chapter begins by defining toxic assets and then describing pricing methods to hedge risk in commodities, concluding with a suggested pricing solution based on Basel 2 capital allocation methods.

4 A General Efficient Framework for Pricing Options Using Exponential Time Integration Schemes

Muddun Bhuruth, Ravindra Boojhawon, Ashvin Gopaul, and Yannick Desire Tangman

We develop a general and efficient framework for pricing options under various models. These include the Black–Scholes, Merton's jump-diffusion, Heston's stochastic volatility (SV), Bates' stochastic volatility with jumps models (SVJ), and infinite activity models such as CGMY as well. The semi-discrete systems arising from numerical discretizations of the pricing equations are solved using exponential time integration (ETI) schemes. For such integrators, it is well known that computation of the matrix exponential can be expensive. We show how the algorithms based on ETI can be speeded using best rational approximations via Carathéodory–Fejér points and that only four sparse linear solves are sufficient for obtaining convergent solutions in the case of European and barrier options. For American options, we combine an exponential forward Euler scheme with Richardson extrapolation in an operator splitting spatial discretization framework.

5 Unconditional Mean, Volatility, and the FOURIER–GARCH Representation

Razvan Pascalau, Christian Thomann, and Greg N. Gregoriou

This chapter proposes a new model called Fourier–GARCH which is a modification of the popular GARCH(1,1). This modification allows for time-varying first and second moments via means of flexible Fourier transforms. A nice feature of this model is its ability to capture both short- and long-run dynamics in the volatility of the data, requiring only that the proper frequencies of the Fourier transform be specified. Several simulations show the ability of the Fourier series to approximate breaks of an unknown form, irrespective of the time or location of breaks. This chapter shows that the main cause of the long-run memory effect seen in stock returns is the result of a time-varying first moment. In addition, the study suggests that allowing only the second moment to vary over time is not sufficient to capture the high persistence observed in lagged returns.

6 Essays in Nonlinear Financial Integration Modeling: The Philippine Stock Market Case

Mohamed El-Hedi Arouri and Fredj Jawadi

This chapter aims to study the hypothesis of stock-market integration with the world equity market for the Philippines in a nonlinear framework using recent developments of nonlinear cointegration, nonlinear cointegration tests, and nonlinear error correction models. Our results suggest further evidence of nonlinear integration and show an asymmetrical and time-varying stock-price mean-reverting phenomenon toward the world index.

7 A Macroeconomic Analysis of the Latent Factors of the Yield Curve: Curvature and Real Activity

Matteo Modena

This chapter extends the strand of literature that examines the relation between the term structure (TS) of interest rates and macroeconomic variables. The yield curve is summarized by three latent factors which are obtained through Kalman filtering the Nelson–Siegel TS model. In this chapter, we address the challenging issue of attributing a macroeconomic interpretation to the curvature factor, finding evidence that curvature reflects the cyclical fluctuations of the economy. Interestingly, this result holds in spite of whether curvature is extracted from the nominal or the real TS. A negative shock to curvature seems either to anticipate or to accompany a slowdown in economic activity. The curvature effect thus complements the transition from an upward-sloping yield curve to a flat one.

8 On the Efficiency of Capital Markets: An Analysis of the Short End of the UK Term Structure

Andrew Hughes Hallett and Christian Richter

In this chapter, we test the expectations hypothesis (EH) for the term structure of interest rates jointly with the uncovered interest parity (UIP) condition for the short end of the UK term structure. We find that the US interest rate, the UK monetary instrument, and the (spot) exchange rate all affect the short-term interest rate. However, the impact of the US rate is the biggest effect, although that has decreased a little during the recent financial crisis. We also test a bounded rationality approach. We deviate from the contemporary literature by refusing to impose rational expectations. Instead, we assume extrapolative expectations as an obvious behavioral alternative. We show that incorporating extrapolative expectations in both hypotheses turns out to be a significant improvement. Hence, and in contrast to previous work which has assumed rational expectations, we find the UIP and the EHs are not rejected. Thus, the problem seems to have been violations of the rational expectations paradigm, not violations of UIP or the EH of the term structure per se.

9 Continuous and Discrete Time Modeling of Short-Term Interest Rates

Chih-Ying Hsiao and Willi Semmler

The objective of this chapter is to compare continuous and discrete time models of short-term interest-rate data. Regarding the continuous-time method, three discretization methods, the Euler method, the Milstein method, and the new local linearization method are employed to obtain discrete-time approximate models. We estimate the short-term interest rates of Germany, the UK, and the USA. Results of the specification tests of autocorrelation and normality indicate that the model suggested by Chan et al. (1992) is not a very strong candidate to be the data-generating process for these short rates as well as the other continuous-time models suggested by Ait-Sahalia (1996) and Anderen and Lund (1997). Therefore, we turn to adopt discrete-time models.

We find that the ARMA-ARCH model with level dependent volatility performs better than the continuous-time models in terms of the likelihood values and the specification tests as well as the in-sample and out-of-sample forecasts.

10 Testing the Expectations Hypothesis in the Emerging Markets of the Middle East: An Application to Egyptian and Lebanese Treasury Securities

Sam Hakim and Simon Neaime

Despite many rejections, the expectations hypothesis remains the widely accepted premise believed to explain the shape of the yield curve. This chapter investigates the stochastic properties of the term structure of interest rates in Egypt and Lebanon, two emerging bond markets in the Middle East. Our results show that interest rates in each of these two countries are non-stationary and can be modeled as unit root processes. Further, cointegration analysis indicates that the interest rates do not drift apart but move together over time, a finding that supports the expectations hypothesis. Our results shed light for the first time on two of the Middle East's bond markets, a region where interest rates have received little attention before. Furthermore, the results suggest that the expectations hypothesis is strong enough to hold even in infant emerging markets.

Part I

Derivatives Pricing and Hedge Funds

1
The Operation of Hedge Funds: Econometric Evidence, Dynamic Modeling, and Regulatory Perspectives

Willi Semmler and Raphaële Chappe

1.1 Introduction

The Madoff case has all the makings of a Ponzi scheme. Ponzi schemes follow what Hyman Minsky described as Ponzi finance. Do hedge funds, or at least some of them, follow a similar scheme? The best summary of different financing practices, such as hedge, speculative, and Ponzi financing, is given in Minsky (1986). Hedge finance is a situation where operating cash flow can service all payment obligations associated with the financing. Speculative finance involves situations where operating cash flow supports interest payments but not repayment of principal. Ponzi finance describes a situation where operating cash flow is insufficient to cover either principal or interest payments, which can be financed only via an increase in liabilities, thus by a new inflow of funds.

This chapter shows a model that can be used to describe situations in asset management where the use of leverage to generate above-average returns may result in some hedge funds finding themselves, willingly or unwillingly, embroiled in a Ponzi financing scheme. This chapter is organized as follows. After reviewing the literature on hedge funds, this chapter briefly examines empirical data on the structure of the industry: size, assets under management, measures of returns, and leverage. This chapter then proceeds to develop a dynamic model of the wealth-generating process in a hedge fund. If the manager attracts investors by promising high returns, the hedge fund is, in essence, borrowing at a high interest rate. If actual returns do not exceed this high cost of borrowing, on average the hedge fund persistently needs positive inflow of fresh funds in order not to become insolvent. While hedge finance is typically associated with a reasonable debt-to-equity ratio, and speculative finance typically results in liabilities being rolled over, in Ponzi financing the value of net equity is gradually reduced as the company is in essence continually borrowing against the future. In Section 1.4, we run a simulation of the impact of the sudden decline of inflow of funds on the hedge funds, ultimately leading to insolvency.

An economy where Ponzi finance dominates is inherently fragile, unstable, and conducive to financial crisis. The high level of risky credit is unsustainable in the long run, and it is only a question of time before borrowers start to default, potentially bringing the entire financial system to the brink of collapse via contagion effects if there is sufficient interconnectedness. The hedge-fund industry has been opaque, with little regulatory oversight and no reporting requirements, creating opportunities for such Ponzi financing. Given the size of the industry and the gradually increased exposure of ordinary investors and the general public to hedge funds, if indeed some funds follow a pattern of Ponzi financing, there are implications for systemic risk. However, greater transparency is necessary on the part of hedge funds to even begin to understand such implications. A new regulatory framework is needed to that effect and has been proposed by the Obama Administration.

1.2 The literature on hedge funds

The hedge-fund industry is largely unregulated. Hedge funds are under no obligation to report net asset values or income statements, unlike their counterparts in the mutual fund industry, which have to provide daily net asset value calculations and quarterly filings. Hedge-fund managers report performance information to different databases on a purely voluntary basis. This has given rise to several biases and the issue of distorted data, thoroughly examined in the literature. Survivorship bias originates from lower returns being excluded from performance studies if the hedge fund has failed (only surviving funds are analyzed). There are different studies and estimates of survivorships bias in the literature, analyzing different years and different funds. Amin and Kat (2003) estimate survivorship bias at about 2 percent per year. Fung and Hsieh (2000) have calculated that the survivorship bias is 3 percent annually with a 15 percent dropout rate. Malkiel and Saha (2004) found a bias averaging

3.74 percent per year. Self-selecting bias is generated to the extent that hedge-fund managers report performance after the fact and only have an incentive to do so if the hedge fund has performed well. Hedge funds might wait for a good track record before they start to report to a database. Underperforming funds might decide to stop reporting to a database. It is typically the case that hedge funds stop reporting during their last months (as was the case, for example, with Long Term Capital Management losses between October 1997 and October 1998, which were not reported to data providers).

One example of self-selecting bias is the backfill bias (also called instant history bias), which appears when hedge-fund managers add historical results to their files to give a more comprehensive view of the fund's performance once the fund begins reporting to a database. Since fund managers have an incentive to begin reporting only with a good track record, backfilled returns tend to be higher than contemporaneously reported returns. The backfill bias has been addressed for different databases at different times. Fung and Hsieh (2000) estimated the backfill bias to be 1.4 percent for the Tass database over the period 1994–1998. Also on the basis of the Tass database, Posthuma and Van Der Sluis (2003) have found that backfilled returns are on average 4 percent higher annually than normal returns.

Much of the literature is focused on risk analytics for hedge funds. One issue is that the risk/reward profile for most hedge funds is not as well understood as that for traditional investments. It is well established that there is typically a trade-off between risk and return. The high (double-digit) returns historically earned by hedge funds may well present underlying risk exposures that are not well identified and managed by traditional risk-management tools. Lo (2001) has stressed the need for a new set of risk analytics designed to address the unique features of hedge funds and to go beyond traditional value-at-risk analysis. Such unique features include survivorship bias, the non-normal distribution of returns, and the enormous flexibility in trading enabled by the absence of regulatory constraints and by the hedge-fund manager's ability to manage positions very actively. The need to replace traditional risk measures has led to alternative performance measures, such as the downside risk framework designed to take into account the asymmetry of returns. Generally speaking, the downside risk framework attempts to take into consideration returns which are inferior to a target rate of return. Mamoghli and Daboussi (2009) attempt to study the impact of downside risk measures on the performance measurement of hedge funds. More detail can be found in Section 1.3.

Chan et al. (2006) have examined the risk of illiquidity exposure and have attempted to develop quantitative measures of systemic risk. Illiquid exposure is particularly relevant in the event of a forced liquidation, which is a real threat for hedge funds due to the highly leveraged nature of their positions. Quick withdrawals of credit, such as that resulting from adverse changes in the market value of posted collaterals, can directly lead to forced liquidations. In turn, the unwinding of large positions over short periods can create significant market disruptions and a breakdown of the financial system as a whole (systemic risk). Getmansky et al. (2004) show that high levels of serial correlations can be partly explained by illiquidity of the hedge fund's portfolio. Chan et al. (2006) argue that autocorrelation coefficients of the fund's monthly returns are useful indicators of liquidity exposure and show that hedge funds display a significantly higher level of serial correlation than mutual funds.[1]

There is also some analysis in the literature more specifically focused on the factors that explain the failure of a hedge fund. Malkiel and Saha (2004) have developed probit regression analysis to examine the probability of a fund's survival. The results show that the fund's performance in the most recent four quarters is an important determinant of the fund's probability of survival, that large funds have a higher probability of survival, and that a high standard deviation of returns in the fund's most recent year has a negative impact on a fund's probability of survival. Further, Malkiel and Saha (2004) find that a fund is more likely to fail in the first few years of operation and that once a fund has survived the first years and has established a track record, its likelihood of failure declines over time.

1.3 The empirics of hedge funds

The size of the hedge-fund industry has exploded in the past ten years. Its growth, for the most part attributable to large returns generated through high leverage and use of derivatives, has had a significant impact on the structure of U.S. and global markets. The industry has grown into a parallel financial universe – a shadow banking system. Hedge funds are involved in virtually every kind of market and invest in every kind of asset: from equity, loans, mortgages, and distressed debt to project finance, derivatives, etc. But the industry is still to a large extent unregulated. As a result, it is an even greater factor of systemic risk than it was in 1998 when the Fed intervened to prevent the near collapse of Long Term Capital Management.

The term "systemic risk" is used to describe the possibility that the failure of one financial institution can disrupt the financial system as a whole through a series of correlated defaults. For example, the banking panics in the USA during 1930–1933 caused many failures of banks which were otherwise financially sound.[2] Originally created as a temporary regulatory agency under the Banking Act of 1933, the Federal Deposit Insurance Corporation (FDIC) was designed to insure deposits in all national banks. The coverage is limited to a given amount per depositor, recently increased to $250,000 until December 31, 2013. By protecting depositors from bank failures, the FDIC significantly reduces the risk of bank runs and thus minimizes systemic risk. Yet the size and evolution of the hedge-fund industry raises the question of whether it is generating a new form of systemic risk: that of market disruptions caused by hedge-fund failures.

There are different estimates of the current size of the hedge-fund industry, partly due to the fact that there is no single definition of what constitutes a hedge fund, resulting in different views of what comprises the hedge-fund universe. Hedge Fund Intelligence has estimated total assets under management to be at $1.808 trillion[3] at the end of 2008, while Hedge Fund Research (HFR) estimates a lower figure of $1.4 trillion.[4] As would be expected in light of recent market disruptions, the year 2008 was a difficult one for the hedge-fund industry. Hedge funds experienced losses and redemptions and saw $155 billion of net outflows.[5] The outlook for 2010 is more promising. Deutsche Bank estimates the industry to attract $222 billion of new funds, and thus nearly return to pre-crisis levels.[6]

In spite of the recent downturn, these figures show that the industry has grown approximately by a factor of five from $387 billion in 1998. By comparison, over the same period, U.S. GDP has grown by only 64 percent and the mutual fund industry has grown by only 69 percent.[7] Citadel Investment Group, which manages approximately $15 billion of investment capital, accounts for nearly 10 percent of the daily trading volume of U.S. equities and is the largest market maker of options in the U.S., executing approximately 30 percent of all equity options trades daily (Griffin 2008). Further, ordinary investors and the general public have been increasingly exposed to the hedge-fund industry. This is due to new entities investing in hedge funds, such as pension funds, universities, endowments, charitable organizations, foundations, and "funds of hedge funds" that offer shares to the general public. In addition, hedge funds have been somewhat loosening their investment requirements.

As for other active investment portfolios, hedge-fund performance is typically measured by alphas and betas, as well as by Sharpe ratios. Hedge-fund returns are often characterized through the use of a model giving the following expected return-beta expression: $E(r_i) = \alpha_i + \beta_i E(r_m)$, where $E(r_i)$ is the expected return of the fund and $E(r_m)$ is the expected return of the market as a whole. Alternatively, the model may use only the return in excess of the risk-free rate. The beta coefficient measures the tendency of the return to rise or fall in correlation with the market. The $\beta_i E(r_m)$ term should correspond to the return that can be obtained with any diversified investment, for example an index fund. The alpha term captures the return in excess of this amount. Roughly, the beta and alpha terms measure the passive and active components of the return respectively. Normally the beta should be the right benchmark for a hedge fund, and one would expect to be paying fees only for the alpha. However, this is not the case, as the compensation structure of hedge funds is unique, typically a 2 percent management fee based on assets under management and a 20 percent performance fee on total return.[8] Presumably, it is the ability to deliver high alphas that is responsible for the rise of the hedge-fund industry. While some papers have confirmed statistically significant positive alphas, others have shown that a substantial part of the return can be explained by simple stock, bond, and cash betas (Ibbotson and Cheng 2005). Gilli et al. (2010) even show it may be possible to replicate the attractive features of hedge-fund returns using liquid assets. They are able to construct a portfolio that closely follows the CSFB/Tremont Hedge Fund Index but is less sensitive to adverse equity-market movements. This would allow investors to avoid high management fees, limited liquidity, potential redemption fees, and ultimately the lack of transparency associated with hedge funds.

The beta term can also be defined as the coefficient in a multiple regression on the return of the market portfolio (typically Standard & Poor's 500 or other index), with the alpha term the intercept (Cochrane 1999). Some alpha and beta estimates from the literature are presented in Table 1.1. In addition, Ibbotson and Cheng (2005) have also estimated a breakdown of return between the management and performance fee, and the alpha and beta returns. Working on an equally weighted index of 3,000 hedge funds over the period January 1995 to March 2004,[9] they estimated the pre-fee return from the fund to be 12.8 percent, which consisted of 3.8 percent going to fees, 3.7 percent from the alpha, and 5.4 percent from the beta. Table 1.2 summarizes the findings of Cochrane (2005), based on regressions using CFSB/Tremont indices.

Table 1.1 Alpha and beta estimates

Source	Alpha	Beta	Database used	Years examined	Number of hedge funds examined
Agarwal and Naik (2000)	0.54 to 1.25 depending on hedge fund strategy	N/A	Hedge Fund Research	January 1994 to December 1998	1,000
Malkiel and Saha (2004)		0.231 (contempo-raneous betas) 0.393 (lagged betas)	TASS	1996–2003	2,024
Goetzmann and Ross (2000)	0.3	N/A	TASS	January 1994 to May 2000	1,221
Brunnermeier and Nagel (2004)	N/A	0.42	merged database of MAR, TASS, and HFR	April 1998 to December 2000	N/A

Table 1.2 Alpha and beta estimates across investment styles

Style	Expected Return (%/mo)	a	b	a3	b3
Index	0.64	0.46	0.28	0.36	0.44
Std. errors	0.20	0.17	0.04		
Short	–0.53	0.10	–0.94	0.13	–0.99
Emerging markets	0.39	0.00	0.58	–0.07	0.69
Event	0.61	0.46	0.22	0.38	0.37
Global Macro	0.93	0.82	0.17	0.74	0.31
Long/Short equity	0.73	0.42	0.47	0.32	0.65

Source: Cochrane (2005) regressions using CFSB/Tremont indices at hedgeindex.com.

The Sharpe ratio is widely used in financial analysis (Sharpe 1994). The Sharpe ratio *SR* measures the ratio of the return of an investment earned in excess of the risk-free rate (typically the appropriate T-bill rate) over the return volatility (measured by the standard deviation), for a given time period. Sharpe ratios are often characterized with the following expression: $SR = \frac{E(R_t)-R_f}{\sigma}$, where $E(R_t)$ is the expected return of the fund portfolio between dates $t-1$ and t, R_f is the risk-free return (or other benchmark) in that period, and σ the standard deviation of the portfolio return (see Lo 2002). Roughly speaking, the ratio can be described as the reward per unit of risk. Xiong (2009) finds that funds with more assets tend to produce higher returns at lower levels of volatility, resulting in superior risk-adjusted performance. Lo (2002) has shown that annualized hedge-fund Sharpe ratios figures, computed by multiplying monthly Sharpe ratios by $\sqrt{12}$, are often overstated or understated by as much as 65 percent due to positively or negatively serially correlated returns. These estimation errors are attributable to approaches that do not take into consideration the statistical properties of the underlying returns. Lo (2002) shows that a more accurate measure of annualized Sharpe ratios requires the use of appropriate statistical distributions for the return history. There are many estimates of Sharpe ratios in the literature. Tables 1.3 and 1.4 summarize a few findings.

One criticism is that the traditional measures of alphas, betas, and Sharpe ratios are static in nature (i.e. based on return distributions at a given point in time). Lo (2001) has identified the need for a new set of risk analytics better suited to address the unique features of hedge-fund

Table 1.3 Sharpe ratio estimates

Source	**Sharpe ratio**	**Risk-free rate per annum**	**Database used**	**Years examined**	**Number of hedge funds examined**
Agarwal and Naik (2000)	0.10 (quarterly for directional strategies) 0.46 (quarterly for non-directional strategies)	5%	Hedge Fund Research	January 1994 to December 1998	1,000
Goetzmann and Ross (2000)	1.16 (annual)	US 30 day T-bill	TASS	January 1994 to May 2000	1,221

Table 1.4 Sharpe ratio estimates across investment styles

	Annualized Sharpe ratio		**Annual adjusted Sharpe ratio**				
	Mean	**SD**	**Mean**	**SD**	**Database used**	**Years examined**	**Number of hedge funds examined**
Convertible arbitrage	2.38	3.66	1.85	2.55	TASS	February 1977–August 2004	176
Dedicated short-seller	0.05	0.59	0.19	0.46	TASS	February 1977–August 2004	29
Emerging markets	0.86	1.63	0.84	1.31	TASS	February 1977–August 2004	263
Equity market neutral	0.97	1.24	1.06	1.53	TASS	February 1977–August 2004	260
Event driven	1.71	1.48	1.49	1.48	TASS	February 1977–August 2004	384
Fixed-income arbitrage	2.59	9.16	2.29	5.86	TASS	February 1977–August 2004	175
Global macro	0.60	0.92	0.70	0.90	TASS	February 1977–August 2004	232
Long/short equity	0.82	1.06	0.81	1.07	TASS	February 1977–August 2004	1,415
Managed futures	0.34	0.91	0.50	0.88	TASS	February 1977–August 2004	511
Multistrategy	1.67	2.16	1.49	2.05	TASS	February 1977–August 2004	139
Fund of funds	1.27	1.37	1.21	1.22	TASS	February 1977–August 2004	952

Source: Chan et al. (2006).

investments as the “next challenge in the evolution of the hedge-fund industry.” Lo (2008a) has proposed new measures of performance that capture both static and dynamic aspects of decision making on the part of the hedge-fund manager, in an attempt to consider forecasting skills that are also essential to successful active investment strategies. This new methodology decomposes a portfolio’s expected return into two distinct components. The first is a static weighted average of individual securities’ expected returns, which measures the portion of expected return due to static investments in the underlying securities. The second component is the sum of covariances between returns and portfolio weights, which captures the forecast power implicit in the manager's dynamic investment choices. The key assumption is that at any given date portfolio weights are functions of the manager's prior information. Thus the correlation between portfolio weights and returns is used as a measure of the predictive power of the information used by the manager to select his portfolio weights. This methodology is particularly relevant for hedge-fund strategies where both components are significant contributors to their expected returns.

Another criticism of the alpha, beta, and Sharpe ratios measures is that hedge funds may display nonlinearities that are not captured by linear regression models. The distribution of hedge-fund returns displays non-normal characteristics, which have been analyzed in the literature. For example, Brooks and Kat (2002) have found that published hedge-fund indexes exhibit high kurtosis, indicating that the distribution has "fat" tails. Examining data from the Tass database from 1995 to 2003 for various hedge-fund categories, Malkiel and Saha (2004) have confirmed that hedge-fund returns are characterized by high kurtosis and that many hedge-fund categories have considerable negative skew, implying an asymmetric distribution. The shape of the probability distribution of returns affects the Sharpe ratio. For example, Bernardo and Ledoit (2000) show that Sharpe ratios are misleading when the shape of the return distribution is far from normal.

One possible approach is to accept these limitations, recognizing that the measures are not robust to manipulations, and instead to focus on identifying strategies that "game" these performance measures, for example techniques for maximizing the Sharpe ratio. Other approaches attempt to come up with alternatives to the Sharpe ratio, for example stochastic-discount factor-based performance measures (Chen and Knez 1996), or downside variance, a new concept in risk analysis that can be used even when return series are not normally distributed. Downside risk analysis focuses on downward exposure rather than total volatility. The basic premise is that the investor has a threshold of desired performance called the minimum acceptable rate of return (MAR), which can be an index or a risk-free rate. Downside volatility is the volatility of returns below this benchmark. The Sortino ratio can be formed as the return in excess of the benchmark over the downside volatility and used as an alternative to the Sharpe ratio. The Sortino ratio (Sortino and Price 1994) is a measure of risk-adjusted return based on the MAR of each investor. Mamoghli and Daboussi (2009) show that hedge-fund strategies change ranks when the performance measure changes from the Sharpe ratio to the Sortino ratio or to other alternative performance measures. The use of the Sortino ratio allows taking into account the asymmetric nature of the returns, which is not captured by the measure of variance in the Sharpe ratio. Investors are also using maximum drawdown as a measure of expected loss. The maximum drawdown reflects the biggest past loss as a percentage of the prior high watermark. It makes no assumptions regarding distribution, and hence is robust to skewed distributions. For an example of the use of maximum drawdown to analyze performance of a fund of hedge funds, see Heidorn et al. (2009).

The main source of leverage for hedge funds is collaterized borrowing through repo markets and prime brokers. Because hedge funds have been virtually unregulated, they are not required to report their use of leverage and face no capital adequacy requirements. Leverage can be defined in balance-sheet terms, as the ratio of borrowing to net worth (equity). However, this definition fails to capture another implicit source of leverage, the additional embedded leverage of derivative products in which hedge funds have traditionally invested. This definition is also static and fails to capture the links between the ease with which a hedge fund can borrow and the fund assets' market liquidity. Brunnermeier and Pedersen (2009) provide a model that links an asset's market liquidity with traders' funding liquidity, showing the interconnectedness between the two. In this model, market liquidity is defined as the difference between the transaction price and the asset's fundamental value, while funding liquidity is defined as the ability to raise cash at short notice. Brunnermeier and Pedersen (2009) show that a mechanism of mutual reinforcement between market liquidity and funding liquidity may result in liquidity spirals. For example, the tightening of funding lowers market liquidity, leading to higher margin requirements, which in turn further accelerates the tightening of funding standards (margins are the difference between the security's price and collateral value and must be financed by the investor's own capital). More generally, while solvency itself is typically determined as a "stock" value, funding liquidity can be expressed as a flow constraint, whereby at each point in time money outflows are less than inflows and money held by the fund (see for example Drehmann 2007 and Drehmann and Nikolaou 2009). This flow constraint captures the liquidity spiral described by Brunnermeier and Pedersen (2009). Further, leverage levels also rise and fall on the basis of the fluctuations of the measured riskiness of existing assets, which may be tied to macroeconomic cycles (see Adrian and Shin 2007).

In this sense, leverage and its evolution can be thought of in terms of economic risk relative to capital.

For these reasons, it has been difficult to estimate accurate leverage levels in the hedge-fund industry. In the absence of clear reporting of hedge-fund leverage, the size of market positions managed by hedge funds compared to funds received from investor subscriptions is a good preliminary indicator of leverage. As a whole, hedge funds controlled approximately $3.7 trillion of market positions as of the end of 2008, indicating on average leverage at a ratio of 1 to 1.[10] However, there are potentially vast differences between funds: for example, Long Term Capital Management was very highly leveraged at a ratio of 28 to 1 (debt to equity) at the end of 1995,[11] as part of a strategy to multiply profit on very tiny spreads,[12] while Paulson & Co., an investment advisory firm that currently manages approximately $36 billion in assets, is only leveraged up to 33 percent of its equity capital.[13] Prior to the current financial crisis, around 70 percent of hedge funds was levered at less than 2 to 1.[14] The industry also experienced some deleveraging in 2009. In a survey conducted by Deutsche Bank at the end of 2008, only 12 percent of participants said that they used leverage, and only a further 4 percent indicated that they were interested in doing so, as opposed to 24 percent and 12 percent respectively 12 months before.[15] Thus, on average, the industry is not as levered as some of the banking institutions.[16] This said, leverage is undisclosed to investors and to the government, and, as we have mentioned, additional leverage embedded in certain products is not really being measured.

1.4 A dynamic model

In order to show how an unregulated hedge fund follow a speculative or Ponzi financing as defined in Minsky (1986), we want to introduce a stylized model. To study this problem in a simplified way (a more sophisticated dynamic decision model is sketched in the Appendix), we propose a dynamic model of wealth accumulation where the stochastic returns take on the form of a Brownian motion with mean reverting. This type of model has been used by, among others, Campbell and Viceira (2002), Wachter (2002), Munk et al. (2004), Platen and Semmler (2009), and Brunnermeier and Sannikov (2009). Details of such a model can also be found in Chiarella et al. (2007). As it is modified here, it can be seen to reflect well the risk-taking operations of hedge funds.

1.4.1 Returns to size

Standard economic theory on wealth accumulation suggests that the difference in wealth arises from income differences (shocks) and differences in saving rates, discount rates, and risk appetite and tax treatment. Cagetti and De Nardi (2006) give a comprehensive overview of the standard intertemporal economic models and the standard causes for the increase in wealth.

Here we want to use a dynamic portfolio approach of the Campbell and Viceira type, which includes not only a wealth equation but also Brownian motions for returns. Yet, we want to argue that there are scale effects from wealth, which might have to be considered when sketching such a model. In general we could imagine three types of scale effects:

First, some investors have more or better information about expected returns than others. Though it is commonly assumed that markets are efficient and that no money is to be made by forecasting (all information is already built into asset prices and returns), industry, firm, and product knowledge, as well as knowledge on innovations, product development, and future market share, is likely to give rise to better information and higher expected returns.[17] This has often been the case in industry developments such as the oil boom before World War I (WWI), the boom in the auto industry after WWI, the boom in the steel industry during and after WWI, the high-tech boom in the 1990s, the commodity price boom after 1995, the real-estate boom after 2001, and the recent banking and finance industry boom. Note, however, that there is also often manipulation of information, in particular when there are stock options as remunerations for executives.

Second, for large investors there are scale advantages, not only with respect to information but also with respect to leveraging. Investment opportunities can be explored more extensively with greater access to, and lower cost of, credit. Levering, and over-leveraging, up to a ratio of roughly 30 to 1 (see Section 1.3) has become one of the most common and well-known strategies to harvest large gains from traditional as well as new financial instruments in the recent past.[18] Yet, these gains are often only temporarily harvested, as discussed in Section 1.3.

Third, one might reasonably assume that larger income will imply lower consumption rates and higher saving rates. This will result in a higher proportion of funds being reinvested with the fund. Numerous studies in the literature have used this assumption, showing that the wealth will be built up faster with higher saving ratios. The faster build-up of wealth is an expected outcome of this strategy. There is lots of evidence that higher income leads to higher saving rates.[19]

An important mechanism of faster accumulation of wealth works with borrowing and lending schemes. As mentioned in the introduction, Minsky (1986) has introduced the financing practices such as hedge finance, speculative finance, and Ponzi finance. In recent times, in particular with the Madoff case, a Ponzi scheme has become a well-known term to the public. In testifying in front of Congress, Frankel (2008) mentions the following elements of the Ponzi scheme:

> Ponzi schemes are simple. A con artist offers obligations that promise very high returns at seemingly very low risk from a business that does not in fact exist or a secret idea that does not work out. The con artist helps himself to the investors' money, and pays the promised high return to earlier investors from the money handed over by these and later investors. The scheme ends when there is no more money from new investors. (Frankel 2008)

Ponzi schemes, by their very nature follow Ponzi financing. However, Ponzi financing does not necessarily imply the existence of a Ponzi scheme. Ponzi schemes are deceptive in nature, relying on fraudulent misrepresentations, while Ponzi finance simply entails an inability, in the long run, to sustain payments with operating cash flow. We are interested to explore the possibility of funds adopting Ponzi financing. Let us explain how this works in terms of a dynamic model of wealth accumulation. We can use a model with mean reversion in returns where the expected returns move in the long run to some mean. This is one of the most common assumptions in recent asset return modeling (see Munk et al. 2004). Consumption is neglected here. For a more complex extended version of the model, which incorporates stochastic consumption and portfolio choice and can only be solved numerically with a dynamic programming algorithm, see the Appendix.

1.4.2 The wealth-generating process

Let us first define the wealth-generating process. It starts with the investor turning over funds to a hedge manager and paying a fee,[20] $c_t W_t$, with c_t a percentage of the invested wealth, W_t. To explain the income and wealth process we use, as stochastic processes, Brownian motions – a process for wealth, a process of a mean-reverting asset returns and an interest-rate process. Models with mean reversion in returns are now frequently used in the portfolio modeling literature. The simplified dynamic model can be written as follows (see Munk et al. 2004 and Wachter 2002):

$$dW = \{[\alpha_t(r_t + x_t) + (1 - \alpha_t)r_t]W_t - c_t W_t - r^p W_t\}dt + \sigma_\omega W_t dz_t \quad \text{(A.1)}$$

$$dx_t = \lambda(\overline{x} - x_t)dt + \sigma_x dz_t \quad \text{(A.2)}$$

$$dr_t = k(\theta - r_t)dt + \sigma_r dz_t \quad \text{(A.3)}$$

Hereby, r_t, denotes the short term interest rate, αt, the fraction of wealth invested in risky assets, x_t is the premium on risky assets, $\overline{x}$ is an expected mean of the premium, which is assumed to be constant. We here have assumed a time-varying equity premium, following a mean reversion process, as in Campbell and Viceira (1999). There is a stochastic shock imposed on each dynamic equation. A similar model is used in Campbell and Viceira (2002: Chapter 5), where, however, a constant expected premium is used. Moreover, θ is the mean interest rate and λ and κ are adjustment coefficients. They represent adjustment speeds of the equity premium and interest rates toward the mean. Moreover, the dzt are the increments in the Brownian motions, possibly different for all three Brownian motions.

We can think of the wealth process of equation (A.1) as representing (long) bonds, equity, real estate, options, or commodities which could be managed by a hedge fund. In addition, some investment is undertaken with a risk-free interest rate, r_t. We then interpret the wealth as obtained from accumulated assets, a fraction being invested in risky hedge-fund assets, indicated by α_t, and a fraction invested in a risk-free investment, $(1 - \alpha_t)$. Risk-free here means that the return is known over the holding period. Thus, $\alpha_t(r_t + x_t)W_t$ represents the expected return from risky assets. On the other hand, the risk-free asset generates an expected return of $(1 - \alpha_t)r_t W_t$. Moreover, r^p is the return that the hedge fund promises to pay the investor. We may assume, for example, that the hedge fund attempts to attract investors by delivering high returns. Those promised returns will have to be paid the next time period, at least on average, if the hedge fund wants to be perceived as credible. This, however, means that the hedge fund has actually borrowed at a rate r^p.

Thus, in the long run, and on average, the wealth W_t of the hedge fund grows if the positive returns on both the risky assets and the risky free assets exceed the fee and the promised return r^p. Yet, if mostly the promised returns r^p are set too high, on average the hedge fund persistently needs positive inflow of fresh funds in order not to become insolvent.

1.4.3 The hedge fund management

The hedge-fund manager can view the above equation in a different way. As mentioned above, the hedge fund promises returns in order to obtain an inflow of funds. If r^p is set too high, the fund must persistently achieve a higher return usual market return, $r_t + x_t$, which is the asset return for risky assets (premium plus risk-free rate), in order to deliver the promised return r^p – or it must rely on a positive inflow of funds $\sigma_w W_t dz_t$ to be able to survive. Thus, it must persistently attract investors to the fund. But as mentioned above, this means that the promise to pay high returns, to attract investors, turns into the actual payment to the existing investors. The hedge fund thus has to pay the existing investors those returns.

So, in fact, one can say that the hedge fund borrows funds from investors (that guarantees returns for new investors) at a rate r^p but receives possibly only a smaller return, possibly in excess of r_t, and below, equal to, or above $\alpha_t(r_t + x_t) + (1 - \alpha_t)r_t$. When, positive returns $r_t + x_t$ and r_t are generated for the hedge fund, wealth rises for the investor on average at the rate $r_t + x_t$. The value of the hedge fund can rise too as long as the term in the Brownian motion $\sigma_w W_t dz_t$ is positive (or has an expected mean greater than zero).

1.4.4 The collapse of the hedge fund

If indeed r^p is set too high, when the inflow of new funds dries up and the increment in Brownian motion $\sigma_w W_t dz_t$ on average becomes negative, the hedge-fund wealth must necessarily decline, go to zero, and finally turn negative. The latter results will hold when suddenly funds are withdrawn. This actually happened since the middle of the year 2008. Figure 1.1 represents typical results of our simulations of equations (A.1)–(A.3). Curve W (wealth) represents the evolution through time of the market value of hedge-fund assets. Additional borrowing by the hedge fund from capital markets can also come in.[21] Curve D (debt) represents the evolution of hedge-fund borrowing over time. Figure 1.1 shows that although the hedge fund starts positive, it will end up in insolvency. The higher the promised returns are, the more likely it is that the hedge fund will become insolvent. Depending on whether the fund is leveraged, insolvency will come at different points in time. If the fund has borrowed from capital markets, the fund is technically insolvent as soon as the market value of

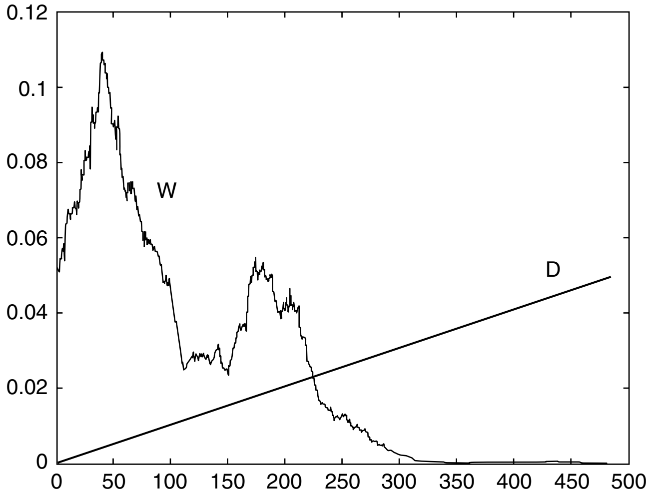


Figure 1.1 Simulation of the insolvency of the hedge fund

assets falls below the value of debt (where curve W intersects curve D). If the fund has no leverage, insolvency happens when the market value of assets falls close enough to zero that the fund cannot deliver promised return r^p.[22] At this point, with no further borrowing available or no new inflow of funds from new investors, original investors have lost their initial investment as the market value of hedge-fund assets falls to zero. The same results hold when there is a sudden decline of inflow of funds or when large withdrawals occur. Finally, note that if hedge fund only earns return at the rate of r_t, or less, the bankruptcy will come earlier, as in the Madoff case.

A recent article published in the *Financial Times* (Brewster 2008) evidenced that many investors in hedge funds had to withdraw funds suddenly in the period of the financial market meltdown, 2008–2009. This was due to other payment obligations, in order to make up for losses somewhere else, or owing to forced deleveraging, making 2008 a record year in its experience of such major redemptions. On the other hand, the hedge funds have tried to impose "gates" against fast exits (for example, some funds have split their assets into liquid and nonliquid baskets to make it harder for investors to get money back immediately). The news in the years 2008–2009 was full of those stories. If such a trend persists continued withdrawals are likely to precipitate hedge-fund bankruptcies and reveal the existence of Ponzi financing.

1.5 The regulatory context and implications

Having shown that promised high returns and use of leverage to generate above-average returns can result in some hedge funds running a Ponzi financing scheme, we turn to the regulatory context. The lack of regulatory reporting requirements facilitates such financing in that managers have the ability to claim to deliver returns in excess of actual returns. Further, as already described, the size of the hedge-fund industry raises the question of whether a new form of systemic risk is present, with the added issue that the risk profile of hedge funds is not fully understood.

A new set of risk analytics is required to develop more accurate quantitative measures of things such as performance, illiquidity exposure, probability of the fund's survival, and systemic risk. For this purpose, transparency and access to information is critical. In his written testimony to the U.S. House of Representatives, Andrew Lo (2008b) emphasized that any attempt to understand and measure systemic risk required greater transparency on the part of hedge funds, and he stressed the need to develop a new regulatory framework to that effect:

> The first order of business for designing new regulations is to develop a formal definition of systemic risk and to construct specific measures that are sufficiently practical and encompassing to be used by policymakers and the public. Such measures may require hedge funds and other parts of the shadow banking system to provide more transparency on a confidential basis to regulators, e.g., information regarding their assets under management, leverage, liquidity, counterparties, and holdings. (Lo 2008b: 2)

1.5.1 The hedge fund: definition and structure

There is no legal definition of a hedge fund. The term is absent from federal securities laws. As many as fourteen different definitions have been identified in government and industry publications (Vaughan 2003). A hedge fund can be thought of as a catch-all provision designating any privately organized pooled investment vehicle administered by professional investment managers whose interests are not sold in a registered public offering and which, as such, avoids registration under the Investment Company Act of 1940. The Dodd-Frank bill has now adopted this catch-all approach (hedge funds are largely defined by what they are not) and defines a hedge fund as any investment company that has avoided regulation under the Investment Company Act of 1940.

Hedge funds are typically structured as limited partnerships, a legal entity characterized by the presence of both general and limited partners, to achieve maximum separation of ownership and management. General partners have the authority to legally bind the partnership. As opposed to limited partners, general partners have joint and several liability for the debts of the partnership. A general partner typically manages the fund for a fixed fee (usually a percentage of assets under management) and a percentage of the gross profits from the fund (the "carry" or "carried interest"). The investors are limited partners with no managerial oversight. Investors might invest directly in the fund, or via another fund called a feeder fund (itself an offshore or U.S. fund), thus resulting in a layered structure that allows each investor to obtain the best possible tax treatment while allowing the hedge-fund manager to keep all assets in a single entity (the master fund).

1.5.2 Lack of regulatory oversight

The Investment Company Act is legislation passed by the U.S. Congress dealing with the regulation of investment companies. It applies to any fund that issues securities and "is or holds itself out as being engaged primarily … in the business of investing, reinvesting, or trading in securities."[23] Among other things, the Investment Company Act requires registration with the Securities and Exchange Commission (SEC) and sets reporting requirements to investors.

Unlike their mutual fund counterparts, most hedge funds fell outside the scope of the Investment Company Act by availing themselves of applicable exemptions, such as having 100 or fewer beneficial owners and not offering their securities in a public offering,[24] or because investors are all "qualified" high net-worth individuals or institutions.[25] This explains why hedge funds have deliberately chosen not to raise capital on public markets. Two recent examples of substantial private initial offerings are that of Oaktree Capital Management LLC, which sold approximately 14 percent of its equity for more than $800 million in May 2007, and Apollo Management LP, which privately raised $828 million in August 2007. Both transactions listed equity securities on Goldman Sachs & Co.'s nonpublic market, the GS Tradable Unregistered Equity OTC Market (GSTRuE).[26]

Hedge-fund advisers have also been exempt from regulation under the Investment Advisers Act of 1940.[27] The Investment Advisers Act is a companion statute to the Investment Company Act and was primarily designed to introduce record keeping and antifraud standards to the investment advisory profession. Under the Act, investment advisers must register with the SEC. An investment adviser is someone who, for compensation, engages in the business of advising others about the value of securities or the advisability of investing in, purchasing, or selling securities.[28] The SEC has authority to conduct audits and to examine the records of investment advisers.[29]

Hedge-fund managers have been typically exempt from registration under the private adviser exemption, whereby any investment adviser is exempt from registration who during the course of 12 months (1) has had fewer than fifteen clients and (2) neither holds himself out generally to the public as an investment adviser nor acts as an investment adviser to any investment company registered under the Investment Company Act.[30] For this purpose, only the hedge fund itself (as opposed to the investors in the fund) is considered to be a client of the hedge-fund manager. The rationale is that a hedge-fund manager will typically manage the assets of the fund on the basis of collective investment objectives rather than individual investors' financial goals and profiles. Because investors do not directly receive customized advice, it appears appropriate to view the fund itself rather than each investor as a client of the manager. Because of this technical point, most hedge-fund managers have been exempt from registration, because hedge-fund managers usually run fewer than fifteen hedge funds.

1.5.3 The SEC's attempt to impose mandatory registration

Concerned with the growth of the industry, and with the increased exposure of ordinary investors to such funds, the SEC sought to establish greater transparency through mandatory registration of hedge-fund advisers under the Advisers Act, taking the position that one can look through to the investors in the fund as clients of the adviser for the purposes of the private adviser exemption.[31] Under this interpretation, most hedge-fund managers would have had more than fifteen clients and would have had to register by February 1, 2006.

The expectation was that through registration the SEC would be able to gather "basic information about hedge fund advisers and the hedge fund industry," as well as "oversee hedge fund advisers," and "deter or detect fraud by unregistered hedge fund advisers."[32] Under this proposal, advisers would have had to adopt record-keeping procedures subject to periodic audits by the SEC and to supply information (financial statements) to investors concerning their results of operations.

Challenged by a hedge-fund manager, the SEC was eventually struck down for lack of statutory grounding by a DC circuit decision on June 23, 2006.[33] The decision rejected the SEC's interpretation of the word "client" as establishing a direct client relationship between investors and hedge-fund managers that would have been inconsistent with the

Advisers Act as a whole. The Court also highlighted that regulating investment companies on the basis of the number of investors would appear somewhat misaligned with the SEC's proclaimed concern with the national scale of the fund's activities. Reliance on some financial metric, such as the volume of assets under management, or the level of leverage, would seem more appropriate. This required comprehensive new legislation rather than twisted interpretation of the existing statutes.

1.5.4 Consequences of lack of regulatory oversight

One consequence of the lack of regulatory oversight has been that hedge funds could participate in highly complex financial transactions inaccessible to other regulated market participants, thus furthering the growth of the industry and exacerbating the potential for systemic risk. Another consequence was that hedge funds typically remained secretive about their positions and strategies, even to their own investors, asserting the proprietary nature of trading techniques and algorithms.

Even though they were not obligated to, some hedge-fund managers did voluntarily register with the SEC so as to give the market a higher level of confidence and potentially to minimize the amount of due diligence performed by new investors. It turns out that Madoff himself had registered with the SEC in September 2006. One issue is that the SEC may not have the resources needed to examine investment advisers on a regular basis. If so, mandatory registration could lull investors into a false sense of security. Madoff himself avoided attracting scrutiny in spite of repeated letters to the SEC accusing him of running a large Ponzi scheme. The SEC did not pay specific attention to the basic fact that the fund relied on a small auditing firm with no other large clients and no reputation on Wall Street. The SEC has tended to focus on funds with high-risk profiles and high-risk trading strategies. Madoff avoided drawing attention by supposedly engaging in plain-vanilla trading.

The issue is that in the absence of detailed financial statements and the full disclosure of detailed trading strategies, there is typically not enough information to allow investors to conduct basic due diligence regarding the performance of funds. The lack of a regulatory framework facilitated the ability for managers to claim to deliver returns in excess of actual returns, in that there is no transparency regarding the fund's financials. One rationale for limited regulatory oversight of the hedge-fund industry has traditionally been that high-stakes investors should be capable of protecting themselves. However, this is untenable given the trend toward retailization, the involvement of institutional investors, and the potential contribution to systemic risk.

1.5.5 Guiding principles for regulatory reform

Today, in the context of the current financial crisis, in a post-Enron, post-Worldcom, post-Madoff world, the trust in financial markets has been severely damaged. There is no reason to believe that private-sector actors such as lawyers, accountants, internal-risk managers, and external rating agencies can on their own ensure the reliability of the system and well-functioning markets. There are two distinct goals of financial regulation. One goal is to maintain market integrity and to protect investors from fraud. The other is to monitor systemic risk and to preserve the stability of the financial system. Ponzi schemes of the magnitude of the recent Madoff scandal have implications for both.

To elaborate on that point, hedge-fund failures can contribute to destabilize the economy and the financial sector because of the size of the industry and the potential liquidation and unloading of positions, which may contribute to creating and accelerating downward price spiral. Kundro and Feffer (2008) and Christory et al. (2008) have shown that a majority of hedge-fund failures occur because of operational issues or/and fraud (such as misappropriation of funds, false valuations, and Ponzi schemes), rather than unsuccessful investment strategy.

Both adequate protection against fraud and regulation of systemic risk require that the regulator have access to better financial information. Every financial firm of a significant size should be disclosing timely financials (balance sheets and profit-and-loss statements). Specifically, with respect to hedge funds, transparency regarding valuation policies, performance attributes, portfolio exposures, and risk metrics, and limited disclosure of positions has also been suggested (Mertzger 2008). For this purpose, new inspection systems need to be developed for the regulator to be able to process this vast quantity of information.

One issue is whether there should be different regulators for protection against fraud and guaranteeing the stability of the financial system. The general consensus is that the Fed is more inclined to monitor systemic risk, whereas the SEC has traditionally been focused on providing investors with accurate financial information and the prevention of fraud so as to protect investors. In some jurisdictions, in the U.K. for example, there is a unique regulator for the entire financial services industry, the Financial Services Authority (FSA). A single regulator may be in a better position to address both fraud and systemic risk concerns more efficiently. Either way, all reported financial information should be consolidated for the purpose of monitoring systemic risk.

The regulatory framework had been focused on form over substance and had not kept up with the development of markets and sophistication of new products that have contributed to blurring the traditional roles of firms in the financial sector. If hedge funds are performing banking functions (e.g., extending loans to troubled companies), they should be regulated as such. If banks are operating like hedge funds (e.g., merchant banking and proprietary trading), or have hedge funds as operating branches, those activities should be subject to banking regulation as well. It is inconceivable that the asset-management operations of Bear Stearns were left essentially unregulated, with no strict requirements for transparency, to ultimately bring the firm down.

1.5.6 Proposals under the Obama administration

The Obama administration has recently passed the Dodd-Frank Act, a major legislation designed to reform the financial sector and prevent future bailouts. The Dodd-Frank Act passed the House of Representatives on June 30, 2010, and was approved by the Senate on July 15, 2010. There are now registration requirements for investment advisers to private funds with assets under management of $150 million or more. Hedge funds will now have to register with the SEC and disclose assets under management, use of leverage (including off- balance sheet leverage), counterparty credit risk exposures, trading practices, valuation policies, types of assets held, and side arrangements or side letters (whereby certain investors in a fund obtain more favorable rights or entitlements than other investors). Hedge funds will also have to have assets audited by public accountants. Finally, the Act also specifically contemplates that the SEC be empowered to collect systemic risk data, reports, examinations and disclosures. The SEC would make this information available to the Financial Services Oversight Council, a newly created Council designed to look after the stability of the financial system as a whole.

The Act also significantly restricts proprietary operations undertaken by commercial banks (provision known as the Volcker rule). Banks can place up to 3 percent of their Tier 1 capital in hedge fund and proprietary trading investments. The other aspect of the rule is that banks are prohibited from holding more than 3 percent of the total ownership interest of any private equity investment or hedge fund. This falls short of a complete disallowance of proprietary desks, which had been originally suggested and would have been equivalent to restoring Glass-Steagall.

1.5.7 Proposals in the European Union

While systemic risk is cross-border in nature, in the European Union (EU) regulatory oversight is still currently largely national. Data reporting is voluntary and incomplete, and there is currently no mechanism for sharing the information between regulators of member countries. Aware of this issue, the European Commission has proposed a hedge-fund directive, the Directive on Alternative Investment Fund Managers that would create a comprehensive regulatory and supervisory framework for hedge funds at the European level. The proposed directive will provide harmonized regulatory standards for all hedge funds (above a threshold of €100 million of assets) and managers. These should include minimum standards in relation to governance, ongoing capital requirements, and processes, as well as enhanced transparency requirements for supervisors and the general public. More specifically, some of the key requirements the directive provides for include

1. valuation of fund assets to be undertaken by an independent appraiser;
2. an annual report to be made available to investors and the regulator with audited balance sheet and income and expenditure account, as well as a report on the activities of the financial year;
3. disclosure of uses and sources of leverage (to be aggregated and shared with other regulators in the EU), as well as limits to the maximum amount of leverage;
4. disclosure of investment strategies and objectives of the fund;
5. description of the fund's liquidity risk management;
6. disclosure of dominant or controlling interests in listed or non-listed companies.

Regulators are invited to communicate information to regulators of other member states where this is relevant for monitoring systemic risk. Further, aggregated information relating to the fund activities are to be communicated on a quarterly basis by each regulator to the Economic and Financial Committee for the Council of the European Union.

1.6 Conclusion

This chapter has shown a model that can be used to describe situations in asset management where the use of leverage and the promise of above-average returns to investors can result in some hedge funds following a pattern of Ponzi financing. Unlike in a pure Ponzi scheme, this situation could develop inadvertently without necessarily involving fraudulent misrepresentation. The illiquidity of some positions, the computation of the most accurate mark of a security under mark-to-market accounting rules, for example, might lead to a declared return that in the end exceeds actual return. This issue is further accentuated when there are biases in the data and a lack of transparency regarding valuation policies and leverage. Given that hedge funds are investing in every kind of assets, and that the size of assets under management has exploded in the past ten years, this situation carries systemic risk that is not properly understood nor measured. In that respect, the new regulatory framework developed by the Obama administration is a welcome development.

One main challenge, going forward, is the measuring and monitoring of systemic risk. The data-gathering effort proposed by the Obama Administration would allow the regulator to consider the financial system as a whole, potentially aggregating the information across institutions to develop systemic scenario analyses. However, regardless of which regulatory agency is in charge on monitoring systemic risk (for example, a new entity such as the Financial Services Oversight Council), some key issues will include the ability to streamline the collection of a comprehensive consolidated systemic-risk database and to develop new formal measures of systemic risk.

Appendix: A stochastic dynamic model with preferences

The background of the stochastic dynamic model of wealth accumulation by a hedge fund in Section 1.4 is an intertemporal decision model with preferences that allows for consumption choice. In the model, we have modeled the payoffs $c_t W_t$ and $r^p W_t$ as well as the fraction allocated to risky assets, as determined by some law of motion depending on time. Yet the payoffs as well as the fraction allocated to risky assets can be determined by an intertemporal decision model. In doing so, we use, as in Section 1.4, a model with mean reversion in returns where the expected returns move in the long run to some mean. This is one of the most common assumptions in recent asset return modeling (see Munk et al. 2004). We want to discuss here a stochastic consumption and portfolio choice model with preferences which has two choice variables and three state variables. The basics of such model are analytically treated in Semmler (2006: Chapter 16). The herein presented extended version, as compared to the model in Section 1.4, also needs to be solved numerically, yet by a dynamic programming algorithm. As state equations, we again can use three Brownian motions as stochastic processes, a process for wealth, a mean-reverting interest-rate process, and an equity premium process. This model explores consumption and portfolio choices at each point of the state space. We hereby can obtain the decision variables. To avoid a third state equation one could presume a constant expected equity premium as in Campbell and Viceira (2002: Chapter 3), see also Semmler (2006), Munk et al. (2004), and Wachter (2002). Next, let us introduce the full dynamic model, which can be written for power utility as:

$$\max_{\alpha,C} \int_0^{\infty} e^{-\delta t} \frac{(c_t W_t)^{1-\gamma}}{1-\gamma} dt \tag{A.4}$$

s.t.

$$dW = \{[\alpha_t(r_t + x_t) + (1-\alpha_t)r_t]W_t - c_t W_t - r^p W_t\}dt + \sigma_\omega W_t dz_t \tag{A.5}$$

$$dx_t = \lambda(\overline{x} - x_t)dt + \sigma_x dz_t \tag{A.6}$$

$$dr_t = \kappa(\theta - r_t)dt + \sigma_r dz_t \tag{A.7}$$

Hereby, W_t denotes total wealth, r_t, the short term interest rate, α_t, the fraction of wealth held as equity, x_t, the equity premium, and $\overline{x}_t$, an expected mean equity premium which we assume to be a constant (with a stochastic shock imposed on it). The parameter θ is the mean interest rate, λ and κ are adjustment coefficients, and dz_t the increment in Brownian motion. The terms $c_t W_t$ and $r^p W_t$ represent the payoffs for consumption and guaranteed returns. Such a model can be solved through dynamic programming (see Semmler 2006). Section 1.4 has introduced and solved a simplified version of this model without choice variables.

Acknowledgments

This chapter was presented at the International Business and Economy Conference (IBEC) 2010 and awarded the Best Finance Paper Award. This chapter was also presented at the Eastern Economic Association 36th Annual Conference.

Notes

1. Illiquidity exposure can induce serially correlated returns to the extent that mark-to-market accounting requires that a market value be assigned to portfolio securities, even when there is no active market and comparable transactions as is the case for illiquid investments. There is some discretion on the part of the hedge-fund manager to compute the most accurate mark for the security. Returns calculated on the basis of such computations typically exhibit lower volatility and higher serial correlation than true economic

returns. There is the issue of "smoothed" returns, where, for example, a hedge-fund manager might obtain different quotes for a given illiquid security from different broker-dealers, and pick a linear average as the most accurate mark, thereby minimizing volatility. There is the issue of serial correlation if the broker-dealers do not frequently update their quotes, as is likely when there is a low trading volume. Another source of serial correlation is performance smoothing, where hedge-fund managers deliberately fail to report a portion of gains in profitable months to offset potential future losses.

2. This led to major regulatory overhaul designed to restore confidence in capital markets, to promote the disclosure of important market-related information, and to protect investors against fraud. The Securities Act of 1933 was the first major legislation to regulate the new issue of securities to the general public and to protect investors against fraud. The Securities Exchange Act of 1934, designed to regulate secondary markets, created the Securities and Exchange Commission and required ongoing disclosure for publicly traded securities.
3. Hedge Fund Intelligence March 2009.
4. HFR Global Hedge Fund Industry Report, Fourth Quarter, 2008.
5. HFR Global Hedge Fund Industry Report, Fourth Quarter 2008.
6. Deutsche Bank, 2010 Alternative Investment Survey.
7. As per the Bureau of Economic Analysis, U.S. GDP was $8.793 trillion in 1998 and $14.441 trillion in 2008.
8. This has been described as having option-like features in that the manager has an incentive for volatility in returns (Cochrane 2005).
9. Using the TASS database.
10. On the basis of HFR industry report through 2008, Q3, and Credit Suisse for 2008, Q4, projections, Lo (2008b) estimated $1.6 trillion of net assets and $3.7 trillion of total market positions at the end of 2008.
11. The firm's equity had grown to $3.6 billion while assets had grown to $102 billion at the end of 1995. See Lowenstein (2001) p78.
12. See Lowenstein (2001) p26: from the very start, it was always contemplated that Long Term Capital Management would be heavily leveraged, twenty to thirty times of capital or more, in order to make a decent profit on very tiny spreads.
13. John Paulson's verbal testimony to the U.S. House of Representatives, Committee on Oversight and Government Reform, November 13, 2008 Hearing on Hedge Funds.
14. UBS Hedge Fund Report (February 2009).
15. The participants in the survey are investors in hedge funds that collectively manage more than $1.1 trillion in hedge-fund assets.
16. See to that effect Philip Falcone's verbal testimony to the US House of Representatives, Committee on Oversight and Government Reform, November 13, 2008, Hearing on Hedge Funds.
17. Information flux could include insider information obtained through informal networks, as the recent case of the Galleon Group may demonstrate. The FBI arrested the founder of this hedge fund on allegations of insider trading.
18. There is the well-known Paradox pointing to those scale effects: "The people who most need the money are worst credits risks and thus cannot get a loan, whereas people who least need the money are best credit risks and thus once again the rich get richer" (Tooby and Leda 1996).

19. Early contributions include Fisher 1930; Keynes 1936; Vickrey 1947; Duesenberry 1949; Hicks 1950; Pigou 1951; Friedman 1957; Friend and Kravis 1957; and Modigliani and Ando 1960. For more recent work, see for example Saltz 1999; Dynan et al. 2004; and Chakrabarty et al. 2008.
20. Typically between 1% and 2% of assets under management. Performance-based fees can be up to 20% of returns.
21. Indeed, we have seen in Section 1.3 that hedge funds use leverage, albeit in more limited proportions than banks. Prior to the current financial crisis, around 70 percent of hedge funds were levered less than 2 to 1. At the end of 2008, on average, the hedge-fund industry had a 1-to-1 debt-to-equity ratio. But there are great differences in leverage ratios between funds, sometimes going up to 30 to 1.
22. We use the term "insolvency" rather loosely in that technically the fund is not insolvent simply because it cannot deliver promised returns to investors. However, as we have discussed, if the hedge fund wants to be perceived as credible, on average it is committed to delivering r^p to investors. In essence, in terms of understanding long-term sustainability of the fund, this can be interpreted as the hedge fund having borrowed at a rate of r^p.
23. 15 USC ß 80a-3 (a) (1) (A).
24. 15 USC ß 80a-3(c)(1).
25. 15 USC ß 80a-3(c) (7).
26. Apollo subsequently announced it would transfer the securities to the New York Stock Exchange. It has been suggested that Apollo's GSTRuE offering may therefore have been a transitory step designed to save time initially and to delay the time-consuming registration process until after capital had been raised (Davidoff 2008).
27. 15 USC ß 80b-1.
28. 15 USC ß 80b-2 (11).
29. 15 USC ß 80b-4.
30. 15 USC ß 80b-3 (b) (3).
31. See Rule 203 (b) (3)–2 under the Investment Advisers Act, 17 CFR ß 275.203 (b) (3)–2.
32. Id.
33. Goldstein v. SEC, 451 F.3d 873 (DC Cir. 2006). Interestingly enough, the Fed was opposed to this mandatory registration proposal. The SEC subsequently tightened restrictions on investors who could invest in both hedge funds and private-equity funds (investors must have owned at least $2.5 million in investments). This measure failed to address systemic risk and was only concerned with individual risk to investors.

References

Adrian, T., and Shin, H. S. (2008) "Liquidity, Financial Cycles and Monetary Policy: Current Issues and Economics and Finance," *Federal Reserve Bank of New York*, 14 (1).

Agarwal, V. and Naik, N. (2000) "On Taking the 'Alternative' Route: The Risks, Rewards, and Performance of Persistence of Hedge Funds," *Journal of Alternative Investments*, 2 (2): 6–23.

Amin, G. and Kat, H. (2003) "Stocks, Bonds and Hedge Funds: Not a Free Lunch!" *Journal of Portfolio Management*, 29 (4): 113–118.

Brewster, D. (2008) "Money Flows Out of Hedge Funds at Record Rate," *Financial Times*, December 30.

Brooks, C. and Kat, H. (2002) "The Statistical Properties of Hedge Fund Index Returns and Their Implications for Investors," *Journal of Alternative Investment*s, 5 (3): 26–44.

Brunnermeier, M. and Nagel, S. (2004) "Hedge Funds and the Technology Bubble," *The Journal of Finance*, 59 (5): 2013–2040.

Brunnermeier, M and Pedersen, L. (2009) "Market Liquidity and Funding Liquidity," *Review of Financial Studies*, 22 (6): 2201–2238.

Brunnermeier, M. and Sannikov, Y. (2009) "A Macroeconomic Model with a Financial Sector," Working Paper, Princeton University, Princeton, NJ.

Bernardo, A. E. and Ledoit, O. (2000) "Gain, Loss, and Asset Pricing," *Journal of Political Economy*, 108 (1): 144–172.

Cagetti, M. and De Nardi, M. (2006) "Entrepreneurship, Frictions, and Wealth," *Journal of Political Economy*, 114 (5): 835–870.

Campbell, J. and Viceira, L. (1999) "Consumption and Portfolio Decisions When Expected Returns Are Time Varying," *The Quarterly Journal of Economics*, 114 (2): 433–495.

Campbell, J. and Viceira, L. (2002) *Strategic Asset Allocation*, Oxford: Oxford University Press.

Chakrabarty, H., Katayama, H. and Maslen, H. (2008) "Why Do the Rich Save More? A Theory and Australian Evidence," *Economic Record*, Supplement, 84: S32–S44.

Chan, N., Getmansky, M., Haas, S., and Lo, A. (2006) "Do Hedge Funds Increase Systemic Risk?" *Economic Review*, 91 (4): 49–80.

Chen, Z. W. and Knez, P. J. (1996) "Portfolio Performance Measurement: Theory and Evidence," *Review of Financial Studies*, 9 (2): 551–556.

Chiarella C., Hsiao, C. and Semmler, W. (2007) "Intertemporal Investment Strategies under Inflation Risk," Research Paper Series 192, Quantitative Finance Research Centre, University of Technology, Sydney.

Christory C., Daul, S. and Giraud, J. R. (2006) "Quantification of Hedge Fund Default Risk," *The Journal of Alternative Investments*, 9 (2): 71–86.

Cochrane, J. (1999) "New Facts in Finance," *Economic Perspectives*, 23 (3): 36–58.

Cochrane, J. (2005) "Betas, Options, and Portfolios of Hedge Funds," available online at http://www.slidefinder.net/B/Betas_Options_ Portfolios_ Hedge_Funds/2799231 (accessed September 18, 2009).

Davidoff, S. (2008) "Paradigm Shift: Federal Securities Regulation in The New Millennium," Brook. J. Corp. Fin. & Com. L. 339 (367).

Drehmann, M. (2007) "Discussion on *Banks, Markets and Liquidity* by F. Allen and E. Carletti," in C. Kent and J. Lawson, *The Structure and Resilience of the Financial System*, Reserve Bank of Australia RBA Annual Conference Volume 2007, available at http://ideas.repec.org/s/rba/rbaacv.html (accessed on September 18, 2010).

Drehmann, M. and Nikolaou, K. (2009) "Funding Liquidity Risk: Definition and Measurement," ECB Working Paper No. 1024. Available online at http://ssrn.com/abstract=1338092 (accessed September 18, 2010).

Duesenberry, J. (1949) *Income, Saving, and the Theory of Consumer Behavior*, Cambridge, Mass.: Harvard University Press.
Dynan, K., Skinner, J., and Zeldes, S. (2004) "Do The Rich Save More?" *Journal of Political Economy*, 112 (2): 397–444.
Fisher, I. (1930) *The Theory of Interest*, New York: Macmillan.
Frankel, T. (2009) Statement before the Committee on Financial Services of the US House of Representatives, Washington, DC, January 5.
Friedman, M. (1953) "Choice, Chance, and the Personal Distribution of Income," *Journal of Political Economy*, 42 (4): 277–290.
Friend, I. and Kravis, I. B. (1957) "Consumption Patterns and Permanent Income," *American Economic Review*, 47 (2): 536–555.
Fung, W. and Hsieh, D. (2000) "Performance Characteristics of Hedge Funds and CTA Funds: Natural versus Spurious Biases," *Journal of Quantitative and Financial Analysis*, 35 (3): 291–307.
Getmansky, M., Lo, A. and Makarov, I. (2004) "An Econometric Model of Serial Correlation and Illiquidity in Hedge Fund Returns," *Journal of Financial Economics*, 74 (3): 529–609.
Gilli, M., Schumann, E., Cabej, G., and Lula, J. (2010) "Replicating Hedge Fund Indices with Optimization Heuristics," May 6. Available online at http://ssrn.com/abstract=1601708 (accessed September 18, 2010).
Goetzmann, W. and Ross, S. (2000) "Hedge Funds: Theory and Performance," Yale School of Management Working Paper No. F-52B, New Haven, Conn.
Goetzmann, W., Ingersoll, J., Spiegel, M., and Welch, I. (2003) "Sharpening Sharpe Ratios," *Financial Planning*, 33 (1): 49–58.
Griffin, K. (2008) "Testimony to the House Committee on Oversight and Government Reform," Prepared for the US House of Representatives, Committee on Oversight and Government Reform, Hearing on Hedge Funds, Washington, DC, November 13.
Heidorn, T., Kaiser, D. and Roder, C. (2009) "Performance Measurement of Hedge Funds Portfolios in a Downside Risk Framework," *The Journal of Wealth Management*, 12 (2): 101–112.
Hicks, J. R. (1950) *A Contribution to the Theory of the Trade Cycle*, London: Oxford University Press.
Ibbotson, R. and Cheng, P. (2005) "Sources of Hedge Fund Returns: Alphas, Betas, and Costs," Yale ICF Working Paper No. 05–17, New Haven, Conn.
Keynes, J. M. (1936) *The General Theory of Employment, Interest, and Money*, New York: Harcourt, Brace.
Kundro, C. and Feffer, S. (2003) "Hedge Funds Fail Due to Operational Risk," study by Capco, abstract available at http://www.edge-fund.com/Capco03.pdf (accessed September 18, 2010).
Lo, A. (2001) "Risk Management for Hedge Funds: Introduction and Overview," *Financial Analysts Journal*, 57 (6): 16–33.
Lo, A. (2002) "The Statistics of Sharpe Ratios," *Financial Analysts Journal*, 58 (4): 36–52.
Lo, A. (2005) "The Dynamics of the Hedge Fund Industry," *CFA Digest*, 35 (4): 87–90.
Lo, A. (2008a) "Where Do Alphas Come From? A New Measure of the Value of Active Investment Management," *Journal of Investment Management*, 6 (3): 6–34.

Lo, A. (2008b) "Hedge Funds, Systemic Risk, and the Financial Crisis of 2007," written testimony prepared for the U.S. House of Representatives, Committee on Oversight and Government Reform, Hearing on Hedge Funds, Washington, DC, November 13.

Lowenstein, R. (2001) *When Genius Failed: The Rise and Fall of Long-Term Capital Management*, New York: Random House.

Malkiel, B. and Saha, A. (2004) "Hedge Funds: Risk and Return," Princeton University Working Paper, Princeton, NJ.

Mamoghli, C. and Daboussi, S. (2009) "Performance Measurement of Hedge Funds Portfolios in a Downside Risk Framework," *The Journal of Wealth Management*, 12 (2): 101–112.

Minsky, H. (1986) *Stabilizing an Unstable Economy*, New Haven, Conn.: Yale University Press.

Modigliani, F. and Ando, A. (1960) "The 'Permanent Income' and 'Life Cycle' Hypothesis of Saving Behavior: Comparison and Tests," in Irwin Friend and Robert Jones (eds.), *Consumption and Saving*, Philadelphia, Pa.: University of Pennsylvania, vol. II, pp. 49–174.

Munk, C., Sørensen, C., and Nygaard Vinther, T. (2004) "Dynamic Asset Allocation under Mean-Reverting Returns, Stochastic Interest Rates and Inflation Uncertainty: Are Popular Recommendations Consistent with Rational Behavior?" *International Review of Economics and Finance*, 13 (2): 141–166.

Pigou, A. C. (1951) "Professor Duesenberry on Income and Savings," *Economic Journal*, 61 (244): 883–885.

Platen, E. and Semmler, W. (2009) "Asset Markets and Monetary Policy," Research Paper Series 247, Quantitative Finance Research Centre, University of Technology, Sydney.

Posthuma, N. and Van der Sluis, P. (2003) "A Reality Check on Hedge Fund Returns," Free University Amsterdam.

Ruder, D. (2008) "Testimony to the House Committee on Oversight and Government Reform," prepared for the U.S. House of Representatives, Committee on Oversight and Government Reform, Hearing on Hedge Funds, Washington, DC, November 13.

Saltz, I. (1999) "An Examination of the Causal Relationship between Savings and Growth in the Third World," *Journal of Economics and Finance*, 23 (1): 90.

Semmler, W. (2006) *Asset Prices, Booms, and Recessions*, Heidelberg and New York: Springer.

Sharpe, W. (1994) "The Sharpe Ratio," *Journal of Portfolio Management*, 21 (1): 49–58.

Sortino, F., and Price, L. (1994) "Performance Measurement in a Downside Risk Framework," *Journal of Investing*, 3 (3): 59–64.

Tooby, J. and Cosmides, L. (1996) "Friendship and the Banker's Paradox: Other Pathways to the Evolution of Adaptations for Altruism," *Proceedings of the British Academy*, 8 (2): 119–143.

Vaughan, D. (2003) "SEC Roundtable on Hedge Funds," May 13, comments available online at http://www.sec.gov/spotlight/hedgefunds/hedge-vaughn.htm (accessed September 18, 2010).

Vickrey, W. (1947) "Resource Distribution Patterns and the Classification of Families," *Studies in Income and Wealth*, NBER, Volume 10.

Wachter, J. (2002) "Portfolio and Consumption Decisions under Mean-Reverting Returns: An Exact Solution for Complete Markets," *Journal of Financial and Quantitative Analysis*, 37 (1): 63–91.
Xiong, J. (2009) "Impact of Size and Flows on Performance for Funds of Hedge Funds," *Journal of Portfolio Management*, 35 (2): 120–130.

2
Inferring Risk-Averse Probability Distributions from Option Prices Using Implied Binomial Trees

Tom Arnold, Timothy Falcon Crack, and Adam Schwartz

2.1 Introduction

We generalize the Rubinstein (1994) risk-neutral implied binomial tree (R-IBT) model to a physical-world risk-averse implied binomial tree (RA-IBT) model. The R-IBT and RA-IBT trees are bound together via a relationship requiring a risk premium (or a risk-adjusted discount rate) on the underlying asset at any node. The RA-IBT provides a powerful numerical platform for many empirical financial option and real option applications; these include probabilistic inference, pricing, and utility theory applications.

For ease of exposition, we have estimated a constant risk premium[1] RA-IBT using Standard & Poor's 500 index options. In our implied tree, this is consistent with assuming a representative agent with a power-like utility function where the constant relative risk aversion (CRRA) parameter is allowed to vary across states and through time. With these assumptions, we estimate the pricing kernel (marginal rate of substitution) and implied relative risk aversion (RRA) of our agent and compare and contrast our results with other authors' findings. Other empirical applications can be found in Arnold et al. (2009).

2.2 Motivation and literature review

2.2.1 Motivation/introduction of risk-averse trees

The traditional binomial tree model of Cox, Ross, and Rubinstein (CRR) (1979) is very powerful, but it is constrained in many respects. The CRR model cannot, for example, reproduce some well-known empirical results (e.g., fat tails, skewness, volatility smiles, etc.), and any probabilistic inferences from a CRR tree must be risk-neutral inferences not risk-averse probability inferences from the physical economy.

The Rubinstein (1994) implied binomial tree (R-IBT) generalizes the CRR model to fit the prices of a series of traded options of the same maturity. The R-IBT allows significant deviations from lognormality of prices (and from normality of continuously compounded returns), allows the up-jump probability to vary throughout the tree, and allows the local volatility structure to vary throughout the tree. The R-IBT model is still constrained, however, in that the probability structure is risk-neutral; and, therefore, like the CRR model, it does not allow probabilistic inferences about the physical economy.

In this chapter, we begin with a generalization of the risk-neutral CRR model to a risk-averse binomial tree (RA-BT). We then present a generalization of the risk-averse RA-BT to a risk-averse implied binomial tree (RA-IBT). Like the R-IBT model, the RA-IBT model captures volatility smiles and varying local volatility. It should be noted, however, that both the RA-BT and the RA-IBT have one extra input compared with the CRR and R-IBT models: The risk-averse trees need to be supplied with a risk premium (or a risk-adjusted discount rate) at every node to make them estimable. This risk premium feeds into a relationship that drives a transformation from the risk-neutral trees to the risk-averse trees.

For empirical ease, we impose a constant risk premium in the RA-IBT estimations in this chapter. Whether the risk premium is imposed directly, or derived from utility assumptions for a representative agent, to each such risk premium function over the nodes of an RA-IBT there corresponds a different implied risk-averse probability distribution for the future prices of the underlying asset; this, in turn, implies a unique, fully specified stochastic process for the underlying asset prices. Sensitivity analysis is thus essential for any inferences.

The CRR, RA-BT, R-IBT, and RA-IBT models can each be used to value and hedge both European-style and American-style options. Neither the R-IBT model nor the RA-IBT model (under the conditions we develop in this chapter) can be calibrated using American-style options.

2.2.2 Literature review

Rubinstein (1994: 793, footnote 25) describes implied trees developed by Hayne Leland (1980) that use "subjective probabilities" (i.e., an individual investor's non-risk-neutral probabilities). Stutzer (1996) also infers "subjective" (i.e., risk-averse) probability densities from options data. Stutzer differs from us, however, in that he uses diffusions rather than binomial trees, he requires historical data that are not needed here, and he uses the risk-averse density to estimate the risk-neutral density for risk-neutral pricing (the focus of his paper), whereas our focus is the risk-averse density itself.

Jackwerth and Rubinstein (1996) infer probabilities from option prices using binomial trees. However, Jackwerth and Rubinstein differ from us because they use risk-neutral probabilities, whereas we use risk-averse probabilities. We do though use their "smooth" objective function in estimating our R-IBT.

Jackwerth (2000) recovers risk-neutral densities from S&P 500 options and uses historical realized index returns to approximate subjective (i.e., risk-averse) densities. Jackwerth is thus able to infer aggregate absolute risk-aversion functions for different states. When comparing pre- and post-Crash of 1987 data, changes in the implied risk-neutral densities that are not accompanied by changes in the risk-averse densities imply changes in absolute risk-aversion functions that appear inconsistent with any sensible economic theory (e.g., Jackwerth finds significantly negative absolute risk aversion). Jackwerth suggests that overpriced put options may explain the inconsistency, and he is able to construct profitable trading strategies that appear to exploit this potential mispricing. Jackwerth (1999) acknowledges that another explanation for inconsistent risk-aversion functions is that his risk-averse distributions built using historical observations may differ from true ex-ante risk-averse distributions (Jackwerth 1999: 445–446). However, he rejects this as an explanation because his trading strategies appear profitable (supporting the hypothesis that his inconsistent risk-aversion functions are driven by exploitable mispriced options, not by poor estimates of subjective probability). In contrast, we show that we get wholly positive risk-aversion functions for the same period and underlying index as Jackwerth's, but using implied techniques for both the risk-averse and risk-neutral densities (see Section 5 for details).

Jackwerth (2004) is an expanded and updated version of Jackwerth (1999). Jackwerth (2004) also has an expanded section on implied risk aversion and useful summary tables of categorized literature. Jackwerth (2004) labels his earlier findings (i.e., those in Jackwerth 2000) as a "pricing kernel puzzle." The puzzle is that, although the implied marginal utility of wealth function should be monotonically decreasing in wealth, Jackwerth's empirical estimates of it after the Crash of 1987 are locally increasing in wealth near the initial wealth level. This suggests that, in these ranges of wealth, the representative agent is risk-seeking not risk-averse. Ait-Sahalia and Lo (2000: 36, Fig. 3) find a similar locally humped plot of the scaled marginal rate of substitution using S&P 500 futures prices. Our chapter is similar to Jackwerth (2000, 2004) and Ait-Sahalia and Lo (2000) in that we recover both risk-neutral and risk-averse distributions and compare them, but we differ in that we infer our ex-ante risk-averse distributions from option prices using the RA-IBT rather than from a backward-looking historical time series of the underlying, as do these authors.

Another strand of the literature includes Bliss and Panigirtzoglou (2004) and Alonso et al. (2009) and several other papers they cite. Like Ziegler (2007), these authors exploit the Breeden and Litzenberger (1978) result, rather than implied binomial trees, to derive risk-neutral densities. They then calibrate the parameters of a chosen utility function that is used to risk-adjust the risk-neutral density. The objective of the calibration is that the risk-averse density should best explain subsequently realized returns (see further discussion in Section 5). This allows them to discuss implied risk aversion. Their work is closely related to our own, except that rather than use implied binomial trees, they use numerical smoothing techniques to account for volatility smiles over a range of option strikes, and they use the Breeden and Litzenberger (1978) result. Our approach has two advantages over these approaches. First, our implied tree is guaranteed to be arbitrage-free (assuming there are no-arbitrage opportunities among the quoted option prices), whereas the Bliss and Panigirtzoglou (2004) numerical smoothing techniques are not guaranteed to be arbitrage-free. Second, our approach uses a simple numerical estimation without splines or smoothing (we estimate an R-IBT and then apply a simple direct transformation to it).

Blackburn (2006) focuses on the time-series properties of risk aversion and whether the representative agent's utility is time-separable or not. Like Bliss and Panigirtzoglou (2004) and Alonso et al. (2009), he exploits the Breeden and Litzenberger (1978) result and uses splines to derive the risk-neutral density. Unlike these authors, Blackburn argues that a calibration that maximizes the forecast ability of the risk-averse density is inappropriate (because it looks ahead in a way not possible when the agent is making decisions). Instead, Blackburn obtains a risk-aversion estimate using five years of historical data. Blackburn (2006) thus differs from us in that he does not use trees at all, and he uses historical data to estimate the risk-aversion parameter.

2.3 A risk-averse binomial tree (RA-BT) model

We now derive the RA-BT model, so that the implied version (i.e., the RA-IBT) can be developed in Section 4. We begin by noting that generating the RA-BT or RA-IBT trees relies upon three interrelated technical steps. First, we derive the functional form of the transformation between the risk-neutral and risk-averse trees. Second, we need to generate a risk premium or risk-adjusted discount rate at every node of the RA-IBT to feed into the transformation in the first step. Third, we need to combine the first two steps and propagate risk-averse probabilities through the risk-averse tree.

To work our way toward establishing the first two steps in the case of the RA-BT, recall that a continuous-time option pricing model using the risk-adjusted probability measure requires a stochastic path-dependent risk-adjusted discount rate; no single risk-adjusted discount rate can capture the changes in the option's risk associated with the moneyness of the option.[2] Black and Scholes recognize this with their "instantaneous CAPM" approach to deriving the Black–Scholes partial differential equation (Black and Scholes 1973: 645–646; Ingersoll 1987: 323–324). However, the (Black–Scholes) model that emerges is difficult to interpret with respect to the physical world because the risk-averse probability parameters fall out of the calculation.

The risk-averse RA-BT model is similar to a discretized version of the original Black–Scholes instantaneous CAPM derivation that allows for changing risk-adjusted discount rates. The numerical discretization allows us to infer physical-world parameters from the tree – an inference not explicitly available in the closed-form continuous-time (i.e., the Black–Scholes world) limit of the RA-BT pricing model.

To generate the RA-BT model, begin with the assumptions of the CRR model as follows. Consider an asset with spot price S_t at time t. From time t to time $t+\Delta t$, the asset price either moves up by a multiplicative growth factor $u = e^{\sigma\sqrt{\Delta t}}$, or moves down by a multiplicative growth factor $d = e^{-\sigma\sqrt{\Delta t}}$. Assume a constant continuously compounded riskless interest rate r, so the riskless growth factor is $R = e^{r\Delta t}$ over the time step. Let V_t be the time–t price of a European-style derivative that, at time $t+\Delta t$, has the value V_u in the up state and V_d in the down state. Let $q = (R-d)/(u-d)$ denote the fixed CRR risk-neutral probability of an up-move. Then, to avoid arbitrage, the CRR model says that equation (2.1) holds for the value of the European-style derivative over the time step Δt:

$$\begin{aligned} V_t &= \frac{1}{R}E_{RN}\left(V_{t+\Delta t}\right) \\ &= \frac{1}{R}\left[qV_u + (1-q)\,V_d\right] \\ &= \frac{1}{R}\left[\left(\frac{R-d}{u-d}\right)V_u + \left(\frac{u-R}{u-d}\right)V_d\right], \end{aligned} \tag{2.1}$$

where $E_{RN}(\cdot)$ is the risk-neutral probability expectations operator.

Arnold et al. (2009) give both equilibrium and no-arbitrage derivations to show that, if $K = e^{k\Delta t}$ is a risk-adjusted compounding factor over time step Δt, then $p = (K-d)/(u-d)$ is the risk-averse probability of the up state and equation (2.2) holds for the value of the European-style derivative over the time step Δt:

$$\begin{aligned} V_t &= \frac{1}{R}\left[E\left(V_{t+\Delta t}\right) - \left(\frac{V_u - V_d}{u-d}\right)(K-R)\right] \\ &= \frac{1}{R}\left\{\left[pV_u + (1-p)V_d\right] - \left(\frac{V_u - V_d}{u-d}\right)(K-R)\right\} \end{aligned} \tag{2.2}$$

where $E(\cdot)$ is the risk-averse probability expectations operator. That is, given an estimated CRR tree and a risk-adjusted discount rate k (or a risk premium $k - r$), we can deduce the risk-averse probability p for any up-move in the CRR tree. The derivative pricing is identical between equations (2.1) and (2.2), but the probabilities are risk-neutral in (2.1) and risk-averse in (2.2).

Equation (2.2) is, essentially, a certainty equivalent formula. It is derived from the relationship in equation (1), and it forms the basis for our first technical step. It provides a means of deducing risk-averse probabilities to overlay on the price structure of an underlying risk-neutral tree. Our RA-BT tree is, thus, a CRR tree transformed by replacing risk-neutral probabilities with risk-averse probabilities (see Arnold et al. 2009 for more details and Arnold and Crack 2004 for an application).

The second step is to generate a risk premium or risk-adjusted discount rate to feed into each node of the no-arbitrage transformation between risk-neutral and risk-averse trees. The risk premium can be derived from an asset pricing model, such as the CAPM. In the case of a macro asset (e.g., the S&P 500 index portfolio), a representative agent argument frees us of the CAPM assumptions, and the risk-adjusted discount rates that feed into each node of this transformation can be derived using general assumed utility functions for the representative agent (Arnold et al. 2009).

Note that, for any given node, each admissible k produces the same option valuation at that node (admissible k requires $-\sigma\sqrt{\Delta t} < k\Delta t < \sigma\sqrt{\Delta t}$ or, equivalently, $d < K < u$ to avoid negative risk-averse probabilities). That is, the risk-adjusted discount rate k determines the risk-averse probability p and, by construction, k and p offset each other within the pricing equation to leave the option value unchanged. In fact, one must be cautious in interpreting p as the risk-averse probability of an up-move at a node, because any admissible k produces the same option valuation at that node (see Arnold et al. 2009 for more details). This leads us to a clause. The fidelity with which p reproduces the true risk-averse probability of an up-move at a given node of a one-step RA-BT tree depends upon the accuracy in the choice of k as the true risk-adjusted discount rate for the underlying security.

This fidelity clause is both a major strength and a slight weakness of our chapter. It is a major strength because, once supplied with the risk-adjusted discount rate k at any node, we can take that node on an existing risk-neutral tree and transform it into a node on a risk-adjusted tree, while retaining the derivatives pricing at that node. That is, we get existence of a solution to the risk-averse tree driven by existence of a solution to the no-arbitrage-driven risk-neutral tree, and we get it via a no-arbitrage transformation between the two trees at that node. The fidelity clause is a slight weakness because the risk-averse probabilities inferred from the RA-BT (and subsequently from the RA-IBT, as discussed below) are only as good as the discount rate fed into the transformation between the trees.

The third step is to propagate probabilities through the tree. It is almost trivial in the case of the RA-BT. The multi-period RA-BT model follows immediately from the single-period model in equation (2.2) simply by applying the single-period model iteratively backwards through the tree. The values for the underlying asset price in the risk-averse tree are identical to the values for the underlying asset price in the CRR tree. In the special case where $k = r$ at every node or, equivalently, $K = R$ at every node, the risk-averse model reduces to the CRR model. In the limit where step size tends to zero, Black–Scholes-world pricing is obtained.

Assuming a constant risk premium in the multi-period RA-BT places a restriction on the most general form of the RA-BT, where the risk premium varies freely with state and time. A constant risk premium in the RA-BT is consistent with assuming power utility for a representative agent.

2.4 The risk-averse implied binomial tree (RA-IBT)

We need to establish the same three technical steps that we established for the RA-BT in building an RA-IBT. We establish the first two steps by starting with an R-IBT. Rubinstein's R-IBT is a multi-step binomial tree. Each step in the R-IBT is a one-step CRR model (though it need not possess the traditional CRR property that $u = 1/d$). The R-IBT (and thus each of the one-step CRR trees of which it is composed) is estimated using a calibration to market prices of traded European-style options. This calibration allows the parameters of each one-step CRR model to be different at every node within the R-IBT. By doing so, the final distribution of asset prices in an R-IBT need not be lognormal, even when step sizes tend to zero. The R-IBT is well behaved and robust (Chriss 1997: 431). In practice, as long as no arbitrage violations exist among the option prices, then a solution exists for the R-IBT (Rubinstein 1994: 783).

We can both establish that a solution to the risk-averse implied binomial tree (RA-IBT) model exists and demonstrate how the solution is related to the other binomial trees by asking what happens if, at each node within an R-IBT, we transform the one-step CRR model there into a one-step RA-BT model. That is, given risk-adjusted discount rate k at that node and focusing on just one internal node of the tree, we replace risk-neutral probability of an up-move $q = (R-d)/(u-d)$ with risk-averse probability of an up-move $p = (K-d)/(u-d)$, where R and K are as defined earlier. We do not change u or d, or the values of the underlying, or the value of the derivative at this node, just the probabilities. The valuation formula at this node, looking ahead to the next two nodes, then changes from equation (2.1) to equation (2.2). This is the first step we needed to establish.

If we do exactly the same transformation for every internal node in the tree, we create a new tree full of one-step risk-averse tree (RA-BT) models. The new tree provides the same pricing as the R-IBT at each node. That is, the underlying asset and any derivative have the same values at each node on the new tree as they had in the R-IBT we started with – we have a new tree that has risk-averse probabilities of an up-jump at any step, risk-adjusted discount rates for the underlying, and a new pricing formula (equation [2.2]) at each node. This new tree is our RA-IBT. If a solution exists for the R-IBT (and we note above that it does in the absence of arbitrage opportunities between the options), then we can build our new RA-IBT tree using the transformation described above and in detail in Arnold et al. (2009).

The second step is generation of the risk premium at each node to feed into the transformation at the first step. We assume a constant risk premium throughout the tree. This is not a requirement of the model but rather an empirical restriction imposed here for ease of exposition. This restriction is consistent with a representative agent that possesses a power-like utility function (i.e., power utility where the CRRA parameter varies with the state).

This new RA-IBT tree captures volatility smiles and excess skewness and kurtosis. The probability structure in this tree is, however, no longer risk-neutral but risk-averse. It has all the benefits of the R-IBT but without the restriction to risk-neutral probabilities. Of course, where the risk premium is zero, the RA-IBT reduces to an R-IBT.

We have not yet mentioned explicitly how the ending nodal risk-averse probabilities are generated in our RA-IBT. In the original Rubinstein R-IBT, an optimization is performed, and the ending nodal risk-neutral probabilities are the choice variables that are estimated. Rubinstein (1994: 790) supplies a backward recursion that starts at these ending nodal probabilities and works backward through the tree to deduce all the u, d, and q parameters, which typically vary at each node (with constant discount rate r for both the underlying and the option). In our RA-IBT, something quite different is needed. Having already solved for an R-IBT and propagated its ending nodal probabilities backward through the R-IBT, we then apply the transformation described above to arrive at the u, d, and p parameters, which typically vary at each node through the RA-IBT. Recall that the transformation needs a risk-adjusted discount rate k for the underlying at each node either from an assumed utility function or imposed (with utility consequences). The u and d parameters in the RA-IBT tree are unchanged from those in the R-IBT tree. We may then propagate the up-step probabilities p forward through the RA-IBT to obtain nodal probabilities at each node, out to the ending nodes of the RA-IBT. Arnold et al. (2009) demonstrate the propagation algorithm in detail. This completes the derivation of the RA-IBT.

We now discuss another distinction between the R-IBT and the RA-IBT. Rubinstein's R-IBT possesses binomial path independence (BPI). That is, each path leading to any given node arrives there with equal probability (Chriss 1997: 417). The nodal probability at any node in an R-IBT is thus simply the path probability times the number of paths arriving at that node. Our transformation from the risk-neutral R-IBT to the risk-adjusted RA-IBT does not, however, preserve BPI. It is not true that paths through our risk-averse RA-IBT tree have equal path probability. Rubinstein forces his R-IBT tree to have BPI so as to reduce the degrees of freedom enough to be able to solve the problem and arrive at a solution that propagates naturally backwards through his R-IBT tree. The fact that Rubinstein enforces BPI in his tree yields his solution, which guarantees the existence of ours. Our direct transformation of Rubinstein's tree to ours is followed by a forward propagation of the probabilities without ever needing to explicitly work out path probabilities.

In summary, we rely upon three distinct but interrelated technical steps. The first step is the derivation of the functional form of the transformation that binds the risk-neutral and risk-averse trees. At any given node in the tree, this transformation is a function of the risk-adjusted discount rate or risk premium at that node. The second step is the ability to derive this risk-adjusted discount rate at any given node. The third step is our demonstration of how to combine and implement these first two steps and how to propagate risk-adjusted probabilities through the risk-averse binomial tree.

2.5 Empirical analysis

2.5.1 The options data

We use intraday data on S&P 500 index options over the period from January 1993 to September 1995. Twenty out of 33 possible sets are usable after we eliminate data that do not provide an adequate cross-section of prices.

Each month we select 10 call option quotes, with bid and ask prices in excess of $0.50. The options are of different strikes but the same maturity (two months, but varying between 59 and 61 days throughout the period). We select options that are closest to the money and as close as possible to 11:00 A.M. CST. For a given calibration, all of the quotes are usually collected within a quarter of an hour. The index level is sampled as close as possible to 11:00 A.M. CST. The index level is adjusted for dividends (i.e., the discounted value of future dividends during the life of the option is subtracted from the index price based on historic dividend payouts collected from the *S&P 500 Information Bulletin*). Given the short maturity of the options and the stability of the index over this time period, using the actual dividends as a substitute for anticipated dividends appears reasonable. Further, for the same reason, the riskless rate is used to discount the dividends. The effect of using an assumed higher discount rate corresponding to the index for discounting the dividends is negligible. Finally, we screen option quotes for arbitrage violations, using a risk-free rate inferred from closing quote midpoints of US Treasury securities that straddle the option maturity date (collected from the *Wall Street Journal*).

Given the criteria for the data, 20 sets of options are available for empirical analysis. Using a 200-step binomial tree, the RA-IBT model is estimated using imposed risk premiums of 0.0 percent (this is the R-IBT), 3.7 percent, 7.5 percent, and 11.3 percent (i.e., RA-IBTs consistent with power utility with varying CRRA). In total, 20 binomial trees are estimated via optimization using 10 option quotes each (these are the R-IBT trees); an additional 60 RA-IBT trees are derived as transformations of the R-IBTs (three different risk premiums for each R-IBT). A CRR tree for each of the 20 sets of options data is also computed for comparison purposes.

Table 2.1 Model price comparison with bid and ask quotes of 60-day options (January 19, 1993)

Strike price	Bid price	Model price	Ask price
405	32.0000	32.7518	34.0000
410	27.3750	27.9949	29.3750
415	23.0000	23.4068	25.0000
420	19.0625	19.0629	19.8125
425	15.0000	15.0581	16.0000
430	11.2500	11.4714	12.2500
435*	8.1250	8.3688	8.6250
440	5.1250	5.7877	6.1250
445	3.3750	3.7434	3.8750
450	1.7500	2.2497	2.2500

* Indicates the at-the-money option. The risk-free rate is 2.88 percent per annum. The annual implied volatility is 11.2 percent. The objective function is minimized to a value of 0.0000425. These model prices are from a 200-step R-IBT using a lower bound on nodal probabilities of 0.0000005. Identical model prices (to more than 10 decimal places) are obtained from the three RA-IBT models with different risk premiums.

All the risk-averse RA-IBT models in this chapter are estimated by transforming the solution to a risk-neutral Rubinstein R-IBT. The Rubinstein R-IBT is estimated using an optimization routine that minimizes the Jackwerth and Rubinstein (1996) smooth objective function subject to pricing the traded options within the spread. In practice, the pricing is very good. Of the 200 options we price in calibrating the R-IBTs (i.e., 10 options priced in each of 20 periods), only two are not priced within the spread. One is priced at 1/1000th of a penny below the bid and the other at 12/1000ths of a penny below the bid. Otherwise, all other options are priced strictly within the spread. The RA-IBTs, by construction as direct transformations of the R-IBTs, produce identical pricing to the R-IBTs from which they are derived. Table 2.1 displays, as an example, the R-IBT prices versus the bid and ask prices for the first set of options (January 19, 1993). The dividend adjusted level of the S&P 500 around 11:00 a.m. CST was 433.78.

2.5.2 Marginal rate of substitution and implied relative risk-aversion

We now consider the relationships between utility functions and probability densities. There are many reasons for doing this. For example, stochastic volatility type and jump diffusion type option pricing models typically have skewed and kurtotic return distributions for the underlying security. In these models, a utility function is necessary to map from the theoretical risk-averse probability return distribution of the underlying security to the associated risk-neutral return distribution. Generally, the utility function is chosen with convenience in mind and not necessarily due to a particular economic rationale. Empirical work that harvests information about aggregate utility from option prices can help to shape the assumptions made in such models.

To use our estimated density functions for the S&P 500 to make inferences about representative agent utility functions, we assume that some unspecified equilibrium asset pricing model holds, that it applies in a representative agent setting, and that the S&P 500 as a broad market index serves as a proxy for aggregate consumption. For related discussion, see Ait-Sahalia and Lo (2000), who in turn cite Brown and Gibbons (1985) and also discuss the limits of these assumptions; see also Bliss and Panigirtzoglou (2004).

Let $f_{R-IBT}(S_T)$ and $f_{RA-IBT}(S_T)$ denote the implied risk-neutral and risk-averse density functions, respectively, that are inferred from our IBTs for the future level S_T of the S&P 500.[3] We exploit two relationships between utility functions and probability densities. The first relationship is that the ratio of the risk-neutral density to the risk-averse density gives up to a constant that is independent of the index level, the marginal rate of substitution (MRS) of the representative agent between consumption at time T and time t, as shown in equation (2.3) (Ingersoll 1987: 187; Ait-Sahalia and Lo 2000: 27; Jackwerth 1999: 72, 2000: 436, and 2004: 53):

$$f_{R-IBT}(S_T)/f_{RA-IBT}(S_T) \propto U'(S_T)/U'(S_t) = MRS \tag{2.3}$$

The MRS is the "pricing kernel." We will use the terms "MRS" and "pricing kernel" interchangeably.

If the representative agent is risk-neutral, then we expect the MRS to be unity. If the representative agent is risk-averse, we expect the MRS to be downward-sloping as a function of wealth. Jackwerth (2000, 2004) reports several authors finding that the MRS is locally upward-sloping for some wealth levels near the initial wealth– a "pricing kernel puzzle" because it means the representative investor is locally risk-seeking.

We find that the MRS is downward-sloping and well behaved for each level of risk premium fed into the RA-IBT model (see Figure 2.1 for the 7.5 percent risk premium case). The time period for our data sample encompasses that of Ait-Sahalia and Lo (2000). Our results for the MRS are

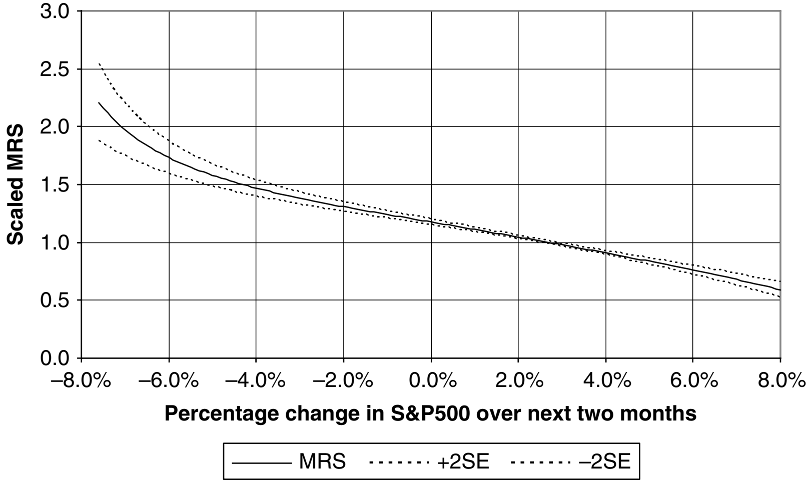


Figure 2.1 Scaled marginal rate of substitution (RP = 7.5%)
Note: For the 7.5 percent risk premium case, and for each of the 20 months' optimizations from 1993 to 1995, we estimate the scaled marginal rate of substitution (MRS) (i.e., the pricing kernel) as the ratio $f_{R-IBT}(S_T)/f_{RA-IBT}(S_T)$. We then average these MRS numbers across the 20 months' estimations by first associating each with the percentage change in the S&P 500 relative to that month's dividend adjusted index level S^*. We calculate the average only where each individual density possesses a contiguous range of values that does not use the optimization's lower bound on the density of 0.0000005. Each tree's ending nodes are different; so, for each return level on the plot, we linearly interpolate between the individual estimates before taking the average. We show the average plus and minus two empirical standard errors of the mean (SE).

quite similar to those in Ait-Sahalia and Lo (2000: 36, figure 3), though their confidence interval is bordering on negative territory at high values of the index, whereas ours is clearly positive everywhere. Our results are, however, quite unlike the oddly shaped, locally increasing pricing kernels in Jackwerth (2004: 57, figure 11). The Jackwerth (2004) data are from a much later time period than ours, and this could partially explain differences in results.

The Ait-Sahalia and Lo (2000) MRS is calculated using nonparametric techniques for the numerator and historical data for the denominator in equation (2.3). The Jackwerth (2004) MRS is calculated using IBTs for the numerator and historical data for the denominator in equation (2.3). We differ from each of these authors in that we are the first to use wholly implied techniques for both numerator and denominator. Our MRS results are much smoother than Ait-Sahalia and Lo's and wholly downward-sloping, unlike Jackwerth (2004).

The second relationship between utility and probability densities we exploit is that we can estimate the implied Arrow–Pratt measure of RRA using equation (2.6) (Ingersoll 1987: 38–39):[4]

$$RRA = S_T \left[f'_{RA-IBT}(S_T)/f_{RA-IBT}(S_T) - f'_{R-IBT}(S_T)/f_{R-IBT}(S_T) \right] \qquad (2.4)$$

Typical empirical estimates of the RRA range from about 0 to 55 (see good summaries of prior findings in Ait-Sahalia and Lo 2000: 39, Table 5; and Jackwerth 2004: 53–54). Jackwerth finds, however, clearly negative values for absolute (and thus also for relative) risk aversion near initial levels of wealth (Jackwerth 2000: 442, figure 3: 442). Negative RRA ties in with locally upward-sloping MRS and forms his pricing kernel puzzle – inconsistent with economic theory.

Our implied RRA numbers are wholly positive across all levels of wealth, and they are of the order of 3–9, 7–18, and 9–27 for the 3.7 percent, 7.5 percent, and 11.3 percent risk premium cases, respectively (middle case only shown). Our implied RRA numbers (Figure 2.2) are similar to Ait-Sahalia and Lo's (2000: 38, figure 4) in sign, size, and behavior across states, but our plots are, again, much smoother than theirs. Like Ait-Sahalia and Lo, we find economically and statistically significant evidence against CRRA, and we find that RRA increases with increasing wealth beyond current levels. Of course, varying CRRA is consistent with our assumption of an imposed constant risk premium. Our results are quite different from the clearly negative results in Jackwerth (2000: 442, figure 3). Unlike our MRS results, the time period that Jackwerth uses to calculate his risk aversion in Panel D of his figure 3 (Jackwerth 2000: 442) overlaps substantially with the time period we use. Therefore, different time periods are not the explanation.

The major difference between our analysis and Jackwerth's is that we use implied trees for the risk-averse density estimation and he uses historical data. Although Jackwerth (2000: 445–446) dismisses his use of historical data to estimate the risk-averse density as a cause of the puzzle, we suspect that this may be, at least in part, responsible for his economically unintuitive results.

Ait-Sahalia and Lo (1998) criticize IBTs as possessing inherently nonstationary estimates relative to nonparametric techniques, but we certainly do not see that in the results we present. In particular, in comparing our

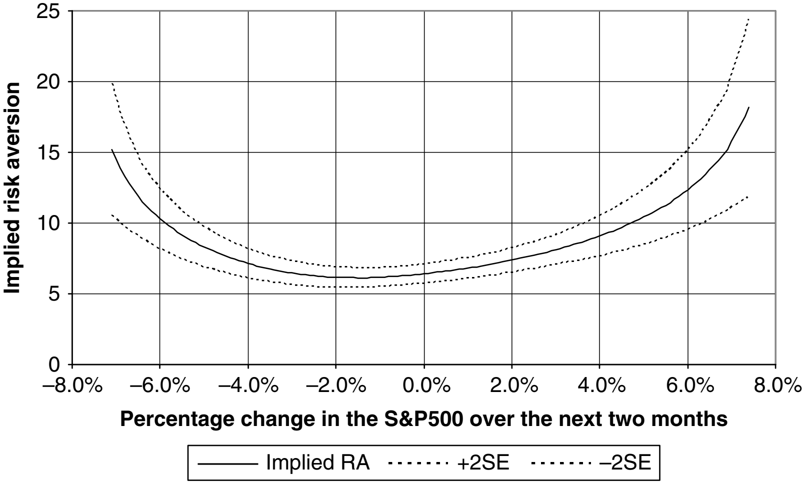


Figure 2.2 Implied Arrow–Pratt relative risk aversion (RP = 7.5%)
Note: For the 7.5 percent risk premium case and for each of the 20 months' optimizations from 1993 to 1995, we estimate the implied Arrow–Pratt RRA for each two-month-ahead level S_T of the S&P 500 as $S_T[f'_{RA-IBT}(S_T)/f_{RA-IBT}(S_T) - f'_{R-IBT}(S_T)/f_{R-IBT}(S_T)]$. We average these estimates over the 20 months' estimations by first putting them on an equal footing by associating each with the percentage change in the S&P 500 relative to that month's dividend adjusted index level S^*. We calculate the mean only where each month's density (and its slope estimate) possesses a contiguous range of values that does not use the optimization's lower bound on the density (0.0000005). Each tree's ending nodes are different; so, for each return level on the plot, we linearly interpolate between the individual estimates before taking the average. We show the average plus and minus two empirical standard errors of the mean.

MRS and RRA estimations with those in Ait-Sahalia and Lo (2000), we note that our plots are much smoother. This difference in smoothness may be because their nonparametric kernel estimations use a bandwidth that is too small, though Jackwerth suggests that they may in fact have oversmoothed their results (Jackwerth 2004: 54). Alternatively, the difference in smoothness may be because Ait-Sahalia and Lo estimate their plots as a snapshot over one year, whereas we re-estimate our trees 20 times over three years and then average the results. Our averaging may produce smoother results than their nonparametric technique, but we would not expect this if our IBTs were inherently nonstationary, as suggested by Ait-Sahalia and Lo.

Bliss and Panigirtzoglou (2004) approach the stationarity issue from yet another angle. Rather than using implied binomial trees, they use Breeden and Litzenberger's well-known result (1978) to deduce risk-neutral densities from option prices. They then assume a utility function (either power or exponential) and use it to obtain a risk-adjusted density function. By fixing the utility function and allowing the density functions to be time-varying, they avoid having to assume that the density functions are stationary. The aim of their chapter is to calibrate the parameters of the utility function so as not to be able to reject the risk-adjusted implied densities as forecasts of subsequently realized returns distributions. They find RRA estimates that decrease with increasing horizon and little evidence of pricing kernel anomalies.

The bottom line is that we find no pricing kernel puzzle using wholly implied techniques. Therefore, the representative agent is risk-averse with no local risk-seeking behavior. We do find significant variation in RRA across states that is inconsistent with an assumption of CRRA. We also find that risk aversion increases with increasing wealth beyond current wealth.

2.6 Conclusion

Our risk-averse implied binomial tree (RA-IBT) model generalizes Rubinstein's risk-neutral implied binomial tree (R-IBT) model by allowing for a nonzero risk premium on the underlying asset. The RA-IBT accommodates a risk premium that is time-varying and/or state-dependent, depending upon either the assumed utility function of the representative agent (in the case of a macro asset such as the S&P 500 index portfolio) or a CAPM beta (in the case of an individual stock). We have imposed a constant risk premium in our empirical work with S&P 500 index options (consistent with assumed power utility with varying CRRA for a representative agent). Our S&P 500 index options data run between 1993 and 1995. We estimate the pricing kernel (marginal rate of substitution) and implied RRA and compare and contrast our results with other researchers' results. In particular, we are the first to use implied techniques for both the risk-neutral and risk-averse densities, and we find no "pricing kernel puzzle" using these techniques (compared with other authors who use historical data to generate the risk-averse density).

Acknowledgments

We thank David Alexander, Ravi Bansal, Alex Butler, Scott Chaput, Robin Grieves, Robert Hauswald, Jimmy Hilliard, Stewart Mayhew, Susan Monaco, Sanjay Nawalkha, Mark Rubinstein, Louis Scott, Richard Shockley, seminar participants at the University of Georgia and University of Otago, and an anonymous referee. Any errors are ours.

Notes

1. In Arnold, Crack and Schwartz (2009) we show how to relax the constant risk premium assumption and generate a risk premium at any node using an assumed utility function of a representative agent, and we provide examples for power utility and negative exponential utility.
2. Cox and Rubinstein (1985: 324) discuss a related problem with a discount rate that is correct on average.
3. Strictly speaking, our discrete trees yield discrete probability masses associated with ending discrete nodal values of the index. For our 200-step trees, we associate the probability mass with the width of a range of index values about the node, and we deduce the density f as the constant value of mass/width over that range about that node.
4. In fact, we use $RRA = (1+\rho)[f'_{RA-IBT}(\rho)/f_{RA-IBT}(\rho) - f'_{R-IBT}(\rho)/f_{R-IBT}(\rho)]$, where $\rho = S_T/S^* - 1$ is the simple net return, so that we can put each of the 20 months' estimations on an equal footing and average across them to get the mean RRA and its empirical standard error. This form of the RRA is mathematically identical to equation (2.4)—although we have never seen it published.

References

Ait-Sahalia, Y. and Lo, A. W. (1998) "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices," *Journal of Finance*, 53 (2): 499–547.

Ait-Sahalia, Y. and Lo, A. W. (2000) "Nonparametric Risk Management and Implied Risk Aversion," *Journal of Econometrics*, 94 (1/2): 9–51.

Alonso, F., Blanco R., and Rubio, G. (2009) "Option-Implied Preferences Adjustments, Density Forecasts, and the Equity Risk Premium," *Spanish Economic Review*, 11 (2): 141–164.

Arnold, T. and Crack, T. F. (2004) "Using the WACC to Value Real Options," *Financial Analysts Journal*, 60 (6): 78–82.

Arnold, T., Crack T. F., and Schwartz, A. (2009) "Inferring Risk-Averse Probability Distributions from Options Prices Using Implied Binomial Trees: Additional Theory, Empirics, and Extensions," Working paper. Available online at SSRN: http://ssrn.com/abstract=749904 (accessed June 2, 2009).

Black, F. and Scholes, M. (1973) "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81 (3): 637–659.

Blackburn, D. W. (2006) "Option Implied Risk Aversion and Elasticity of Intertemporal Substitution," Working paper. Available online at SSRN: http://ssrn.com/abstract=927440 (accessed June 2, 2009).

Bliss, R. R. and Panigirtzoglou, N. (2004) "Option-Implied Risk Aversion Estimates," *Journal of Finance*, 59 (1): 407–446.

Breeden, D. T. and Litzenberger, R. H. (1978) "Prices of State-Contingent Claims Implicit in Options Prices," *Journal of Business*, 51 (4): 621–651.
Brown, D. P. and Gibbons, M. R. (1985) "A Simple Econometric Approach for Utility-Based Asset Pricing Models," *Journal of Finance*, 40 (2): 359–381.
Chriss, N. A. (1997) *Black–Scholes and Beyond: Option Pricing Models,* New York: McGraw-Hill.
Cox, J. C., Ross, S., and Rubinstein, M. (1979) "Option Pricing: A Simplified Approach," *Journal of Financial Economics*, 7 (3): 229–263.
Cox, J. C. and Rubinstein, M. (1985) *Options Markets,* Englewood Cliffs, NJ: Prentice-Hall, Inc.
Ingersoll, J. E. (1987) *Theory of Financial Decision Making*, Savage, Md.: Rowman & Littlefield.
Jackwerth, J. C. (1999) "Option-Implied Risk-Neutral Distributions and Implied Binomial Trees: A Literature Review," *Journal of Derivatives*, 7 (2): 66–82.
Jackwerth, J. C. (2000) "Recovering Risk Aversion from Option Prices and Realized Returns," *Review of Financial Studies*, 13 (2): 433–451.
Jackwerth, J. C. (2004) *Option-Implied Risk-Neutral Distributions and Risk Aversion*, Charlottesville, VA: AIMR Research Foundation Monograph (CFA Institute).
Jackwerth J. C. and Rubinstein, M. (1996) "Recovering Probability Distributions from Option Prices," *Journal of Finance*, 51 (5): 1611–1631.
Leland, H. E. (1980) "Who Should Buy Portfolio Insurance?" *Journal of Finance*, 35 (2): 581–594.
Rubinstein, M. (1994) "Implied Binomial Trees," *Journal of Finance*, 49 (3): 771–818.
Stutzer, M. (1996) "A Simple Nonparametric Approach to Derivative Security Valuation," *Journal of Finance*, 51 (5): 1633–1652.
Ziegler, A. (2007) "Why Does Implied Risk Aversion Smile?" *Review of Financial Studies*, 20 (3): 859–904.

3
Pricing Toxic Assets

Carolyn V. Currie

3.1 Introduction: The importance of pricing toxic assets

The term "toxic asset" is a nontechnical term used to describe certain financial assets whose value has fallen significantly so that there is no longer a functioning market for these assets. That is, there is no liquidity in the market, and the market cannot clear. This term became common during the financial crisis that began in August 2007 but predated the global financial crisis, as it was used in 2006 by Angelo Mozilo, founder of Countrywide Financial, who used the term "toxic" to describe certain mortgage products in emails in spring of 2006, as revealed in SEC filings:[1]

> "[The 100% loan-to-value subprime loan is] the most dangerous product in existence and there can be nothing more toxic." (March 28, 2006)

The majority of these assets were connected with residential mortgages which had been securitized, that is, bundled into groups of assets and then onsold. The buyer could then borrow on the basis that these assets had been assigned high credit ratings. Often they were onsold into levered special-purpose vehicles, which had a mixture of equity, senior, and junior debt, with subordinated debt. This subordinated debt was then onsold into another special-purpose vehicle, which was similarly structured. All derived their top credit ratings from the underlying parcel of mortgages. As a side effect, bets on credit ratings became rampant. These credit derivatives were largely written by only three players – principally a state-supervised insurance company, AIG. The value of these assets was very sensitive to economic factors such as housing prices, default rates, and financial-market liquidity. At the slightest faltering of economic growth, the value of these assets started to deteriorate.

The pricing problem with these assets is related to how to comply with accounting standards demanding mark to market, which could have resulted in huge write-downs and technical insolvency for some bank and nonbank financial institutions. The term "zombie bank" was introduced to describe banks that would have become bankrupt if their assets had been revaluated at realistic levels.[2] This resulted in a credit crunch or in excessively speculative lending to compensate for past risk-taking. The net result was a failure in the pricing mechanism, with buyers and sellers unwilling to transact.

On March 23, 2009, US Treasury Secretary Timothy Geithner announced a public-private investment partnership (PPIP) to buy toxic assets from banks' balance sheets. The government hype was,

> An attractive feature of the program is that it will allow the marketplace to establish values for the assets – based, of course, on the auction mechanism that will signal what someone is willing to pay for them – and thus might ease the virtual paralysis that has surrounded those assets up to now. For a relatively small equity exposure, the private investor thus stands to make a considerable return if prices recover. The government will make a gain as well.[3]

The PPIP has two primary programs. The Legacy Loans Program will attempt to buy residential loans from banks' balance sheets. The FDIC will provide nonrecourse loan guarantees for up to 85 percent of the purchase price of legacy loans. Private-sector asset managers and the US Treasury will provide the remaining assets. The second program is called the Legacy Securities Program, which will buy residential mortgage-backed securities (RMBS) that were originally rated AAA, and commercial mortgage-backed securities (CMBS) and asset-backed securities (ABS), which are rated AAA. The funds will come in many instances in equal parts from the US Treasury's Troubled Asset Relief Program monies, from private investors, and from loans from the Federal Reserve's Term Asset Lending Facility (TALF). The initial size of the PPIP is projected to be $500 billion.[4] Banking analyst Meridith Whitney argues that banks will not sell bad assets at fair market values because they are reluctant to take asset write-downs.[5] Removing toxic assets would also reduce the upward volatility of banks' stock prices. Because stock is a call option on a firm's assets, this lost volatility will hurt the stock price of distressed banks. Therefore, such banks will only sell toxic assets at above-market prices.[6] Hence, the pricing issue is critical. Will the US Government set up a clearing house? Or will it design some type of open outcry, or managed open outcry?[7] The pricing methodology will be the biggest factor in whether the credit system recovers.[8]

3.2 Hedging the prices of agricultural and mining products[9]

Agricultural protection plans are commonly used by banks to hedge for price risk using commodity broking services strategies. The different type of plans include commodity swaps, minimum priced contracts, currency-protected commodity swaps, physical and forward sales, collars, basis, limited liability commodity swaps, and commodity protected commodity swaps.

A commodity swap plan is a derivative product formed by converting the global price of the commodity into local weights and measures. In the case of the commodity swap plan, contracts are mostly cash-settled. A product can be easily bought and sold without having a commitment to provide physical settlement. A commodity swap plan requires an understanding in initial and variation margins. Buyers get the right but not the obligation to buy or sell an underlying commodity. A minimum priced contract plan is designed as a minimum floor contract wherein clients can design the floor.

Commodity price instability has a negative impact on economic growth, income distribution, and the poor. Early attempts to deal with commodity price volatility relying on direct government intervention (for example, price stabilization schemes, floor prices, and guaranteed prices) were generally unsuccessful. Although there may be a case for limited direct intervention in some circumstances, liberalization of markets has resulted in the need for market-based instruments to help manage commodity price volatility. Large commodity exchanges typically offer such products (for example, futures and options), but there are substantial barriers to developing these markets for all commodities and for helping farmers to access existing markets. Key investments needed to expand access to these services include public goods (price information systems, data management systems), strengthening supply chain relationships, strengthening technical capacity in private service providers, and educating potential users.

3.2.1 Managing commodity price risk

Governments in many countries have intervened in markets, often through state economic enterprises, to insulate producers and consumers from world prices. Most interventions have taken a nonmarket approach in the form of quota or buffer stock programs organized through state marketing boards. However, government interventions have been costly and have crowded out private-sector initiatives.

Price-risk management that relies on market-based products rather than government guarantees and subsidies will involve a substantially reduced overall role for government in administration. The long-term objective must be for government to assume a regulatory role, overseeing markets for risk management tools. However, the public sector can facilitate initial development of these markets and/or improve access to established foreign or international markets for these tools, thus ensuring that needs of the poor are adequately addressed. Market-based systems are most relevant for standardized commodities traded internationally in large volumes, mainly coffee, cocoa, rubber, cotton, grains, sugar, and oilseeds (and some livestock products). They are less applicable to high-value, highly differentiated, or perishable products for which price risk is managed through forward contracts, often in the context of integrated supply chains.

The rationale and theoretical underpinnings of formal mechanisms for managing price risk are reasonably simple. There are two basic types of price-risk management tools: physical instruments and financial instruments:

- *Physical instruments* involve strategic pricing and timing of physical purchases and sales (such as "back-to-back" trading), forward contracts, minimum price forward contracts, price-to-be fixed contracts, and long-term contracts with fixed or floating prices.
- *Financial instruments* are exchange-traded futures and options, over the counter (OTC) options and swaps, commodity-linked bonds, and other commodity derivatives.
- *Futures contracts* involve the buyer (or seller) of a futures contract agreeing to purchase (or sell) a specified amount of a commodity at a specified price on a specified date. Contract terms (for example, amounts, grades, delivery dates) are standardized, and transactions are handled only by organized exchanges. Profits and losses in trades are settled daily through margin funds deposited in the exchange as collateral. Futures contracts are usually settled before or at maturity, and they do not generally involve physical delivery of the product.
- *Options contracts offer the right* – but not the *obligation* – to purchase or sell a specified quantity of an underlying futures contract at a predetermined price on or before a given date. Like futures contracts, exchange-traded options are standardized, OTC options offered by banks and commodity brokers. Purchase of an option is equivalent to price insurance and therefore requires that a price (premium) be

paid. Options include calls (which give the buyer the right to buy the underlying futures contract during a given period and are purchased as insurance against price increases) and puts (which give the buyer the right to sell the underlying futures contract during a given period and are purchased as an insurance against price declines).

3.2.2 Benefits

Market-based price risk management instruments have the potential to provide producers with more certainty about the minimum price they will receive for their crop (at the cost of higher revenues forgone), and they may help producers make more efficient farm-management decisions regarding output mix and input use. The elimination of worst-price scenarios can provide incentives for investment in promising sectors (which are often high risk/high return). Reducing market distortions fosters diversification to new and more profitable agricultural enterprises. Further, eliminating the primary reason for nonrepayment of loans – an adverse move in commodity prices – can reduce the risk exposure of producers or market intermediaries in the eyes of lenders and is likely to result in improved access to (and terms of) credit for the sector as a whole.

3.2.3 Policy and implementation issues

Targeting use of nonmarket mechanisms. Reforming existing nonmarket interventions (such as price bands and floors) so that they are minimally distorting will enable the development of market-based mechanisms that "price stabilization" has tended to impede. Key to success of such nonmarket schemes is the ability to accurately define the threshold price, maintain discipline in implementation and include specific sunset clauses. Such schemes are appropriate only when major barriers to market-based alternatives will persist into the medium term and where there is a true underlying competitive advantage for the commodity selected for the price floor scheme.

Commodity exchanges. Well-functioning commodity exchanges – systems of price discovery – improve marketing efficiency for agricultural products and open up new production and marketing opportunities to producers. They reduce price risk (faced by both producers and buyers) by improving overall market liquidity, enhancing stability of local trading networks and providing farmers with more certainty (through better information) of expected future prices (upon which they can make better managerial decisions).

However, out of all these schemes, by far the most popular is floor-plan pricing.

3.3 Floor-plan pricing

Floor-plan pricing mechanisms are hedging tools designed to assist in mitigating the risk associated specifically with commodity price fluctuations. This allows the farmer, subject to approval from the provider to fix a price, to select a price range or to set a price floor/cap, up to three years in advance for certain commodities. This may assist in planning and budgeting with greater accuracy, achieving better control margins, and reducing the time associated with monitoring price movements on overseas commodity exchanges.

Key features include (worked for an Australian farmer):

1. a hedge limit based on a percentage of underlying production, trade or procurement exposure;
2. the hedge can be established in either Australian dollars or other currency denominations. An Australian-dollar-denominated product will eliminate the need to establish a foreign-exchange hedge as well (refer table below);
3. a maximum hedge term of up to three years (refer table below);
4. firm intraday pricing in the local time zone;
5. *no* exchange-traded brokerage fees or daily margin calls;
6. *no* requirement to physically deliver your commodity;
7. up-to-date market intelligence sourced both domestically and from abroad.

Strategy reports are provided on a regular basis to assist in making the most effective hedging decisions relevant to your long-term business objectives.

3.3.1 Agribusiness price risk management solutions

Contract Specifications

Commodity	**Maximum Term (Years)**	**Minimum Quantity**	**Unit of Measure**	**Choice of Currency**	**Pricing Reference (Futures / Exchange)**	**Hedging Strategies**
Canola	3	50	metric tonnes	AUD or CAD	Winnipeg Commodity Exchange (WCE)	Swap, Floor, Cap, Collar, Forward

Corn	3	100	metric tonnes	AUD or USD	Chicago Board of Trade (CBOT)	Swap, Floor, Cap, Collar, Forward
Cotton	3	100	bales	AUD or USD	New York Cotton Exchange (NYCE)	Swap, Floor, Cap, Collar, Forward
Sugar	3	100	metric tonnes	AUD or USD	Coffee Sugar & Cocoa Exchange (CSCE)	Swap, Floor, Cap, Collar, Forward
Wheat	3	100	metric tonnes	AUD or USD	Chicago Board of Trade (CBOT) or Kansas City Board of Trade (KCBOT)	Swap, Floor, Cap, Collar, Forward
Wheat	2	100	metric tonnes	AUD	ASX Milling Wheat	Swap

3.4 An example of floor-plan pricing using cattle[10]

An example will help illustrate the total process of hedging risk using floor plans. Suppose a cattleman who wants to sell a load of feeder cattle in early October checks the options quotes in June and finds he could purchase an October feeder cattle option to sell (a put) at $60/cwt for $2.75/cwt. To further localize this strike price, he adds a $1.00/cwt (basis) since he normally sells 600-pound steer calves slightly higher in October than the October futures price. Commission and premium interest cost will be about $0.25/cwt so the $60 put would provide an expected minimum selling price of $60 + $1.00 − $2.75 − $0.25 or $58/cwt. By comparing this with his other pricing alternatives and his production cost, he decides that the purchase of this put would be an appropriate strategy for the 83 steers he plans to sell in October. He calls his broker and advises him that he wants to purchase one "$60 October feeder cattle put at $2.75." He then forwards a check for $1,450 (500 cwt × $2.75/cwt plus $75 brokerage fee) to his broker.

As October approaches, one of these three things will happen: prices stay the same; prices rise above the option strike price; or prices fall, making the option valuable.

Alternatives are described below in Tables and Figures:[11]

Table 1. Feeder Cattle Price Decline Example

Cash Market	Feeder Cattle Option Market
June 1	
Expect to sell 83 hd in early October, Expected basis =+1.00, So Expect minimum selling price of $58.00 (Strike price - premium & trade cost + basis)	Buy an October Feeder Cattle put option at a $60 strike price for $2.75 per cwt. Premium, trading cost $.25/cwt.
October 10	
Sell 83 hd. feeder steers locally @ $56.00/cwt	October feeder cattle futures trading at $55. Sell $60 October put and collect $5 premium.
Results	
	Offset premium received - original premium & trading cost paid = $5 - $2.75 - $.25 = $2.00
Cash price + gain or loss in options market = actual price received OR $56 + $2 = $58/cwt.	

Figure 1. Possible outcomes when a $60 October put is purchased, +$1.00/cwt. basis.

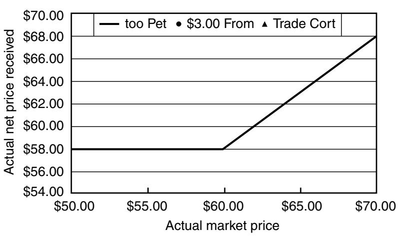
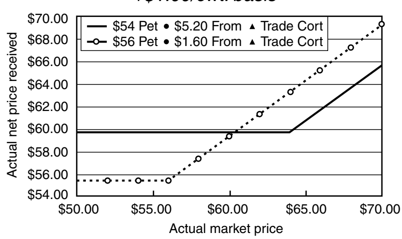


Figure 2. Possible outcomes from a $64 and $56 October feeder cattle put purchase, +$1.00/cwt. basis



However, consider if the cattleman had sold a call as well at an upper price: he would have then collared his price instability and provided for future uncertainty. This method is used to hedge gold price uncertainty associated with the time lags involved in mining gold in Australia.

3.5 Pricing of toxic assets

In this section, a suggested pricing model is expounded that combines the above elements of floor-plan pricing with Basel II methods of allocating capital to a risky portfolio. All that needs to change in the model below is the default probabilities.

Given that the amount of capital allocated to a product, transaction, business unit, or portfolio influences the ultimate profitability, in terms of risk-adjusted return on capital, and given that capital is the most expensive source of funds, capital allocation can be deemed to be a principal driving force in a financial institution. This has been the obstacle in the pricing of toxic assets – if such assets have to be written down against capital then the pricing of risk premium has to be adjusted upwards to compensate for the amount of capital allocated. By combining the method of determining capital to be allocated with a floor-plan pricing mechanism of puts and calls, it should be possible to provide some certainty to the market and effectively to unfreeze it.

We need to understand the concept of economic capital and the use of risk contributions – the risk retained by a facility, or a sub-portfolio, post-diversification – as the foundation of the capital allocation system. Risk contributions can be absolute or marginal, the latter being the changes in risk with and without an additional unit of exposure, a facility or a sub-portfolio of facilities. Whereas absolute risk contributions are allocations of the portfolio risk to the existing individual facilities or sub-portfolios, being embedded in the correlation structure of the portfolio, marginal risk contributions serve essentially for risk-based pricing with an ex-ante view of risk decisions.

Because we need a system of monitoring and pricing risk and return on the basis of risk adjustments (that makes performance comparison across transactions or business units consistent), buyers and sellers of toxic assets should consider systems of ex-post risk-adjusted performance measurement (RAPM) or ex-ante risk-based pricing (RBP) as a method of pricing in line with risk and with the overall profitability goal of a financial institution. In the first case, income is given, whereas the purpose of RBP is to define what is its minimum level. The risk adjustments are the risk contributions, which do not depend on the source of the risk. The same calculations apply to market, credit, and operating risk contributions and performance. The relationship of value at risk (VaR) and risk-based capital is an essential consideration, as the VaR methodology serves to define risk-based capital, or the capital required to absorb potential unexpected losses at a preset confidence level, reflecting the risk appetite of the bank. By definition, it is also the probability that the loss exceeds the capital, triggering bank insolvency. That is, the confidence level is equal to the default probability of the bank.

So what is economic capital? It is the amount of capital that is needed to protect an organization with a chosen level of certainty against insolvency due to unexpected losses over a given time period of, say, one year. Hence, operational economic capital protects the company against insolvency due to unexpected operational losses. To determine the amount of economic capital the firm must decide upon the level of certainty with which it wishes to protect itself against insolvency – the higher the chosen level of certainty, the greater the amount of economic capital required as well as the longer the period, the greater the amount of economic capital needed. Economic capital is, hence, a number that summarizes the current (market, credit, operational or overall) risk profile of the company in a single figure. This figure serves as a measure for understanding the absolute size of risk as well as the change in risk over time, as well as enabling the comparison of risk across different business units. Finally, it is the basis for assessing whether a sufficient return has been earned given the size of the risk being taken.

3.5.1 Calculation of risk contributions[12]

The mechanisms for calculating risk contributions are similar for market risk, credit risk, and operational risk. They require loss distributions but do not depend on the source of the risk. Capital allocations are the absolute risk contributions and sum to total overall portfolio risk, while marginal risk contributions do not. The importance of capital allocations are their input to RAPM on an ex-post basis, while marginal risk contributions are relevant for pricing purposes, i.e. to new transactions.

Capital allocation aims at assigning all types of risk to the business units that generate them, providing top-down and bottom-up links. We can disaggregate and aggregate risk contributions according to any criteria, as long as individual transaction risk contributions are available. For instance, if several business units deal with one client, we can allocate the risk contributions into subsets relative to each business unit according to the type of risk. We can define **standalone risk of a facility**, which is the loss volatility (LV) of a single facility. **Marginal risk** is the change in the portfolio LV when adding a new facility to the existing portfolio. The **absolute risk contribution** is the covariance of the random loss of this single facility with the entire portfolio. *These definitions can be applied to the people, processing, and external events that are or may be connected with facilities and portfolios*.

Of the three definitions above, the two first are intuitive, and the third is mathematical. The absolute risk contribution captures the risk of a facility given that other facilities diversify away a fraction of its stand-alone risk. Risk contributions also depend on the overall measure of portfolio risk to which they contribute. The simplest risk contributions are the contributions to the LV of the portfolio loss. These risk contributions have the attractive property of adding up to the LV of the portfolio.

Since it is common to express capital as a multiple of this portfolio LV, risk contributions are converted into capital allocations through the same scaling factor. Since risk contributions sum to the LV, capital allocations also sum to the portfolio capital. Because they add up to the portfolio risk measure post-diversification effects, the absolute risk contributions are a convenient basis for the overall portfolio capital allocation to individual facilities. The concept of risk contribution is intuitive, but calculations use technical formulas requiring us to specify the notation. Risk contributions always refer to a facility of obligor *i*and a reference portfolio *P*. There are several risk contributions defined below. The portfolio *P* is made up of *N*facilities. Each facility *i*relates to a single obligor. The notation applies to both default models and full valuation mode models. However, all examples use calculations in default mode only for simplicity.

3.5.2 Risk contribution definitions

The stand-alone risk is the LV of a single facility. The *absolute risk contribution* to the LV of the portfolio is the contribution of an obligor *i*to the overall LV. Absolute risk contributions to the portfolio LV differ from the risk contributions to the portfolio capital, which are the capital allocations. The capital allocations relate to the risk contributions through the portfolio LV. To convert absolute risk contributions to portfolio LV into absolute risk contributions to capital, we multiply them by the ratio $m(\alpha)$ of capital to portfolio LV.

The *marginal risk contribution* to LV is the change in portfolio LV when adding an additional unit of exposure, a new facility, a new obligor, or a new portfolio. For instance, the marginal risk contribution of an obligor *f* is the variation of the LV with and without the obligor *f* (or a subset *a*of obligors).

3.5.3 Notation to measure risk contributions[13]

The marginal risk contribution to the portfolio LV and the marginal contribution to capital differ. The first is the variation of the portfolio LV when adding a facility or obligor, or any subset of facilities. The marginal contribution to capital is the corresponding variation of capital. These distinct marginal contributions do not relate in a simple way. The capital is $K(\alpha)$, with confidence level c The two marginal risk contributions are:

$$MRC_f^{P+f}(LV^P) = LV^{P+f} - LV^P$$

$$MRC_f^{P+f}[K(\alpha)] = K(\alpha)^{P+f} - K(\alpha)^P$$

Unless otherwise specified, we use marginal risk contribution as the marginal change in LV of the portfolio. If the multiple $m(\alpha)$ does not change significantly when the portfolio changes, the marginal contribution to LV times the overall ratio of capital to portfolio LV is a proxy of marginal risk contribution to capital. This approximation is not valid whenever the portfolio changes significantly.

3.5.4 Basic properties of risk contributions

The absolute risk contributions serve to allocate capital. The absolute risk contribution to the portfolio LV of a facility *i*to a portfolio *P*is the covariance of the random loss of this single facility *i*with the aggregated random portfolio loss over the entire portfolio (including *i*), divided by the LV of this aggregated random loss. *The formula for calculating absolute risk contributions results from that of the variance of the portfolio. Absolute risk contributions to LV, times the multiple of overall capital to overall portfolio LV, sum exactly to the portfolio capital. This is the key property making them the foundation for the capital allocation system solving the nonintuitive issue of allocating risks.*

Marginal risk contributions serve to make incremental decisions for risk-based pricing. They provide a direct answer to questions such as: What is the additional capital consumed by an additional facility? What is the capital saved by withdrawing a facility or a sub-portfolio from the current portfolio? Marginal contributions serve for pricing purposes.

- *Pricing in such a way that the revenues of an additional facility equal the target hurdle rate of return times the marginal risk contribution of the new facility ensures that the target return of the portfolio on capital remains equal to or above the minimum hurdle rate.*
- *Marginal risk contributions to the portfolio LV are lower than absolute risk contributions to the portfolio LV. However, marginal risk contributions to the portfolio capital can be higher or lower than absolute contributions to the portfolio capital.*

The properties of absolute and marginal risk contributions serve to address different issues. The key distinction is ex-post versus ex-ante applications. Absolute risk contributions serve for ex-post allocations of capital based on effective usage of line, while marginal risk contributions serve to make ex-ante risk-based pricing decisions.

3.5.5 Absolute and marginal risk contributions and their key properties

	To portfolio loss volatility LV^p or capital K^p
Absolute risk contribution ARC (ex post view)	• Sum up to LV^p • Capital allocation (ex post) • Risk-based performance (ex post)
Marginal risk contribution MRC (ex ante view)	Do not sum up to LV^p or K^p • Risk-based pricing (ex ante)

A simple example illustrates these properties, avoiding using complex maths. However, we need to explain the concept of **undiversifiable risk**.

With an existing facility, the absolute risk contribution is proportional to the stand-alone risk. The ratio of the risk contribution of a given facility to the facility LV is lower than 1. It measures the diversification effect at the level of a facility. The ratio represents the "retained risk," or the risk retained within the portfolio as a percentage of the stand-alone risk of a facility. This ratio, RR1, measures the undiversifiable risk of the facility by the portfolio:

$$RR_i = \text{undiversifiable risk/standalone risk} = ACR_i/\sigma_i$$

Risk contributions and retained risk have several attractive properties that we demonstrate below. The RR of an individual facility is simply identical to its correlation coefficient with the entire portfolio. Since all correlation coefficients are lower than or equal to 1, this demonstrates the general result that the risk contribution is always lower than, or equal to, the stand-alone risk. This is intuitively obvious since risk contributions are post-diversification measures of risk.

The facility RR depends on the entire correlation structure with the portfolio. The higher the RR ratio, the higher the undiversifiable risk of the facility. It is important to track the retained risks to identify facilities contributing more to correlation risk. Conversely, for diversification purposes, using low RR guides the choice toward transactions that increase the diversification of the existing portfolio.

We will show in a practical example that the summation of all absolute risk contributions is the aggregated LV of the portfolio. Using the retained risk, this LV is:

$$\sigma p = \sum_i RR_i \times \sigma_i$$

If all cross-correlations are zero, the absolute risk contribution reduces to the variance of the LV squared of the facility over the variance of the portfolio. For a portfolio, the diversification measure is simply the ratio of the aggregated LV of the portfolio to the summation of all individual facility loss volatilities or stand-alone risk.

We now use a simple example to calculate various loss statistics including risk contributions. The example uses a pure default model, building on the example of the two-obligor portfolio with 10 percent default correlation. We do not replicate all detailed calculations. Rather, we detail the comparison of stand-alone and portfolio risk measures.

The examples below provide the details of exposures and the loss distribution.

Standalone default probabilities and default correlations

	Default Probability	**Exposures X_A and X_B**
A	7.00%	100
B	5.00%	50
ρ_{AB}	10.00%	

Loss distribution (default correlation 10%)

	Loss	**Probabilities**	**Cumulated Probabilities**	**Confidence level**
A & B default	150	0.906%	100.000%	$\geq$ 0.906%
A defaults	100	6.904%	99.094%	$\geq$ 7.000%
B defaults	50	4.094%	93.000%	$\geq$ 11.094%
Neither defaults	0	88.906%	88.906%	Not significant

The cumulated loss probabilities provide the loss percentiles. For instance, the loss at the 7 percent confidence level is 50, and the loss at the 0.906 percent confidence level is 100. When we use the second percentile, we consider as a rough proxy of the loss at the 1 percent confidence level. For confidence levels lower than or equal to 0.906 percent, the loss is maximum, or 150. Between 7 percent and less than 0.906 percent, the loss is 100. Between 11.094 percent and less than 7.00 percent, the loss is 50.

3.5.6 Stand-alone expected losses and portfolio expected losses

The expected loss of obligor i is the default risk times the loss given default in value. The expected loss for the portfolio of obligors is the sum of individual obligor expected losses. The expected losses of A and B are the default probabilities times the exposure, or 100×7 percent $= 7$ and 50×5 percent $= 2.5$ respectively for A and B. The portfolio expected loss also results directly from the portfolio loss distribution considering all four possible events, with single default probabilities lower than stand-alone default probabilities. The probability-weighted average of the four loss values is also 9.5.

Stand-alone Loss Volatility and Portfolio Loss Volatility

The stand-alone LV of obligor i in value is: $LV_i = \sigma_i = X_i \times \sqrt{d_i \times (1 - d_i)}$. The convention is that the exposure X is identical to the loss given default or an exposure with a zero recovery rate. The loss volatilities of A and B are $\text{LVA} = 100 \times \sqrt{7\% \times (1 - 7\%)} = 25.515$ and $\text{LVB} = 50 \times \sqrt{5\% \times (1 - 5\%)} =$ 10.897. The unit exposure volatilities are $= \sqrt{\text{P (A) } [1 - P(A)]} = 25.515\%$ for A and $= \sqrt{\text{P (B) } [1 - P(B)]} = 21.794\%$ for B.

Unit exposure volatilities and loss volatilities

Facility	Exposures	Default Probability	Unit exposure volatility	Exposure weighted loss volatility
A	100	7%	25.515%	25.515
B	50	5%	21.794%	10.894

The direct calculation of loss statistics is replicated below. The LV is the square root of the portfolio loss variance, or 28.73.

Loss distribution statistics

EL	9.50
Loss volatility	28.73
Loss variance	825.36

3.5.7 Portfolio capital

Capital derives from the loss distributions and the loss percentiles at various confidence levels. The portfolio loss percentile at 1 percent is approximately $L(1\%) = 100$. The expected losses of A and B are respectively 7.0 and 2.5, totaling 9.5 for the portfolio. Capital is the loss percentile in excess of expected loss, or $100 - 9.5 = 90.5$. If the confidence level changes the loss percentiles change as well.

3.6 Conclusion

Developing a market for caps, collars, and floors for toxic assets as well as a uniform methodology to assess the risk contribution of such portfolios to financial institutions will unfreeze the market and reduce the likelihood of more systemic shocks from write-downs. This, together with regulatory intervention mandating the application of Basel II methodology for assessing capital adequacy, plus suspension of mark to market would aid complete recovery to eventual stability.

Notes

1. "Mozilo knew hazardous waste when he saw it", *Los Angeles Times*, 2009–06–04, http://latimesblogs.latimes.com/money_co/2009/06/the-use-of-toxic-to-describe-high-risk-mortgages-has-been-de-rigueur-for-the-last-two-years-now-it-looks-like-countrywide.html
2. "All the President's Zombies," *New York Times*, February 23, 2009, available online at http://www.nytimes.com/2009/02/23/opinion/23krugman.html. (Accessed March 22, 2009.)
3. "Treasury Details Plan to Buy Risky Assets", New York Times, Edmund L. Andrews and Eric Dash March 23, 2009.
4. "Fact Sheet Public-Private Investment Program." US Treasury. March 23, 2009. Available online at http://www.treas.gov/press/releases/reports/ppip_fact_sheet.pdf. (Accessed March 26, 2009.)
5. "The Put Problem with Buying Toxic Assets." SSRN.com. February 14, 2009. Available online at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1343625. (Accessed February 15, 2009).
6. "The Put Problem with Buying Toxic Assets."

7. Joe Weisenthal, "Banks Still Pricing Toxic Assets Ridiculously High," March 25, 2009, available online at http://www.businessinsider.com/banks-still-pricing-toxic-assets-ridiculously-high-2009–3.
8. Joseph Lazzaro, "Pricing System for Toxic Assets Deemed Key to US Treasury Bank Rescue Plan," February 11, 2009, available online at http://www.bloggingstocks.com/2009/02/11/pricing-system-for-toxic-assets-deemed-key-to-u-s-treasury-bank.
9. This section is largely adapted from the website for National Australia Bank, Agricultural Commodity Broking Services.
10. Adapted from J. C. McKissick, "Commodity Options as Price Insurance for Cattlement," The University Of Chicago Working Paper.
11. Sourced from Mckissick (2006:6).
12. Sourced from Bessis (2002: Chapter 51).
13. Sourced from Bessi (2002: 641–642).

References

Bawa, V. S. and Lindenberg, E. B. (1977) "Capital Market Equilibrium in a Mean-Lower Partial Moment Framework," *Journal of Financial Economics*, 5 (2): 189–200.

Fabozzi, F. J., Focardi, S. N., and Kolm, P. N. (2006) *Financial Modeling of the Equity Market,* Hoboken, NJ: John Wiley.

Hull, J. C. (2006) *Options, Futures, and Other Derivatives,* Upper Saddle River, NJ: Prentice-Hall.

McDonald, R. L. (2006) *Derivatives Markets,* Boston, Mass.: Addison-Wesley.

McNeil, A. J., Frey, R., and P. Embrechts (2005) *Quantitative Risk Management*, Princeton, NJ: Princeton University Press.

Nantell, T. J. and Price, B. (1979) "An Analytical Comparison of Variance and Semivariance Capital Market Theories," *Journal of Financial and Quantitative Analysis*, 14 (2): 221–242.

4

A General Efficient Framework for Pricing Options Using Exponential Time Integration Schemes

Yannick Desire Tangman, Ravindra Boojhawon, Ashvin Gopaul, and Muddun Bhuruth

4.1 Introduction

In numerical option pricing, spatial discretization of the pricing equation leads to semi-discrete systems of the form

$$V'(\tau) = AV(\tau) + b(\tau), \tag{4.1}$$

where $A \in \Re^{m \times m}$ is in general a negative semi-definite matrix and $b(\tau)$ generally represents boundary condition implementations, a penalty term for American option or approximation of integral terms on an unbounded domain in models with jumps. With advances in the efficient computation of the matrix exponential (Schmelzer and Trefethen 2007), exponential time integration (Cox and Matthews 2002) is likely to be a method of choice for the solution of ODE systems of the form (4.1). Duhamel's principle states that the exact integration of (4.1) over one time step gives

$$V(\tau_{j+1}) = e^{A\Delta\tau} V(\tau_j) + e^{A\tau_{j+1}} \int_{\tau_j}^{\tau_{j+1}} e^{-At} b(t)dt,$$

and approximation of the above equation by the exponential forward Euler method leads to the scheme

$$V^{j+1} = \varphi_0(A\Delta\tau)V^j + \Delta\tau\varphi_1(A\Delta\tau)b(\tau_j), \tag{4.2}$$

where $\varphi_0(z) = e^z$ and $\varphi_1(z) = (e^z - 1)/z$.

In a recent work (Tangman et al. 2008b), we investigated the use of exponential time-integration schemes for the numerical pricing of

European options under both the Black–Scholes model (Black and Scholes 1973) and Merton's jump-diffusion (MJD) model (Merton 1976) using a onetime step computation of φ_0. As we have noted, it is well known that only fast evaluations of these φ functions will make implementation of exponential integrators (4.2) efficient. In this chapter, we provide a more general and efficient framework for numerical pricing of options for which the price process is allowed to follow a variety of stochastic dynamics. Since most financial contracts usually have linear boundary conditions, we show that ETI can be easily adapted to price a variety of options under various models and that it is very competitive with existing numerical methods such as Crank–Nicolson. We improve on our previous work by developing fast implementation of the ETI scheme for pricing barrier and American-type options under various models including not only Black–Scholes and MJD but also under stochastic volatility (SV) (Heston 1993), stochastic volatility with jumps (SVJ) (Bates 1996) and CGMY (Carr et al. 2002) processes. Our algorithms rely on efficient techniques for computing the matrix exponential, and we present various numerical results indicating the success of the framework developed here for pricing options. For European options, only four sparse linear systems are required to obtain convergent option prices and hedging parameters making, ETI possibly the fastest Black–Scholes partial differential equation (PDE) solver. For pricing American options, we need to solve the linear complementarity problems (LCP) that arise. We make use here of an operator splitting technique together with the exponential forward Euler scheme to develop an algorithm with linear computational complexity. Extending to barrier options is straightforward by the simple implementation of the boundary condition at the barrier level. Further improvements in accuracy are achieved by employing a simple Richardson extrapolation method, making ETI a robust framework for pricing financial derivatives.

This chapter is structured as follows. In 4.2, we review the option pricing problem formulation for European, barrier, and American options under various models. Next, we describe the second order spatial discretization of the resulting PDEs or partial integro differential equations (PIDEs) and implementation of the boundary condition that leads to semi-discrete systems and show how to apply the ETI scheme. In 4.4, we study efficient evaluation of the matrix exponentiation based on best rational approximations (Schmelzer and Trefethen 2007; Trefethen et al. 2006). Numerical experiments are given in 4.5 followed by concluding remarks in 4.6.

4.2 Option pricing problem

We consider an economy consisting of a single risky underlying asset whose price dynamics are driven by different stochastic differential equations, thus characterizing various models. First, if the price S follows the geometric Brownian motion,

$$dS_t = (r - \delta)S_t dt + \sigma S_t dW_t, \tag{4.3}$$

where r is the interest rate, δ the amount of dividend, σ the volatility and W_t is a standard Wiener process, then a European option which gives the right but not the obligation to buy the asset at expiry T, solves the initial-boundary value problem of the Black–Scholes type

$$\frac{\partial V}{\partial \tau} = LV = \frac{1}{2}\sigma^2 \frac{\partial^2 V}{\partial x^2} + \left(r - \delta - \frac{1}{2}\sigma^2\right)\frac{\partial V}{\partial x} - rV, \; -\infty < x < \infty, \; 0 \le \tau \le T, \tag{4.4}$$

after the log transformation $x = \log(S/E)$ where E is the strike price, $\tau = T - t$ and L represents the spatial operator. The availability of closed form solutions has made Black–Scholes a very popular model. However, empirical evidence showed that this model cannot capture observed market features such as the fat tails and high peaks (asymmetric leptokurtic). Moreover, calibration with market prices showed that the volatility is not constant, as assumed by the Black–Scholes model and that market prices do jump. This is why many other models consisting of jumps (Merton 1976) and stochastic volatility (Bates 1996; Heston 1993) have appeared in the literature.

Merton (1976) was the first to explore option pricing where the underlying stock returns are discontinuous. For the jump-diffusion model, the dynamics of the stock price process are obtained by adding discontinuous Poisson jumps to (4.3) as

$$\frac{dS_t}{S_t} = (r - \delta - \lambda\kappa)dt + \sigma dW_t + (\eta - 1)dN_t,$$

and the Black–Scholes PDE (4.4) becomes a PIDE of the form

$$\begin{aligned}\frac{\partial V}{\partial \tau} &= \frac{1}{2}\sigma^2 \frac{\partial^2 V}{\partial x^2} + \left(r - \delta - \frac{1}{2}\sigma^2\right)\frac{\partial V}{\partial x} - rV \\ &\quad + \int_{-\infty}^{\infty} \left[V(x+z,\tau) - V(x,\tau) - (e^z - 1)\frac{\partial V}{\partial x}(x,\tau)\right] g(z)dz. \end{aligned} \tag{4.5}$$

This is just an extension of the Black–Scholes model and allows for sudden price movements (jumps) that can happen even over a small time step, dt. Naturally, when the mean arrival time λ of the independent Poisson process dN_t is zero, it reduces to the simple Black–Scholes model. Here $\kappa = \mathrm{E}(\eta - 1)$ and $(\eta - 1)$ represent the impulse function causing S to jump to S_η. For Merton's model, the jump size distribution is assumed to follow the normal density function

$$g(z) = \frac{\lambda}{\sqrt{2\pi}\sigma_J} e^{-(z-\mu_J)^2/(2\sigma_J^2)}, \tag{4.6}$$

with mean μ_J, variance σ_J^2 and for this model, κ can be evaluated analytically as $\exp(\mu_J + \sigma_J^2/2) - 1$. A richer model was constructed in Carr et al. (2002) based on the density function

$$g(z) = \frac{Ce^{-Mz}}{z^{1+Y}} \mathrm{I}_{z>0} + \frac{Ce^{Gz}}{|z|^{1+Y}} \mathrm{I}_{z<0},$$

leading to special cases of Kou's double exponential model (Kou 2002) for $Y = -1$, the VG process for $Y = 0$, infinite activity models with finite variation for $Y \in [0, 1]$ and infinite activity with infinite variation for $Y \in [1, 2]$. It is well known that the discontinuity of the kernel at $z = 0$ causes a lower order of convergence for difference schemes (Almendral and Oosterlee 2007; Wang et al. 2007). In addition, for infinite activity models, no diffusion ($\sigma = 0$) is required for the viscosity solution to converge. For a detailed discussion about jump processes, we refer to Cont and Tankov (2004). We confine ourselves to Merton's Gaussian distribution and the CGMY process right now, but the pricing methodology easily extends to other processes as well. When solving Equation (4.5), the important numerical issue is that the nonlocal nature of the convolution integral term present causes a dense matrix inversion for implicit schemes. Discretization of this convolution integral results in a Toeplitz matrix, and it is well known that fast Fourier transform (FFT) needs to be applied for computational efficiency. Carr and Mayo (2007) observed that for Merton's model, part of the integral in equation (4.5) represents the solution to a heat problem. In the next section, we will show how the ETI framework can fully exploit this idea.

Instead of adding a jump component to the stochastic differential equation (4.3), others have preferred to consider a nonconstant volatility. Heston (1993) proposed a model that assumes correlation with the stock process itself. This model is very popular among researchers since it admits a closed-form formula for European options, and it is easy to implement. However, it results in a multidimensional convection-diffusion equation with second-order cross-derivative. While the SV models give good calibrations for longer maturities, jumps are essential to reflect the short maturity patterns. Thus, it seems logical that the most generic model should be a combination of both SV and jump diffusion. The SVJ model was introduced by Bates (1996) and the stochastic differential equations governing the asset price process S and the variance process $y \geq 0$ are given by

$$\frac{dS_t}{S_t} = (r - \delta - \lambda\kappa)dt + \sqrt{y_t}dW_t^1 + (\eta - 1)dN_t,$$

$$dy_t = \alpha(\beta - y_t)dt + \gamma\sqrt{y_t}dW_t^2,$$

where W_t^1, W_t^2 are standard Brownian motions and γ is the volatility of y_t. Following Yan and Hanson (2006), the governing two-dimensional PIDE becomes

$$\frac{\partial V}{\partial \tau} = \frac{1}{2}y\frac{\partial^2 V}{\partial x^2} + \frac{1}{2}\gamma^2 y\frac{\partial^2 V}{\partial y^2} + \rho\gamma y\frac{\partial^2 V}{\partial x \partial y} + \left(r - \delta - \lambda\kappa - \frac{1}{2}y\right)\frac{\partial V}{\partial x}$$
$$+\alpha(\beta - y)\frac{\partial V}{\partial y} - (r+\lambda)V + \lambda\int_{-\infty}^{\infty}[V(x+z,y,\tau)]\,g(z)dz, \qquad (4.7)$$

where $\rho \in [-1,1]$ is the correlation coefficient and $g(z)$ is given by (4.6). For $\lambda = 0$, it reduces to Heston's SV model, and if γ is also zero, we obtain the simple Black–Scholes model.

It is the feature of each option that characterizes the condition at expiry and the boundary conditions at the ends of our computational domain. For a European call option, these conditions are

$$V(x,y,0) = \max(Ee^x - E, 0),$$
$$V_\tau(x,y,\tau) = -rV(x,y,\tau), \quad x \to -\infty,$$
$$V_{xx}(x,y,\tau) = V_x(x,y,\tau), \quad x \to \infty, \qquad (4.8)$$

and we can use one-sided approximations for the partial derivatives in y at the boundaries. For PDEs and PIDEs in one dimension, we simply omit the variable y.

A barrier option depends on whether the stock price hits a barrier or not. For example, an up-and-out-call barrier has the same payoff and lower boundary conditions as that of a European call option except for an upper condition at the barrier x_u as

$$V(x_u,y,\tau) = 0.$$

Nevertheless, most traded options are of American type, and for such options no simple analytical solution exists. An American option that can be exercised at any time up to and including maturity, gives rise to a linear complementary problem of the form

$$V\tau \geq LV, \; V(X,\tau) \geq V(X,0),$$
$$(V\tau = LV) \wedge (V(X,\tau) = V(X,0)), \tag{4.9}$$

where X consists of all the spatial dimensions involved, for example in the SVJ model, $X=(x,y)$. Therefore, for no-arbitrage to hold, the LCP states that the value of an American option must always be at least the payoff due to the early exercise feature.

4.3 Exponential time integration schemes

4.3.1 Black–Scholes

To describe ETI schemes, we first consider a European call option under the Black–Scholes model. For a finite difference discretization of the spatial derivatives in equation (4.4), we need to truncate the infinite x-domain $(-\infty, \infty)$ to a bounded domain $\Omega_x = (x_{\min}, x_{\max})$. We therefore consider a computational grid $\Omega_{\Delta x} \subset \Omega_x$ defined by

$$\Omega_{\Delta x} = \left\{x_i \in \Re : x_i = x_{\min} + i\Delta x, i = 0, 1, \ldots, m, \Delta x = (x_{\max} - x_{\min})/m\right\},$$

and define the central second-order approximations to the first- and second-order spatial derivatives with respect to x by the difference matrices

$$D_x^1 = \frac{1}{2\Delta x}\text{tridiag}[-1,0,1] \text{ and } D_x^2 = \frac{1}{(\Delta x)^2}\text{tridiag}[1,-2,1],$$

respectively. Then, by discretizing the Black–Scholes operator L in (4.4), we obtain the matrix

$$A = \frac{1}{2}\sigma^2 D_x^2 + \left(r - \delta - \frac{1}{2}\sigma^2\right) D_x^1 - rI_x, \tag{4.10}$$

where $I_x \in \Re^{m\times m}$ is the corresponding identity matrix for x. It is easy to implement the boundary conditions for a European call option by setting the only element in the first row of A as $A_{1,1} = -r$ and the linear boundary condition is implemented by using the one-sided second order approximation

$$V_x(x_{\max}, \tau) = \frac{V_{m-2} - 4V_{m-1} + 3V_m}{2\Delta x},$$

to find the last row of D_x^1 and then equate the last row of D_x^2 to that of D_x^1.

4.3.2 Merton's jump diffusion

In order to solve the PIDE for Merton's jump diffusion, we realize, as in Carr and Mayo (2007), that the integral $F = \int_{\Re} V(x+z,\tau)g(z)dz$ is equivalent to the solution of a heat problem, and, for $\mu_J \neq 0$, this solution is translated by μ_J such that the PDE satisfied by F is the convection-diffusion equation

$$\frac{\partial F}{\partial \tau} = \frac{\partial^2 F}{\partial x^2} + \frac{2\mu_J}{\sigma_J^2}\frac{\partial F}{\partial x}, \quad -\infty < x < \infty, \quad 0 \leq \tau \leq \frac{\sigma_J^2}{2},$$

with same initial and linear boundary conditions as for the European call problem (4.8). Hence, using the one-step exponential integration after discretizing the spatial operator to obtain the semi-discretization matrix $B = D_x^2 + \left(2\frac{\mu_J}{\sigma_J^2}\right)D_x^1$, we get

$$F = e^{B\left(\frac{\sigma_J^2}{2}\right)} V(x,0),$$

as the solution to the convection-diffusion problem which is also an approximation to the integral term. Thus the European option price will be

$$V(T) = e^{(A+F_M)T} V(0), \tag{4.11}$$

where $F_M = \lambda \exp(B\sigma_J^2/2)$. Here, we require a double matrix exponentiation, and we will explain in Section 4.4 how this is done efficiently by taking advantage of the commutativity feature of the differential and integral operator.

4.3.3 Carr-German-Madan-Yor (CGMY) model

For the CGMY model, we also consider the PIDE (4.5), but here we need to split the infinite domain of integration into $\Omega_{\Delta Z} = \Omega_{\Delta x}$ and $\Omega_{\Delta Z^*} \backslash \Omega_{\Delta Z}$ where $\Omega_{\Delta Z^*}$ is an extension of our domain such that the truncation error of the integral approximation on the unbounded domain is negligible. For example, we can take the domains $\Omega_Z = (-2,2)$ and $\Omega_{Z^*} = (-4,4)$. For jump processes such as in Merton's model, a simple composite trapezoidal discretization of the integral part will give second-order convergence (Tangman et al. 2008b). However, it is well known that a lower convergence rate is observed for infinite activity processes and the quadrature methods used in (Wang et al. 2007) are essential in order to obtain second-order convergence. In general, a special treatment of the singularity in the integrand at $z = 0$ is required

(Cont and Voltchkova 2005), and we further need to split the integral domain as $\Omega_0 = \{z : |z| \leq \Delta z/2\}, \Omega_1 = \Omega_{\Delta z} \backslash \Omega_0$ and $\Omega_2 = \Omega_{\Delta z^*} \backslash (\Omega_1 \cup \Omega_0)$. Then approximating the integro term in the PIDE (4.5) over Ω_{z^*} gives

$$F_D V_{\Omega_{\Delta z^*}} - \lambda(\Delta z) V(x,\tau) - \kappa(\Delta z) V_x(x,\tau) + \frac{\hat{\sigma}(\Delta z)}{2} V_{xx}(x,\tau),$$

where $\lambda(\Delta z) = \sum_{z_j \in \Omega_{\Delta z^*}} g(z_j)$, $\kappa(\Delta z) = \sum_{z_j \in \Omega_{\Delta z^*}} g(z_j)(e^{z_j} - 1)$, $V_{\Omega_{\Delta z^*}}$ represents the option values in $\Omega_{\Delta z^*}$, $\hat{\sigma}(\Delta z) = \int_{-\frac{\Delta z}{2}}^{\frac{\Delta z}{2}} (e^z - 1)^2 g(z) dz$ (see Tangman et al. 2010 for details) and

$$F_D V \Omega_{\Delta z^*} = F_L V_{\Omega_2^-} + F_M V_{\Omega_0 \cup \Omega_1} + F_R V_{\Omega_2^+},$$

where F_L, F_M and F_R are Toeplitz matrices. Here Ω_2^- and Ω_2^+ represent negative and positive nodes of Ω_2 respectively and can be approximated by the asymptotic option values for V_{Ω_2}.

Then we can formulate (4.5) as the semi-discrete linear system

$$V'(\tau) = (A + F_M) V(\tau) + b(\tau), \quad 0 \leq \tau \leq T, \tag{4.12}$$

where A is constructed in a way similar to equation (4.10) as

$$A = \frac{\sigma^2 + \hat{\sigma}(\Delta z)}{2} D_x^2 + \left(r - \delta - \frac{\sigma^2 + \hat{\sigma}(\Delta z)}{2} - \kappa(\Delta z) \right) D_x^1 - (r + \lambda(\Delta z)) I_x, \tag{4.13}$$

and the remaining integral term $b(\tau) = F_L V_{\Omega_2^-} + F_R V_{\Omega_2^+}$. For a call option, $V_{\Omega_2^-} = 0$ and $V_{\Omega_2^+} = Ee^x - Ee^{-r\tau}$. Integrating (4.12) with respect to time gives

$$V(T) = e^{(A+F_M)T} V(0) + e^{(A+F_M)T} \int_0^T e^{-(A+F_M)\tau} b(\tau) d\tau. \tag{4.14}$$

The special structure of the vector $b(\tau)$ allows a closed-form expression for the integral term in (4.14). It is easy to prove that the second term on the right hand side of (4.14) becomes

$$h = T\phi_1((A+F_M)T)(F_R \varepsilon_1) - ((A+F_M) + rI)^{-1} \left(e^{(A+F_M)T} - e^{-rT} \right) (F_R \varepsilon_2),$$

where $\varepsilon_1 = Ee^{\Omega_2^+}$ and ε_2 is a vector with entries equal to E.

4.3.4 Stochastic volatility and stochastic volatility with jumps

In the SV and SVJ models, we can define the computational grid on y and, hence, the difference matrices D_y^1, D_y^2 and I_y. Then, by using the

Kronecker product $\otimes$ we can easily approximate the spatial operator in equation (4.7) to obtain the block tridiagonal matrix $A \in \Re^{m^2 \times m^2}$ as

$$A = \frac{1}{2}y\left[I_y \otimes D_x^2\right] + \frac{1}{2}\gamma^2 y\left[D_y^2 \otimes I_x\right] + \rho\gamma y\left[\left(I_y \otimes D_x^1\right)\left(D_y^1 \otimes I_x\right)\right] + \left(r - \delta - \lambda\kappa - \frac{1}{2}y\right)\left(I_y \otimes D_{x'}^1\right) + \alpha\left(\beta - y\right)\left(D_y^1 \otimes I_x\right) - (r + \lambda)\left[I_y \otimes I_x\right], \tag{4.15}$$

and construct $V \in \Re^{m^2 \times 1}$ if m grid points are also used for $\Omega_{\Delta y}$. Including jumps will also be trivial using $F_M = \left[I_y \otimes \lambda \exp(B\sigma_j^2/2)\right]$ and the solution is given by equation (4.11).

4.3.5 American options

Finally, to solve American options, we will combine ETI with the operator splitting technique proposed in Ikonen and Toivanen (2004). Their method is based on transforming the inequalities in equation (4.9) into equalities by the addition of an auxiliary term $p(\tau)$and for an American option under the jump-diffusion model, this results in

$$V_\tau = (A + F_M)V + b + p(\tau),$$

where b represents a constant vector since for American options, the payoff values are used for V_{Ω_2}. On the other hand, $b = 0$ if the method given by Carr and Mayo (2007) is used. Then the constraints are enforced as

$$[V(x,\tau) - V(X,0)] \cdot p(\tau) = 0,$$
$$V(X,\tau) \geq V(X,0), \quad p(\tau) \geq 0.$$

The term $p(\tau)$ acts as a penalty term, which is positive if the American constraint is not satisfied and zero otherwise. Using the exponential forward Euler scheme (4.2), we get

$$\overline{V}(\tau_{j+1}) = \varphi_0\left((A + F_M)\,\Delta\tau\right)V(\tau_j) + \Delta\tau\,\varphi_1\left((A + F_M)\,\Delta\tau\right)\left(b + p(\tau_j)\right), \tag{4.16}$$

as a first step split solution with $\tau_{j+1} = \tau_j + T/n$. Then the American option price and the new penalty term are computed by

$$V(\tau_{j+1}) = \max\left(V(\tau_0), \overline{V}(\tau_{j+1}) + \Delta\tau\, p(\tau_j)\right),$$
$$p(\tau_{j+1}) = p(\tau_j) + \frac{1}{\Delta\tau}\left(\overline{V}(\tau_{j+1}) - V(\tau_{j+1})\right). \tag{4.17}$$

It is obvious that the most important part of the implementation will require the evaluation of the different $\varphi_l(A)$ functions for $l = 0, 1$ or, more precisely, their action on the vector V.

4.4 The matrix exponential

In this section, to describe the method for evaluating φ, we assume with no confusion that A is already scaled by $\Delta\tau$. Since Moler and Van Loan (2003), most exponential-type integrators use Padé approximation with scaling and squaring for computing the φ_l functions as suggested in Beylkin et al. (1998) and Minchev and Wright (2005). The same technique is used by the "expm" function in MATLAB®, and, in our case, since the semi-discretization matrix A is already scaled by $\Delta\tau$, the method works faster since less scaling and squaring is required. Recently, Ashi et al. (2009) compared the accuracy and computational time of several methods, and the scaling and squaring technique was found to be efficient. However, the amount of work required is about $O(m^3)$. This approach works well for matrices of moderate dimension, but in practice, when A is large and sparse, we prefer methods that simply approximate the action of the matrix function $\varphi_l(A)$ on the vector V. In this setting, one of the most promising techniques is based on best rational approximations.

4.4.1 Best rational approximations via Carathéodory–Fejér points

For the semi-discretization matrix A with eigenvalues in the left-half plane, it is sufficient to study an approximation for e^z on $z \in (-\infty, 0]$ in order to compute $\varphi_l(A)V$ (Schmelzer and Trefethen 2007). Evaluating the matrix exponentiation based on best rational approximation computed via Carathéodory–Fejér (Trefethen et al. 2006) is very promising since sparse direct solvers can efficiently be used to solve the resulting shifted linear systems of the form $(z_j I - A)x_j = V$. Following (Schmelzer and Trefethen 2007; Trefethen et al. 2006) and using the fact that a rational approximation can be interpreted as a quadrature formula or vice versa, we represent the rational approximation as the partial fraction expansion

$$\varphi_0(A)V \approx \sum_{j=1}^{\eta} c_j (A - z_j I)^{-1} V.$$

The residues c_j and poles z_j can be computed via Carathéodory-Fejér approximations and a MATLAB® code based on a singular value decomposition of a Hankel matrix is given in Trefethen et al. (2006). In

Schmelzer and Trefethen (2007), the code further considers the poles and residues for all φ_l functions showing even better accuracy as l increases. Instead of computing c_j and z_j for φ_1, Schmelzer and Trefethen (2007) used the common poles and residues of φ_0 and obtained the relation

$$\varphi_1(A)V \approx \sum\nolimits_{j=1}^{\eta} c_j z_j^{-1}(A - z_j I)^{-1} V. \tag{4.18}$$

Although this procedure is less accurate, yet it achieves reasonable accuracy for $\eta \geq 8$. This means that we need to solve the same set of linear systems for both φ_0 and φ_1. Also, since for financial problems the matrices that arise are real, the poles and residues come in complex conjugate pairs. Thus, only $\eta/2$ shifted linear systems solutions are required, making rational approximation very efficient. Furthermore, to evaluate equation (4.14), we can thus make use of the Carathéodory–Fejér approximation

$$\left(\frac{e^{A\Delta\tau} - e^{-r\Delta\tau}I}{A\Delta\tau + r\Delta\tau I}\right)V \approx \sum_{k=1}^{\eta} \frac{c_k}{z_k + r\Delta\tau}(A\Delta\tau - z_k I)^{-1} V.$$

For Merton's model, we note that it can prove costly to calculate equation (4.11) directly since the second exponential will have a dense matrix as argument due to the first matrix exponential. Instead, using the fact that integral and differential operators are commutative, we can approximate the two spatial operators in equation (4.11) as

$$\begin{aligned} V(T) &= e^{AT + F_M T} V(0) \approx e^{F_M T} e^{AT} V(0), \\ &= \exp(\lambda T \exp(B\sigma_J^2/2))[\exp(AT)V(0)]. \end{aligned}$$

For the Carathéodory–Fejér method, this means that we need only solve sparse shifted linear systems and thus gain in computational efficiency. To obtain the Carathéodory–Fejér points for the double exponential, we need to modify the MATLAB® code in Trefethen et al. (2006) as

$$F = \exp(\lambda T \exp(scl(t-1)/(t+1+1e^{-16}))) - 1,$$

to obtain $v_2 = \exp(F_M T)v_1 - v_1$ where $v_1 = \exp(AT)V(0)$ and F_M is as in (4.11). Hence $V(T)$ is the sum of v_1 and v_2, t is the Chebychev points, and *scl* represents the scaling factor for stability. Another possibility to prevent the dense matrix inversion is to use the regular matrix splitting (Almendral and Oosterlee 2005) and to solve the linear system using the fixed point iteration

$$V^k = (AT - z_k I)^{-1}[V(0) - (F_M T)V^{k-1}].$$

We can even use FFT to perform the dense matrix vector multiplication in $O(m \log m)$ for F_M as in equations (4.11) or (4.12). We now give the general algorithm for option pricing using ETI.

Algorithm 4.1: General algorithm for option pricing using ETI

1. **Problem start**
   a. Set the model parameters.
   b. Construct $\Omega_{\Delta x}$, $\Omega_{\Delta y}$ and Ω.
2. **Spatial discretization**
   c. Set $D_x^1, D_x^2, I_x, D_y^1, D_y^2, I_y$ using one-sided approximations if necessary.
   d. Implement any linear boundary condition $V_{xx} = V_x$.
   e. Build A using either equation (4.10), (4.13), or (4.15).
   f. Implement any other boundary conditions such as, for example, $A_{1,1} = -r$for a European call and $A_{m+1,1:m+1} = 0, A_{1:m+1,m+1} = 0$, for an up-and-out barrier.
3. **Integral approximation**
   g. Form $F_M = \lambda \exp(B\sigma_J^2/2)$ for Merton's jump-diffusion model.
   h. Form F_L, F_M, F_R and compute $V_{\Omega_2^-}$, $V_{\Omega_2^+}$ and ε_1, ε_2 for the CGMY model.
4. **Matrix exponential**
   i. viii. For the Carathéodory–Fejér method, use $\eta = 8$ to compute c_j, z_j and use (4.18) for odd j.
5. **Solution process: European**
   j. Set $\Delta\tau = T$.
   k. For Black–Scholes, SV, compute $V(T) = \exp(A\Delta\tau)V(0)$.
   l. For Merton's model, compute (4.11).
   m. For CGMY, compute (4.14).
   n. For SVJ, compute (4.11) with $F_M = [I_y \otimes \lambda \exp(B\sigma_J^2/2)]$.
6. **Solution process: American**
   o. Set n and compute $\Delta\tau = T/n$.
   p. for $j = 1, \ldots, n$, compute equations (4.16) and (4.17).

4.5 Numerical experiments

We now present the results of our numerical experiments which were carried out using MATLAB®. For all test cases, we use a computer with 1 GB RAM and 2.21 GHZ AMD Athlon X2 processor. Implementation on other machines and using other programming languages will give

Table 4.1 CPU(s) time in seconds and error at spot price for all methods for pricing a European call option

m	Crank–Nicolson	ETIExpm	ETICF	Error	m	Crank–Nicolson	ETIExpm	ETICF	Error
2^7	0.1720	0.0310	0.0160	0.0060	2^{10}	1.3120	30.422	0.0470	9.42e-5
2^8	0.3280	0.4690	0.0220	0.0015	2^{11}	2.6250	241.86	0.0780	2.36e-5
2^9	0.6410	3.9060	0.0310	3.77e-4	2^{12}	5.7190	–	0.1570	5.89e-6

somewhat different results. Unless specified otherwise, we will use the following table of parameters.

$S_0 = 100, \sigma = 0.2, \tau = 0.05, \delta = 0, E = 100, T = 1, S_B = 160,$
$\mu_J = 0, \sigma_J = 0.2, \lambda = 1, \rho = 0.1, \alpha = 5, \beta = 0.16, x_B = 0.8$

For SV and SVJ models: $T = 0.25, E = 1$
For the CGMY model: $C = 0.5, G = 10, M = 10$
For American call options: $\delta = 0.07$

Table 4.1 shows the speed of execution of different algorithms based on ETI and the Crank–Nicolson scheme for solving a European call option and also include the error at S_0. Clearly, all algorithms will have almost the same accuracy as m varies since the ETI method is exact in time, and, for a sufficiently refined time step, Crank–Nicolson will contain only spatial discretization error. However, we emphasize the huge computational speed improvements of the ETI scheme combined with the Carathéodory–Fejér method that also outperforms, by far, the Crank–Nicolson method. It is observed that the increase in computational time is linearly related to the increasing number of spatial steps, confirming that ETICF has indeed $O(\eta m/2)$ complexity. This is not the case of ETIExpm, which is $O(m^3)$ and thus has the worst CPU timing. This code could not be run for the last value of m due to the huge storage requirements needed by the "expm" function. Since the Crank–Nicolson method lacks L_0 stability, it is essential to restrict the time-step size at least as $O(\Delta x)$, and, for a fair comparison, we will choose $\Delta\tau$ such that the error is approximately the same for all methods. Hence, the Crank–Nicolson scheme, which is $O(nm)$ for n time steps, is slower since it needs to solve much more than the six linear systems required by the ETICF algorithm if $\eta = 12$. Actually, our intensive numerical experiments have shown that only four shifted linear systems solution are needed to obtain very satisfactory accurate prices. Henceforth, for option pricing under the ETI framework, we will use the ETICF with 8 Carathéodory–Fejér points.

With the striking low speed of the ETICF method, we now turn ourselves to the pricing of an American put option in the Black–Scholes framework. Since simple solutions do not exist for this option, we use the monotonically convergent binomial method of Leisen and Reimer (1996) with 15,001 steps as benchmark. Here the free-boundary value problem is solved by enforcing the American constraint at each time step. We consider a short ($T = 0.5$) and a long ($T = 3$) maturity example and compare the ETICF with the operator splitting technique (ETICFOS) with the commonly used Crank–Nicolson with projected successive over-relaxation method (CNPSOR) and Crank–Nicolson with operator splitting (CNOS).

Table 4.2 shows that PSOR converges very slowly for both short and long maturities. Both the ETICFOS and CNOS schemes seem very efficient as they are not only faster, but also give accurate option prices and the two hedging parameters, delta and gamma. In addition, we can see that doubling the number of spatial nodes approximately doubles the CPU time showing that both are algorithms with linear computational complexity (Tangman, Gopaul and Bhuruth, 2008a). This is of $\|p(\tau)\| \leq rE\ S_0$ for CNOS for each time step. This is why CNOS is about ten times faster than ETICFOS for short maturity options. However, for longer T, we can see that CNOS requires more time steps in order to achieve approximately, the same error as ETICFOS. Basically, the operator splitting technique used here is composed of two split steps. In the first one, we need to solve the BS PDE in time and the second step consists of enforcing the American constraint through the addition of a penalty term. While ETI performs the first step exactly in time even for large $V = (4V(2^{N+1}) - V(2^N))/3$, CN will need more time steps to be accurate over large T. Thus for longer maturity, the computational speed is almost the same for both methods.

We note that the ETI scheme solves the PDE part exactly in time and therefore, it is unconditionally stable. Furthermore, adding a penalty term which is always bounded as $\|p(\tau)\| \leq rE$ will not affect the stability of our method. Hence, we can use an extrapolation method to improve the accuracy of the ETICF method. We show in Table 4.3, the absolute extrapolated error at S_0 for pricing European, American and Barrier options under Black-Scholes, Merton's jump-diffusion, Heston's SV and Bates SVJ models. The cpu(s) required to calculate the most accurate extrapolated value $V = (4V(2^{N+1}) - V(2^N))/3$, is also given.

Table 4.2 Error for American put option with $\sigma = 0.4, r = 0.07, \delta = 0.03, E = 100$. The errors for the options $\Delta(\varepsilon_\Delta)$ and $\Gamma(\varepsilon_\Gamma)$ for $m = 720$ and $m = 800$ respectively are also shown

	ETICFOS			CNPSOR			CNOS		
M	**Error**	**Rate**	**CPU (s)**	**Error**	**Rate**	**CPU (s)**	**Error**	**rate**	**CPU (s)**
				$T = 0.5$ uniform grid.					
50	0.0308	–	0.1880	0.0306	–	0.8750	0.0307	–	0.0160
100	0.0079	1.9686	0.3280	0.0078	1.9816	2.9370	0.0076	1.9757	0.0320
200	0.0020	1.9985	0.5930	0.0019	2.0106	13.344	0.0019	2.0070	0.0620
400	4.64e−4	2.0841	1.1410	4.66e−4	2.0453	81.484	4.52e−4	2.1017	0.1100
800	9.26e−5	2.3252	2.2810	1.02e−4	2.1869	571.92	7.37e−5	2.6186	0.2190
$\varepsilon_\Delta = 6.69e-6$		$\varepsilon_\Gamma = 1.28e-7$		$\varepsilon_\Delta = 2.10e-6$		$\varepsilon_\Gamma = 2.90e-7$	$\varepsilon_\Delta = 3.53e-6$		$\varepsilon_\Gamma = 1.34e-7$
reference values		$V = 10.23868$	$\Delta = -0.42294$	$\Gamma = 0.01437$					
				$T = 3.0$ nonuniform grid.					
45	0.0219	–	0.4220	0.0216	–	2.3280	0.0216	–	0.4530
90	0.0053	2.0379	0.7040	0.0051	2.0825	7.9540	0.0051	2.0852	0.7500
180	0.0013	1.9940	1.2970	0.0012	2.0875	37.140	0.0012	2.1446	1.3430
360	3.16e−4	2.0823	2.4370	2.62e−4	2.1955	232.95	2.02e−4	2.5127	2.4690
720	3.01e−5	3.3891	4.8750	2.92e−5	3.1656	1674.1	3.66e−5	2.4640	4.8120
$\varepsilon_\Delta = 2.63e-6$		$\varepsilon_\Gamma = 4.86e-8$		$\varepsilon_\Delta = 2.82e-6$		$\varepsilon_\Gamma = 7.44e-8$	$\varepsilon_\Delta = 5.88e-6$		$\varepsilon_\Gamma = 2.32e-7$
reference values		$V = 20.79322$	$\Delta = -0.33145$	$\Gamma = 0.00630$					

Table 4.3 Extrapolated absolute error for pricing call options under different models

European			**Barrier**		**American**		
Black–Scholes geometric Brownian motion model							
Grid	Error	Ext.Err	Error	Ext.Err	Grid	Error	Ext.Err
$2^6 \times 1$	0.0242	–	0.0287	–	$2^6 \times 2^9$	0.0252	–
$2^7 \times 1$	0.0060	2.56e-5	0.0072	6.46e-6	$2^7 \times 2^9$	0.0064	1.18e-4
$2^8 \times 1$	0.0015	1.78e-7	0.0018	2.90e-7	$2^8 \times 2^9$	0.0016	3.33e-6
CPU(s)	0.0310		0.0310		1.8740		
Merton's jump-diffusion model							
Grid	Error	Ext.Err	Error	Ext.Err	Grid	Error	Ext.Err
$2^6 \times 1$	0.0207	–	0.0805	–	$2^6 \times 2^9$	0.0458	–
$2^7 \times 1$	0.0052	2.06e-5	0.0201	7.78e-5	$2^7 \times 2^9$	0.0114	7.32e-5
$2^8 \times 1$	0.0013	6.06e-7	0.0050	4.74e-6	$2^8 \times 2^9$	0.0016	3.69e-5
CPU(s)	0.1250		0.1250		1.2340		
Heston's stochastic volatility model							
Grid	Error	Ext.Err	Error	Ext.Err	Grid	Error	Ext.Err
$2^4 \times 2^5 \times 1$	0.0035	–	0.0031	–	$2^4 \times 2^5 \times 2^6$	0.0036	–
$2^5 \times 2^5 \times 1$	8.44e-4	5.81e-5	7.32e-4	7.31e-5	$2^5 \times 2^5 \times 2^6$	8.09e-4	1.06e-4
$2^6 \times 2^5 \times 1$	2.15e-4	3.42e-7	1.46e-4	4.99e-5	$2^6 \times 2^5 \times 2^6$	1.55e-4	6.26e-5
CPU(s)	3.6100		1.1560		5.0010		
Bates's stochastic volatility model							
Grid	Error	Ext.Err	Error	Ext.Err	Grid	Error	Ext.Err
$2^4 \times 2^5 \times 1$	0.0056	–	0.0028	–	$2^4 \times 2^5 \times 2^6$	0.0035	–
$2^5 \times 2^5 \times 1$	0.0013	1.61e-5	5.92e-4	1.38e-4	$2^5 \times 2^5 \times 2^6$	7.88e-4	1.10e-4
$2^6 \times 2^5 \times 1$	3.15e-4	3.21e-7	9.82e-5	6.66e-5	$2^6 \times 2^5 \times 2^6$	1.49e-4	6.41e-5
CPU(s)	4.7350		3.8450		5.0120		

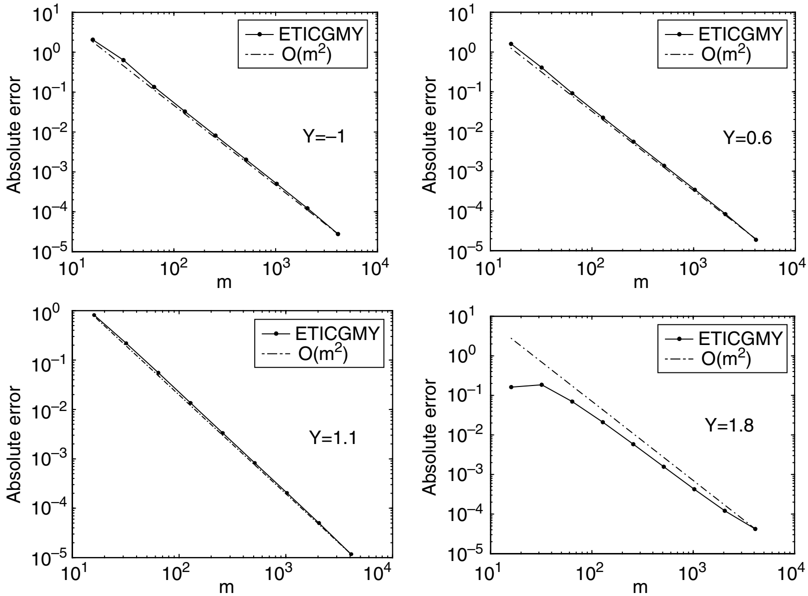


Figure 4.1 Convergence Rates of ETI for The CGMY Process With $C = 0.5$, $G = 10, M = 10$

For the CGMY model, we use $T = 0.5$ and give the convergence plot of a European call option in Figure 4.1 for various values of Y. For Kou's model ($Y = -1$), there is no need to split the integral domain near zero, but for $Y \in (0,2)$, we need to make use of the quadrature used by Wang et al. (2007) to obtain second-order convergence as seen in Figure 4.1. Here, good convergence rates are observed because of the better temporal accuracy of the ETI scheme and also because we did not require any interpolation in S since the integro and PDE grids coincide. But, as observed in Wang et al. (2007), the convergence decays as $Y \to 2$. Due to space limitations, we do not report in detail on computed option prices that have been seen to work well for the simple Richardson extrapolation technique also.

The case $Y = 0, \sigma = 0$ results in the VG model (Madan et al. 1998), and the PIDE becomes convectively dominated. Almendral and Oosterlee (2006) suggested the use of a simple Lax–Wendroff update by adding the diffusive term

$$\frac{\Delta\tau}{2}\left(r - \delta - \frac{1}{2}(\sigma^2 + \hat{\sigma}(\Delta z)) - \kappa(\Delta z)\right)^2 V_{xx}(x,\tau),$$

Table 4.4 Convergence ratio for European under VG model, American and barrier options

	European VG		American ($Y = 0.5$)		Barrier ($Y = 0.1$)	
M	**Value**	**Ratio**	**Value**	**Ratio**	**Value**	**Ratio**
2^5	0.3292	–	5.4351	–	7.2414	–
2^6	0.3795	–	6.0171	–	7.2758	–
2^7	0.3966	2.9357	6.1427	4.6353	7.2843	4.0390
2^8	0.4010	3.9509	6.1729	4.1520	7.2854	4.0586
2^9	0.4021	3.7236	6.1805	3.9737	7.2869	4.1226
2^{10}	0.4024	4.0399	6.1824	4.0181	7.2871	4.2661
Reference value	0.40239		6.18275		–	

to obtain more accurate solutions. Another approach is to combine the semi-Lagrangian discretization (d'Halluin et al. 2005; Wang et al. 2007). However, we do not pursue this here and show in Table 4.4 that the Lax–Wendroff update does improve the accuracy and convergence ratio. We also give in the same table the convergence ratio for an American option with the benchmark solution obtained by using the extrapolated FFT method for Bermudan options (Lord et al. 2008). The convergence ratio for a double knockout barrier at $|x| = 0.5$ is also included to show that second order is achieved for infinite activity processes. This option has not been priced before, and this is why we did not include any reference value. The numerical results obtained for the large variety of models and options characterizes the ETICF scheme as a method of choice for fast option pricing.

4.6 Conclusion

The ETI framework is shown to be general and robust enough to price a variety of options under various models used in literature. Using a recently developed method for matrix functions, only four sparse tridiagonal shifted linear systems are required to be solved to obtain accurate option prices for European and barrier options. Together with an operator splitting technique, American options can also be priced very efficiently. Further accuracy improvements can be obtained by using a simple Richardson extrapolation formula. With a general, fast, and accurate pricing algorithm, the calibration task can now be performed.

Current investigations include a study of other spatial discretization such as finite element and spectral methods (Tangman et al. 2008b) together with ETICF. The implementation of efficient solvers for shifted linear systems (Frommer 2003; Simoncini 2003) is primordial, especially for multidimensional option pricing problems.

Acknowledgments

The authors thank Thomas Schmelzer for his help concerning the efficient computations of the matrix exponentials for Merton's model.

References

Almendral, A. and Oosterlee, C. W. (2006) "Highly Accurate Evaluation of European and American Options under the Variance Gamma Process," *Journal of Computational Finance*, 10 (1): 21–42.

Almendral, A. and Oosterlee, C. W. (2007) "Accurate Evaluation of European and American Options under the CGMY Process," *SIAM Journal of Scientific Computing*, 29 (1): 93–117.

Almendral, A. and Oosterlee, C. W. (2005) "Numerical Valuation of Options with Jumps in the Underlying," *Applied Numerical Mathematics*, 53 (1): 1–18.

Ashi, H. A., Cummings, L. J., and Matthews, P. C. (2009) "Comparison of Methods for Evaluating Functions of a Matrix Exponential," *Applied Numerical Mathematics*, 59 (3–4): 468–486.

Bates, D. (1996) "Jump and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark in Options," *Review of Financial Studies*, 9 (1): 69–107.

Beylkin, G., Keiser, J. M., and Vozovoi, L. (1998) "A New Class of Time Discretization Schemes for the Solution of Nonlinear PDEs," *Journal of Computational Physics*, 147 (2): 362–387.

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

Carr, P. and Mayo, A. (2007) "On the Numerical Evaluation of Option Prices in Jump Processes," *The European Journal of Finance*, 13 (4): 353–372.

Carr, P., Geman, H., Madan, D. B., and Yor, M. (2002) "The Fine Structure of Assets Returns: An Empirical Investigation," *Journal of Business*, 75 (2): 305–332.

Cont, R. and Tankov, P. (2004) *Financial Modelling with Jump Processes,* Boca Raton, Fla.: Chapman & Hall/CRC Press.

Cont, R. and Voltchkova, E. (2005) "Finite Difference Methods for Option Pricing in Jump-Diffusion and Exponential Lévy Models," *SIAM Journal of Numerical Analysis*, 43 (4): 1596–1626.

Cox, S. M. and Matthews, P. C. (2002) "Exponential Time Differencing for Stiff Systems," *Journal of Computational Physics*, 176 (2): 430–455.

d'Halluin, Y., Forsyth, P. A. and Vetzal, K. R. (2005) "Robust Numerical Methods for Contingent Claims under Jump Diffusion Processes," *Journal Numerical Analysis*, 25 (1): 87–112.

Frommer, A. (2003) "BiCGStab (l) for Families of Shifted Linear Systems," *Computing*, 70 (2): 87–109.

Heston, S. (1993) "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies*, 6 (2): 327–343.

Ikonen, S. and Toivanen, J. (2004) "Operator Splitting Methods for American Option Pricing," *Applied Mathematical Letters*, 17 (7): 809–814.

Kou, S. (2002) "A Jump Diffusion Model for Option Pricing," *Management Science*, 48 (8): 1086–1101.

Leisen, D. and Reimer, M. (1996) "Binomial Models for Option Valuation-Examining and Improving Convergence," *Applied Mathematical Finance*, 3 (4): 319–349.

Lord, R., Fang, F., Bervoets, F., and Oosterlee, C. W. (2008) "A Fast and Accurate FFT-Based Method for Pricing Early-Exercise Options under Lévy Processes," *SIAM Journal of Scientific Computing*, 30 (4): 1678–1705.

Madan, D. B., Carr, P., and Change, E. (1998) "The Variance Gamma Process and Option Pricing," *European Finance Review*, 2 (1): 79–105.

Merton, R. C. (1976) "Option Pricing when the Underlying Stocks Are Discontinuous," *Journal of Financial Economics*, 3 (1–2): 125–144.

Minchev, B. V. and Wright, W. M. (2005) "*A Review of Exponential Integrators for First Order Semi-Linear Problems*," Technical Report: Norwegian University of Science and Technology, Trondheim, Norway.

Moler, C. and Van Loan, C. (2003) "Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later," *SIAM Review*, 45 (1): 3–49.

Schmelzer, T. and Trefethen, L. N. (2007) "Evaluating Matrix Functions for Exponential Integrators via Carathéodory–Fejér Approximation and Contour Integrals," *Electronic Transactions on Numerical Analysis*, 29 (2): 1–18.

Simoncini, V. (2003) "Restarted Full Orthogonalization Method for Shifted Linear Systems," *BIT Numerical Mathematics*, 43 (2): 459–466.

Tangman, D. Y., Gopaul, A., and Bhuruth, M. (2008a) "A Fast High-Order Finite Difference Algorithm for Pricing American Options," *Journal of Computational Applied Mathematics*, 222 (1): 17–29.

Tangman, D. Y., Gopaul, A., and Bhuruth, M. (2008b) "Exponential Time Integration and Chebychev Discretisations Schemes for Fast Pricing of Options," *Applied Numerical Mathematics*, 58 (9): 1309–1319.

Tangman, D. Y., Peer, A. A. I., Rambeerich, N., and Bhuruth, M. (2010) "Fast Simplified Approaches to Asian Option Pricing," *Journal of Computational Finance*, to appear.

Trefethen, L. N., Weideman, J. A., and Schmelzer, T. (2006) "Talbot Quadratures and Rational Approximations," *BIT Numerical Mathematics*, 46 (3): 653–670.

Wang, I. R., Wan, J. W., and Forsyth, P. A. (2007) "Robust Numerical Valuation of European and American Options under the CGMY Process," *Journal of Computational Finance*, 10 (4): 31–69.

Yan, G. and Hanson, F. B. (2006) "Option Pricing for a Stochastic-Volatility Jump-Diffusion Model with Log-Uniform Jump-Amplitudes," *Proceedings of American Control Conference*: 2989–2994.

5
Unconditional Mean, Volatility, and the FOURIER–GARCH Representation

Razvan Pascalau, Christian Thomann and Greg N. Gregoriou

5.1 Introduction

Recently there has been an upsurge interest in modeling the nonstationarities present in the volatility of financial data. The clustering and the persistence of volatility of asset returns have been well documented. The IGARCH model of Engle and Bollerslev (1986), for instance, describes in a parsimonious way the high persistence in the conditional volatility of stock returns while the underlying process remains strictly stationary. Alternatively, Granger (1980) and Granger and Joyeux (1980) model the long memory or the long-range dependence of a series of log returns as a fractionally integrated process to allow the autocorrelation functions to decay very slowly, in a fashion characteristic of stock returns. However, seminal papers from Granger and Joyeux (1986), Lamoureux and Lastrapes (1990), and, more recently, from Diebold and Inoue (2001), Mikosch and Starica (2004), Starica and Granger (2005), and Perron and Qu (2007) argue that the high persistence close to unit root and long memory both in the first and the second moments may actually be caused by structural changes in the level or slope of an otherwise locally stationary process of the long-run volatility. Diebold and Inoue (2001) argue that this is due to switching regimes in the data. Mikosch and Starica (2004) provide theoretical evidence that changes in the unconditional mean or variance induce the statistical tools (e.g., sample ACF, periodogram) to behave the same way they would if used on stationary long-range dependent sequences. Starica and Granger (2005) also deliver evidence against global stationarity. Finally, Perron and Qu (2007) conclude that the Standard & Poor's (S&P) 500 return series is best described as a stationary short memory process contaminated by mean shifts. These results imply that a good model for volatility should take into account the possibility of a time-varying unconditional second moment and, possibly, of a time-varying first moment as well.

Engle and Rangle (2008) propose the Spline-GARCH to model long-run volatility non-parametrically using an exponential quadratic spline. However, they do so only for the second moment. Further, Starica and Granger (2005) use step functions to approximate nonstationary data locally by stationary models. They apply their methodology to the S&P 500 series of returns covering a period of seventy years of market activity and find that most of the dynamics are concentrated in shifts of the unconditional variance.

However, these models pose several problems. While spline functions may lead to overfitting, step functions may not give smooth approximations. Even major breaks, such as the stock-market crash of 1929 and the oil-price shocks of the 1970s did not display their full impact immediately. Structural changes may take longer to extinguish, which suggests they need to be modeled as smooth or gradually changing processes. These arguments motivate the present study to propose a new approach to model the long-run first and second moments as smooth processes. This chapter denotes the new process Fourier–GARCH because it uses the flexible Fourier transform of Gallant (1981) (i.e., an expansion of a periodic function in terms of an infinite sum of sines and cosines). The basic model can be extended to incorporate the long-run volatility in the mean model. Flexible Fourier transforms have been used in the literature to approximate nonlinear structures in several ways. For instance, Becker et al. (2001) use Fourier transforms to model inflation and money demand as having smooth changes in the intercept. Also, Enders and Lee (2006) and Becker et al. (2006) propose new unit root and stationarity tests that use the Fourier approximation to model the unknown shape of the structural breaks in macro time series. The main advantage is that the issue of estimating the shape and location of the breaks reduces to selecting the proper frequency of the Fourier sine and cosine terms. A section below details how Fourier transforms can be used to approximate various types of breaks.

The study applies the new model to several of the largest stocks from S&P 500 to estimate volatility persistence in stock returns. Based on the discussion above, this chapter considers several competing models. The basic Fourier–GARCH model specifies a constant first moment, while the second moment changes smoothly over time. A first extension to the basic model allows both the first and the second moments to vary over time, while a second extension incorporates the long-run volatility in the model for the mean. This chapter checks for each model the sum of the estimated coefficients in the equation for conditional volatility to assess the so-called long-memory effect. The results show that allowing only the second moment to vary over time does not significantly reduce the persistence effect. In fact, the difference between this model and the simple GARCH(1,1) is negligible. However, the extended model that allows the first moment to vary over time as well reduces the persistence effect by more than half of the value suggested by GARCH(1,1). The evidence suggests that the persistence effect seen in stock returns is mainly a result of the misspecification of the model for the mean.

This chapter is structured as follows. Section 5.2 discusses in more detail the performance of the Fourier series to approximate various types of structural breaks. Section 5.3 introduces the basic Fourier–GARCH model and its extensions. Section 5.4 discusses the empirical estimates of the long memory effect using four different models, and Section 5.5 concludes.

5.2 Nonlinear trend approximation with Fourier transforms

The general approach to account for breaks is to approximate them using dummy variables. However, this approach has several undesirable consequences. First, one has to know the exact number and location of the breaks. These are not usually known and therefore need to be estimated. This, in turn, introduces an undesirable preselection bias (see Maddala and Kim 1998). Second, use of dummies suggests sharp and sudden changes in the trend or level. However, for low-frequency data it is more likely that structural changes take the form of large swings in the data which cannot be captured well using only dummies. Breaks should therefore be approximated as smooth processes (see Leybourne et al. 1998 and Kapetianos et al. 2003).

Flexible Fourier transforms, originally introduced by Gallant (1981), are able to capture the essential characteristics of one or more structural breaks using only a small number of low-frequency components. This is true because a break tends to shift the spectral density function toward frequency zero. Below is illustrated the ability of Fourier transforms to capture nonlinear trends.

Using a simple form for the mean model, one can allow the intercept μ_t to be a deterministic function of time:

$$y_t = \mu_t + \gamma_t + \varepsilon_t \tag{5.1}$$

where the drift term is written as:

$$\mu_t = c_0 + \Sigma_{k=1}^{s} c_k \sin\left(\frac{2\pi kt}{T}\right) + \Sigma_{k=1}^{s} d_k \cos\left(\frac{2\pi kt}{T}\right);\ s \leq T/2 \tag{5.2}$$

In the above formulation, ε_t is a stationary disturbance term with variance σ_s^2, s is the maximum number of frequencies, k is a particular frequency, and T is the total number of observations. The drift term represents the Fourier approximation written as a deterministic function of sine and cosine terms. Note that by imposing $\alpha_k = \beta_k = 0$ one gets the constant mean or trend return specification. In contrast to other possible series expansions (e.g., Taylor series), the Fourier expansion has the advantage of acting as a global approximation (see Gallant 1981). This property is obtained even if one specifies a small number of frequencies. In fact, Enders and Lee (2006) argue that a large value of s in a regression framework uses many of the degrees of freedom and leads to an overfitting problem.

To illustrate the approximation properties of a Fourier series, this chapter considers first a single frequency in the data-generating process (DGP):

$$\mu_t = c_0 + c_k \sin\left(\frac{2\pi kt}{T}\right) + d_k \sin\left(\frac{2\pi kt}{T}\right) \tag{5.3}$$

where k is the single frequency selected in the approximation, and c_k and d_k represent the magnitudes of the sinusoidal terms.

This study considers several possible patterns for the occurrence of a break. Thus, for $T = 500$, this chapter simulates one break, two breaks, and trend breaks both in the middle and toward the extremes. This chapter illustrates the cases for temporary, permanent, and reinforcing breaks. We display the results below in panels 1 through 9 (i.e., Figure 5.1). As in Enders and Lee (2006), Panels 1 and 2 illustrate approximations for breaks toward the end of a series. In Panel 3, the series has a temporary, though long-lasting break. Panels 4 and 5 display permanent breaks in opposite directions while in Panel 6 the breaks are in the same direction. Finally, Panels 7–9 depict breaks in the intercept and slope of a trending series. This chapter estimates the coefficients of the sinusoidal terms by performing a simple regression of y_t on μ_t and a time trend.

One can draw several conclusions based on the visual inspection of the graphs. First, a single frequency $k = 1$ or two cumulative frequencies $n = 2$ can approximate a large variety of breaks. Second, the Fourier transform approximates well even when the breaks are asymmetric (see Panels 1 and 2). Third, a Fourier series works best when the break is smooth over

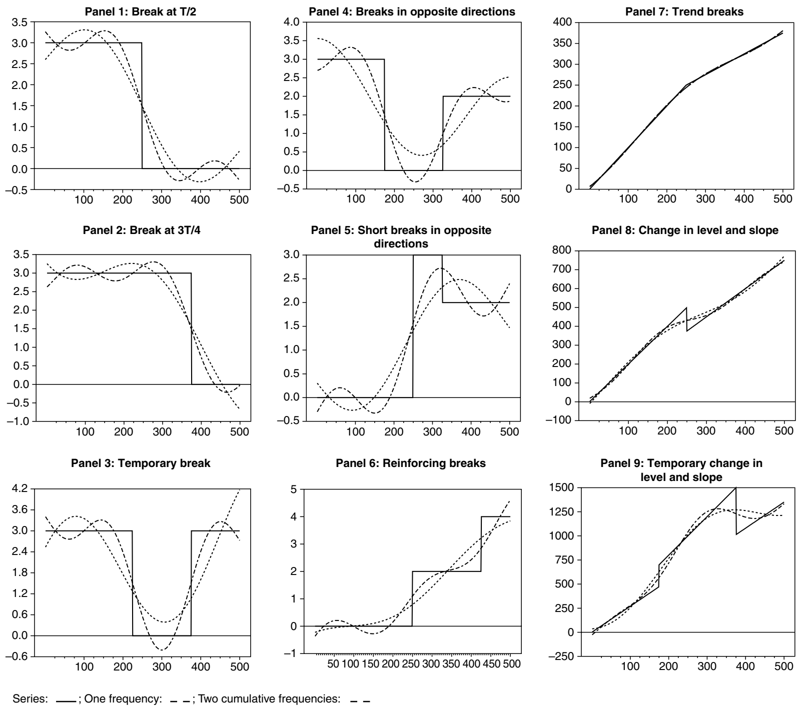


Figure 5.1 Approximation of structural breaks with Fourier Transforms

time, which means it may not be suited for abrupt and sharp breaks of short duration (see Panel 5). An additional frequency of $k = 2$ can improve the fit in this situation. Interested readers are referred to Enders and Lee (2006) and Becker et al. (2006) who have a longer discussion on the properties of the Fourier approximations. The next section introduces a new model to approximate long-run volatility.

5.3 A new model for unconditional volatility

As the introductory part suggested, the simple GARCH (1,1) may not be appropriate because it implies a long-run level of the volatility which is constant. However, previous research regarding the presence of various shifts in stock returns suggests that structural changes in the second moment induce global nonstationarity. This invalidates the use of the simple GARCH(1,1). It is known that breaks shift the spectral density function toward frequency zero. This indicates that the frequencies

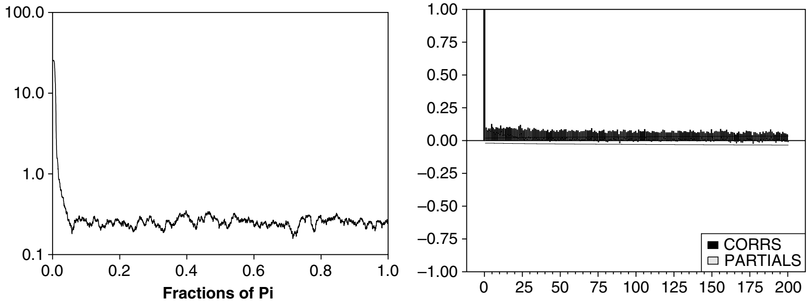


Figure 5.2 Left: Sample spectrum of absolute returns of S&P 500; Right: Sample ACF of returns of S&P 500

to be used are toward the low end of the spectrum (see Enders and Lee 2006). A simple visual inspection of the autocorrelation function and periodogram of absolute returns of S&P 500 confirms this fact (see Figure 5.2).

As you can note from the graphs in Figure 5.2, the most important frequencies that have an impact on the absolute returns are at the low end of the sample spectrum, which is indicative of structural breaks. Both graphs confirm the presence of long memory in financial returns – slow decay with lags still significant at the 200th lag. These findings suggest the use of the following model whose aim is to capture various unknown shifts in long-run volatility. This chapter denotes it the basic Fourier–GARCH:

$$r_t = \mu + \upsilon_t \sqrt{u_t h_t}, \text{ where } \upsilon_t \left| I_{t-1} \sim iid(0,1) \right. \tag{5.4}$$

$$h_t = (1-\alpha-\beta) + \alpha \frac{(r_{t-1}-\mu)^2}{u_{t-1}} + \beta h_{t-1} \tag{5.5}$$

$$u_t = \exp\left[a_0 + \sum_{k=1}^{s} \left(a_k \sin\left(\frac{2\pi kt}{T}\right) + b_k \cos\left(\frac{2\pi kt}{T}\right)\right)\right]; \; s \le T/2 \tag{5.6}$$

The model preserves the parsimony of the GARCH(1,1) model while it allows the unconditional expectation of the volatility to be a function of time and of cycles of different frequencies. A simple extension allows the unconditional mean to be a function of time as well: higher unconditional variance certainly requires higher unconditional mean.

The time-varying first moment is also approximated using a Fourier representation:

$$\mu_t = c_0 + \sum_{k=1}^{s} c_k \sin\left(\frac{2\pi kt}{T}\right) + \sum_{k=1}^{s} d_k \cos\left(\frac{2\pi kt}{T}\right) \tag{5.7}$$

Given its flexible setup, the Fourier–GARCH captures both short- and long-run dynamics. Note that:

$$E(r_t - \mu)^2 = E(\upsilon_t^2 u_t h_t) = u_t E(h_t) = u_t \tag{5.8}$$

The study uses an exponential representation of the Fourier transform to ensure its positivity. Goodness-of-fit measures such as the BIC or AIC criteria are employed to choose the proper number of frequencies exogenously. They are computed as follows:

$$AIC = -lnL + 2n, \quad L = -\sum_{t=1}^{T}\left[\ln(h_t u_t) + \frac{(r-\mu)^2}{h_t u_t}\right] \tag{5.9}$$

$$BIC = -lnL + nln(T), \quad L = -\sum_{t=1}^{T}\left[\ln(h_t u_t) + \frac{(r-\mu)^2}{h_t u_t}\right] \tag{5.10}$$

Here, n denotes the number of parameters estimated by the model. The advantage of using the AIC and BIC criteria is that they include a penalty for the additional estimated parameters. Throughout the estimation, the criteria employ only integer frequencies.

The advantage of using a time-varying first moment for a sample of forty years of daily data of S&P 500 absolute returns is highlighted in Figure 5.3.

Note the better fit of the second model, which augments the basic Fourier–GARCH representation with a time-varying intercept as in equation (5.7). However, given the presumption that a higher long-run volatility requires a higher long-run return, this chapter proposes the Fourier–M model which includes the unconditional time-varying volatility in the equation for the mean:

$$r_t = \gamma u_t + \upsilon_t \sqrt{u_t h_t}, \; where \; \upsilon_t \left| I_{t-1} \sim iid(0,1)\right. \tag{5.11}$$

In this way, both the first and the second moment change over time while the underlying model ensures a parsimonious representation.

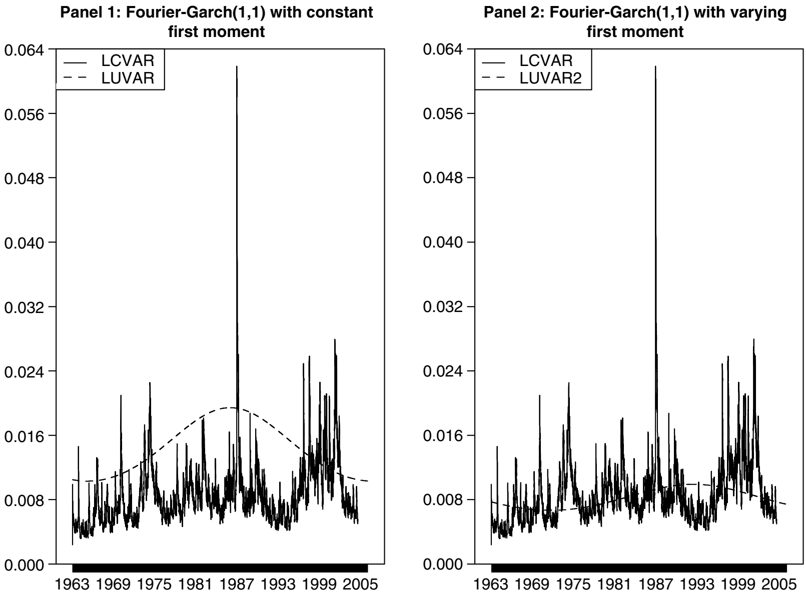


Figure 5.3 Panel 1: LCVAR = conditional volatility; LUVAR = unconditional volatility; Panel 2: LCVAR = conditional volatility; LUVAR2 = unconditional volatility

One way to assess the persistence or long memory in stock returns is to compute the sum of the slope coefficients in conditional volatility. If the sum is close to one, then conditional volatility is said to be almost integrated and it displays very slow time decay. However, the support for long memory is weakened if one finds that a changing first and/or second moment is responsible for the persistence effect. If the sum of the coefficients is significantly less than one after one accounts for shifts in the unconditional mean or volatility, then one can conclude that the volatility process is stationary but suffers from structural shifts (see Perron and Qu 2007).

A sample of daily returns on S&P 500 from 01/02/1963 to 02/30/2005 illustrates this discussion. The best representation is the one that specifies a single frequency both for the mean and for the unconditional volatility (see Figure 5.4).

Note the slow and gradual increase of long-run volatility from the 1960s until the 1980s. Also, note that the estimated long-run volatility of the 1990s is lower than the one for previous decades, which is consistent with market facts.

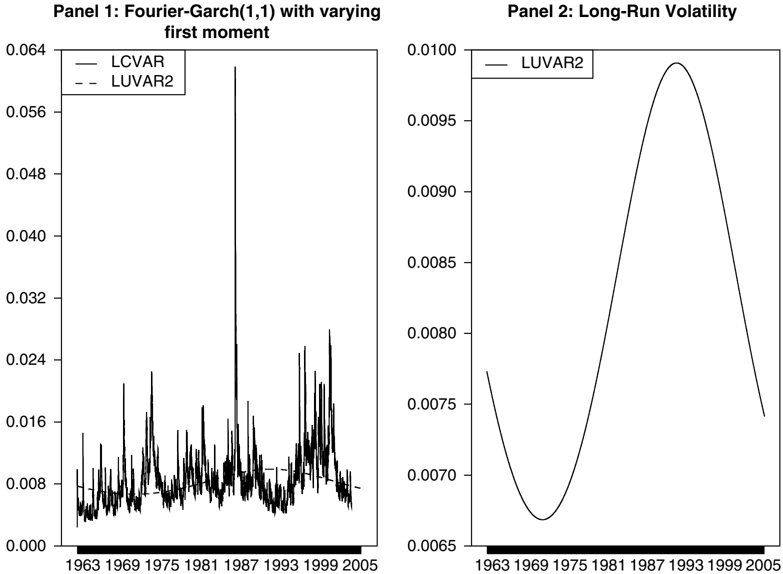


Figure 5.4 Panel 1: LCVAR – conditional volatility, LUVAR2 – unconditional volatility; Panel 2: LUVAR2 – unconditional volatility

5.4 Model validation and persistence effects

This chapter uses several representative stocks of S&P 500 to assess the long-memory effect of stock returns using the new models. The first 12 stocks of the index are selected according to their market percentage participation as of March 2005. Table 5.1 shows their ticker, sector classification and percentage of total assets.

The data has been obtained from the Center of Research in Security Prices made available through the WRDS database. The longest sample period available is 01/02/1926–12/30/2005 and corresponds to Exxon, IBM, Chevron, Philip Morris, and General Electric. Other stock returns have shorter sample periods (i.e., Procter & Gamble from January 2, 1929 onwards, Pfizer and Johnson & Johnson from 1944, and Intel from 1972; while the rest start in 1986). For each stock return, the study chooses exogenously an integer or cumulative frequencies according to the AIC and BIC criteria. According to Enders and Lee (2006), a frequency greater than 5 uses many of the degrees of freedom and leads to an overfitting problem.

Table 5.1 Market capitalization of 13 companies on S&P 500 as of February 28, 2006

Ticker	Issue name	Sector	% of Total assets
XOM	Exxon Mobil Corp	Energy	3.19
GE	General Electric Co.	Industrials	3
MSFT	Microsoft Corp.	Industrials	2.12
C	Citigroup Inc	Financials	2.03
PG	Procter & Gamble	Consumer staples	1.73
PFE	Pfizer Inc.	Health care	1.67
AIG	American Intl. Group Inc.	Financials	1.49
JNJ	Johnson & Johnson	Health care	1.48
MO	Altria Group Inc.	Consumer staples	1.29
CVX	Chevron Corp New	Energy	1.09
IBM	International Business Mach.	Information technology	1.09
INTC	Intel Corp	Information technology	1.07

Table 5.2 displays the results from applying the AIC and BIC criterions to identify the best in sample fitting model. The above mentioned criterions indicate that in most cases the best representation is the basic Fourier-Garch(1,1) model. The coefficients of the sine and cosine terms with up to 5 frequencies are significant at the 5% level both for the basic and for the extended models. However, given that in the model for the mean each additional frequency requires the estimation of two more coefficients, the additional penalty increases the values of the AIC and BIC criterions relative to the ones for the basic model. This is not surprising given that the BIC criterion favors more parsimonious representations. Several exceptions to the finding above are noteworthy. In the case of Microsoft for instance, both criterions select the Fourier-M model to be the optimal representation. Also, the Fourier-M model gives the best fit for Chevron as well. Note that the basic Fourier-Garch(1,1) and the Fourier-M models have very close values for the BIC and SBC criterions. This is true because they estimate the same number of parameters (i.e. six coefficients). In rest, the increased penalty due to the additional coefficients that are estimated in the models with two or more cumulative frequencies is greater than the better fit that is obtained. Therefore, the single frequency representation fits the data best for all models. Figures 5.5 through 5.7 show several graphs of the conditional and long-run volatilities obtained using both a constant and a time varying first moment. Note that for all series the long run volatility changes smoothly over time.

Table 5.2 AIC, BIC, and the log-likelihood

(a)

	AIG			Chevron			Citigroup			Exxon			General Electric		
Frequencies	AIC	BIC	(ℓ)	AIC	BIC	(ℓ)	AIC	BIC	(ℓ)	AIC	BIC	(ℓ)	AIC	BIC	(ℓ)
1	0.989	40.287	11.011	0.044	47.720	11.956	1.609	40.512	10.391	0.003	47.679	11.997	0.048	47.724	11.952
2	4.987	61.217	11.013	4.035	67.603	11.965	5.634	57.505	10.386	4.005	67.574	11.995	4.076	67.643	11.924
3	8.987	79.275	11.013	8.034	87.495	11.965	9.617	74.456	10.383	8.005	87.465	11.995	8.046	87.506	11.954
4	12.987	97.332	11.013	12.035	107.387	11.965	13.614	91.420	10.387	12.005	107.357	11.995	12.047	107.398	11.953
5	16.987	115.390	11.013	16.035	127.278	11.965	17.615	108.389	10.385	16.005	127.249	11.995	16.046	127.290	11.954
1 (mean shifts)	5.132	61.363	10.868	4.645	68.213	11.354	5.936	57.807	10.064	5.488	69.056	10.512	4.375	67.943	11.625
1 (Fourier-M)	1.002	43.174	10.985	0.039	47.714	11.961	4.757	43.660	7.243	0.008	47.683	11.992	0.074	47.750	11.926

(b)

	IBM			Intel			Johnson & Johnson			Microsoft			Pfizer		
Frequencies	AIC	BIC	(ℓ)	AIC	BIC	(ℓ)	AIC	BIC	(ℓ)	AIC	BIC	(ℓ)	AIC	BIC	(ℓ)
1	0.028	47.704	11.972	1.142	48.818	10.858	0.334	46.290	11.66	1.532	40.631	10.468	0.361	46.396	11.639
2	4.026	65.594	11.974	5.153	61.383	10.847	4.334	65.609	11.666	5.629	57.762	10.371	4.379	65.759	11.621
3	8.027	87.487	11.973	9.153	79.441	10.847	8.334	84.927	11.666	9.615	74.558	10.385	8.362	86.087	11.638
4	12.026	107.378	11.974	13.152	97.498	10.848	12.334	104.246	11.666	13.507	91.706	10.493	12.361	104.431	11.639
5	16.026	127.270	11.974	17.150	115.554	10.850	16.333	123.565	11.666	17.539	108.771	10.461	16.360	123.775	11.639
1 (mean shifts)	4.930	68.498	11.070	5.366	61.597	10.634	5.257	66.532	10.743	5.696	57.829	10.304	6.810	68.190	9.190
1 (Fourier-M)	0.029	47.705	11.971	1.162	43.335	10.838	0.335	46.291	11.665	1.494	40.593	10.506	0.364	46.399	11.635

(c)

	Phillip-Morris			Procter & Gamble		
Frequencies	AIC	BIC	(ℓ)	AIC	BIC	(ℓ)
1	0.074	47.750	11.926	0.043	47.718	11.957
2	4.077	67.645	11.923	4.041	67.240	11.959
3	8.076	87.537	11.924	8.041	87.039	11.959
4	12.077	107.428	11.923	12.041	106.839	11.959
5	16.077	127.320	11.924	16.041	126.639	11.959
1 (mean shifts)	4.611	68.179	11.389	4.962	68.161	11.038
1 (Fourier-M)	0.079	47.755	11.921	0.047	47.446	11.953

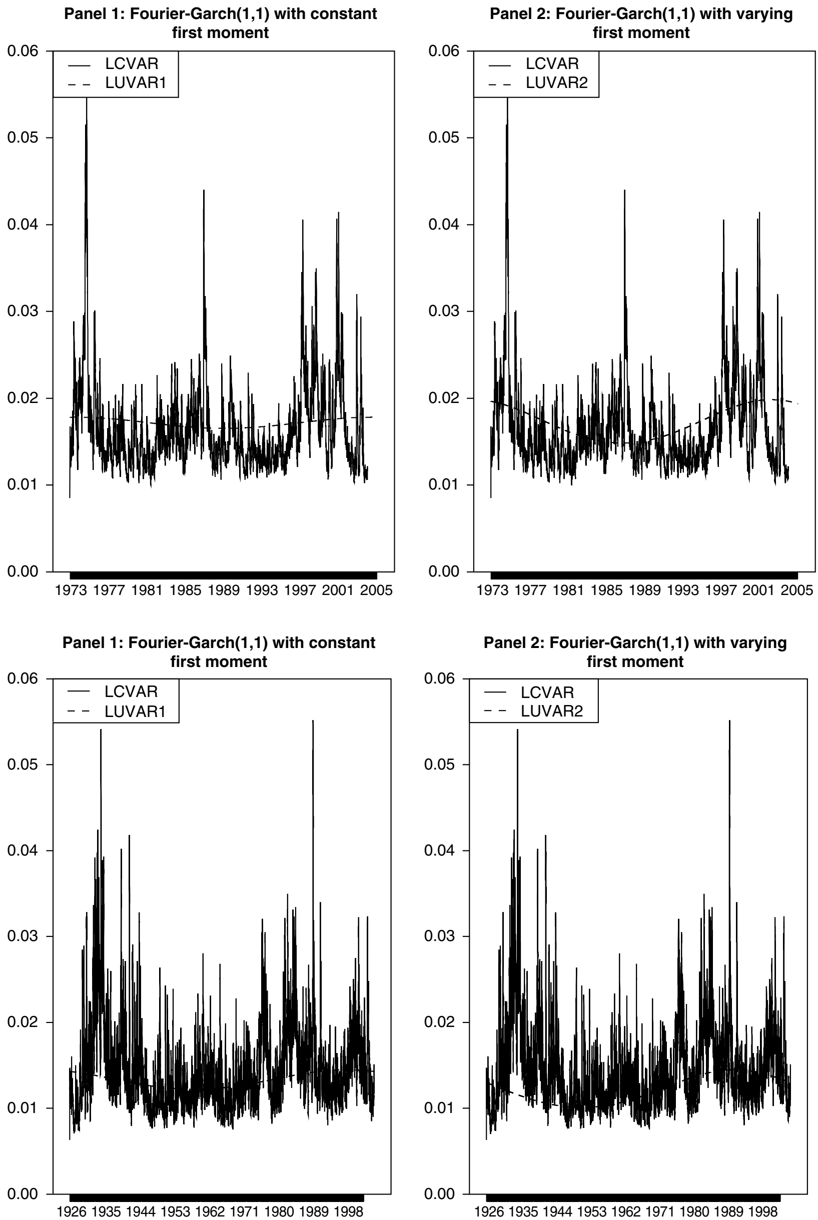


Figure 5.5 Conditional and unconditional volatility for AIG and Chevron

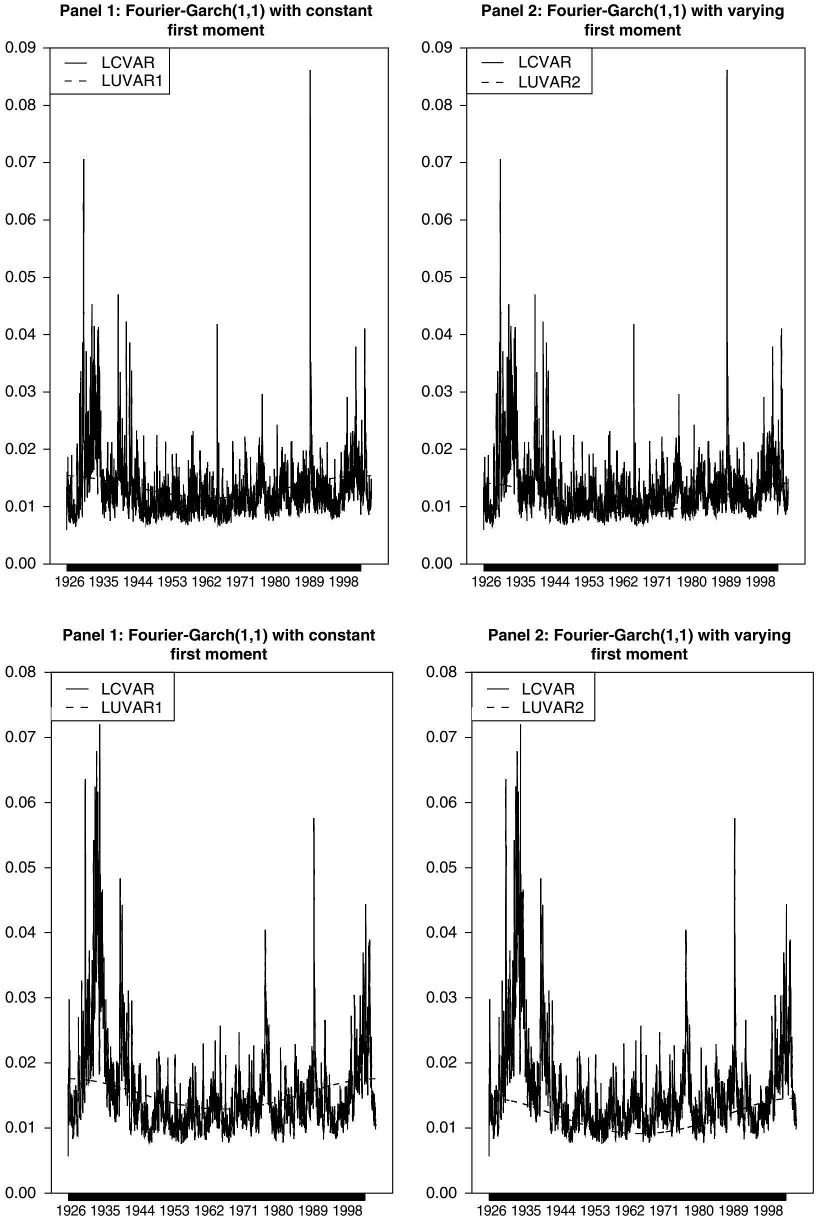


Figure 5.6 Conditional and unconditional volatility for Exxon and General Electric

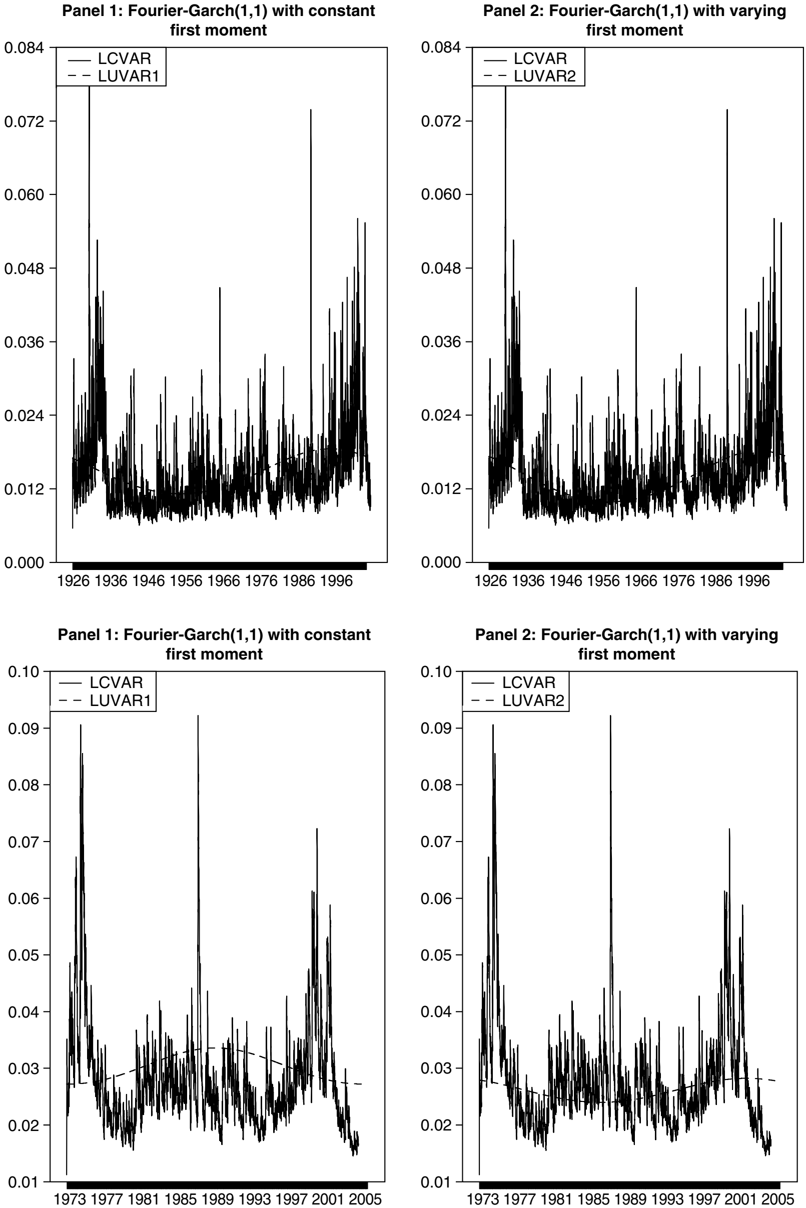


Figure 5.7 Conditional and unconditional volatility for IBM and Intel

Table 5.3 Persistence of financial volatility

Companies	M_0: GARCH(1,1)	M_1: Fourier-Garch(1,1) with constant mean	M_2: Fourier-Garch (1,1) with time-varying mean	M_3: Fourier-M (1,1)
	$\alpha+\beta$	$\alpha+\beta$	$\alpha+\beta$	$\alpha+\beta$
AIG	0.98024	0.98034	0.97753	0.96798
Chevron	0.98704	0.98373	0.75108	0.96577
Citigroup	1.00104	0.90388	0.57667	0.99360
Exxon	0.98333	0.95727	0.54307	0.95894
General Electric	0.99256	0.99013	0.80713	0.99261
IBM	0.99180	0.96291	0.51090	0.95930
Intel	0.99185	0.99206	0.64818	0.98175
Johnson & Johnson	0.95222	0.88250	0.01595	0.90133
Microsoft	0.06820	0.10247	0.22565	0.09550
Pfizer	0.97707	0.90421	0.40066	0.85240
Phillip-Morris	0.99877	0.99251	0.75108	0.98887
Procter & Gamble	0.99595	0.96786	0.29293	0.96698

Next, this chapter investigates whether the selected returns display the long-memory property that is usually observed in financial data. To this end, the study estimates four competing models:

1. the common GARCH(1,1) developed by denoted M_0;
2. the basic Fourier–GARCH(1,1) with constant first moment, denoted M_1;
3. the Fourier–GARCH(1,1) with a time-varying first moment, denoted M_2;
4. the Fourier-M (1,1) with long-run volatility in the mean, denoted M_3.

Table 5.3 shows the results. Clearly, model M_2 provides the best reduction of the persistence effect for most series. For 10 of the 12 stock returns considered, the long-memory effect is dramatically reduced in many instances by half or even more (i.e., General Electric, Pfizer, IBM, Philip Morris, Chevron, Intel, Procter & Gamble, Exxon, Johnson & Johnson, and Citigroup).

Note that the basic representation (i.e. the M_1 model above) has only little impact on overall persistence in the short-run volatility. In most cases, its persistence is only slightly lower than the one of the GARCH(1,1) representation. This is surprising given that this model gives the best fit according the AIC and BIC criteria in 10 out of the 12 stocks considered. Note that model M_3 clearly outperforms model M_1 in terms of reduced long-memory effect as well. The main conclusion is that allowing only for the second moment to vary over time is not enough to account for the strong persistence effect observed in financial returns. However, in contrast to the basic model, a time-varying first moment in the equation for the mean reduces significantly the persistence in short-run volatility.

5.5 Conclusion

This chapter proposes a new model to estimate the short- and long-run dynamics in financial data that takes into account the possibility of a time-varying first and second moment. The flexible Fourier transform of Gallant (1981) approximates the unknown date and shape of any structural break in the first and second moment as smooth processes. The study shows that Fourier series are able to approximate a wide variety of breaks of an unknown form. The basic Fourier–GARCH representation modifies the popular GARCH (1,1) to include a time-varying unconditional variance. This chapter proposes two extensions to the basic model. The first extension specifies a time-varying first moment while the second extension includes the long-run volatility in the equation for the mean. The results suggest that persistence still remains significant in the short-run volatility for the basic model. However, the so-called long-memory effect disappears if one includes a time-varying first moment in the model for the mean. This suggests that conditional volatility persistence is an artifact of the misspecification of the model for the mean.

References

Becker, R., Enders, W., and Hurn, S. (2001) "Modelling Structural Change in Money Demand Using a Fourier-Series Approximation," Research Paper Series 67, Quantitative Finance Research Centre, University of Technology, Sydney.

Becker, R., Enders, W., and Lee, J. (2006) "A Stationarity Test in the Presence of an Unknown Number of Smooth Breaks," *Journal of Time Series Analysis*, 27 (3): 381–409.

Bollerslev, T. (1987) "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," *The Review of Economics and Statistics*, 69 (3): 542–547.

Diebold, F. X. and Inoue, A. (2001) "Long Memory and Regime Switching," *Journal of Econometrics*, 105 (1): 131–159.

Enders, W. and Lee, J. (2006) "Testing for a Unit Root with a Nonlinear Fourier Function," Mimeo, University of Alabama, Tuscaloosa, Ala.
Engle, R. and Bollerslev, T. (1986) "Modeling the Persistence of Conditional Variances," *Econometric Reviews*, 5 (1): 1–50.
Engle, R. F. and Rangel, J. G. (2008) "The Spline-Garch Model for Low-Frequency Volatility and its Global Macroeconomic Causes," *Review of Financial Studies*, 21 (3): 1187–1222.
Gallant, R. (1981) "On the Basis in Flexible Functional Form and an Essentially Unbiased Form: The Flexible Fourier Form," *Journal of Econometrics*, 15 (2): 211–353.
Granger, C. W. J. (1980) "Long Memory Relationships and the Aggregation of Dynamic Models," *Journal of Econometrics*, 14 (2): 227–238.
Granger, C. W. J. and Joyeux, R. (1980) "An Introduction to Long Memory Time Series and Fractional Differencing," *Journal of Time Series Analysis*, 1 (1): 15–30.
Granger, C. W. J. and Joyeux, R. (1986) "Modeling the Persistence of the Conditional Variances: A Comment," *Econometric Reviews*, 5 (1): 51–56.
Kapetianos, G., Shin, Y., and Snell, A. (2003) "Testing for a Unit Root in the Nonlinear STAR Framework," *Journal of Econometrics*, 112 (2): 359–379.
Lamoureux, C. G. and Lastrapes, W. D. (1990) "Persistence in Variance, Structural Change, and the Garch Model," *Journal of Business and Economic Statistics*, 8 (2): 225–234.
Leybourne, S., Newbold, P., and Vougas, D. (1998) "Unit Roots and Smooth Transitions," *Journal of Time Series Analysis*, 19 (1): 83–97.
Maddala, G. and Kim, I.-M. (1998) *Unit Roots, Cointegration and Structural Change*, Cambridge: Cambridge University Press.
Mikosch, T. and Starica, C. (2004) "Non-stationarities in Financial Time Series, the Long Range Dependence and the IGARCH Effects," Econometrics 0412005, EconWPA.
Perron, P. and Qu, Z. (2007) "An Analytical Evaluation of the Log-Periodogram Estimate in the Presence of Level Shifts," Boston University – Economics Department, wp2007-044.
Starica, C. and Granger, C. (2005) "Nonstationarities in Stock Returns," *The Review of Economics and Statistics*, 87 (3): 503–522.

6

Essays in Nonlinear Financial Integration Modeling: The Philippine Stock Market Case

Mohamed El-Hedi Arouri and Fredj Jawadi

6.1 Introduction

Emerging stock markets are one of the best areas for investment and have become more accessible to investors in recent years thanks to successive efforts to open up these markets. These markets not only offer investors generous returns and opportunities but also enable them to better diversify their portfolios. These efforts recently led to a significant increase in capital flows toward the region and a rise in emerging market capitalization that reached around 20 percent of world market capitalization in 2000. This also has a considerable impact on the emerging stock market industry. Indeed, in addition to the significant increase in the financial integration of emerging markets into the world market (Bekaert and Harvey 1995), the adjustment dynamics of their asset prices is almost simultaneously governed by internal, regional, and external economic, financial, and political factors (Adler and Qi 2003; Carrieri et al. 2007).

In this study, we focus on emerging Asian markets. Indeed, a large number of these markets have launched a series of reforms, including their modernization and liberalization. Consequently, integration of Asian stock markets has emerged as an important body of literature (Bekaert and Harvey 1995; Gérard et al. 2003; Carrieri et al. 2007). However, the intensity and efficiency of these reforms and the degree of financial integration differ from one country to another. In addition, the internal and external factors pertaining to the financial markets in this region are also very different, suggesting perhaps multiple asymmetrical regimes of financial integration and segmentation which are interesting to apprehend and investigate.

The main contribution of this chapter is to investigate whether emerging Asian stock markets are integrated into the world market or not. We choose to focus our analysis on the Philippine case for diverse reasons. First, the Philippines is characterized by an overvalued exchange rate, a fragile banking system, and insufficient reserves with regard to the monetary mass (Sachs et al. 1996), suggesting that it may benefit considerably from further financial integration with the world market. Second, the financial markets in the Philippines have only recently, although continuously, been growing, and the Philippines' trade-openness ratio reached an average of 119 percent over the past decade. This is essentially due to the smooth functioning of the ASEAN (Association of South-East Asian Nations), created in 1965 by five countries (Indonesia, Malaysia, the Philippines, Singapore, and Thailand). Furthermore, in order to promote its integration into other international stock markets, the Philippine market underwent several reforms: liberalization and privatization (in 1985) and the introduction of ADR and country funds (in 1989). The Philippine stock market is thus expected to be better integrated during the post-liberalization period than it was during the period prior to the opening up of its market.

Several previous studies in the literature have focused on financial integration in Asian and Latin emerging stock markets. However, the authors' conclusions are not unanimous, and their results are often heterogeneous, perhaps because they define financial integration differently and test it also via different tools. For example, Bekaert and Harvey (1997) and Carrieri et al. (2007) studied Asian and Latin American emerging markets (the Philippines and other emerging countries) using a time-varying partially integrated CAPM. Their main conclusion is that the majority of emerging markets are partially integrated in the world market and that their degrees of integration are time varying. However, these results strongly depend on the validity of the CAPM.

Other studies focus on stock market integration in developed and emerging countries, using cointegration techniques. Masih and Masih (1997) show that the newly industrialized Asian countries of Honk Kong, Singapore, Taiwan, and South Korea share a long-term relationship with the developed markets (the USA, Japan, the UK, and Germany). More recently, Masih and Masih (2001) found significant interdependencies between the established OECD and the emerging Asian markets. Lim et al. (2003) examined the linkages between stock markets in the Asian region over the period 1988–2002 using nonparametric cointegration techniques and found a common force that brings these markets together in the long run. Similar results are suggested by Wang and Nguyen

Thi (2007) and Iwatsubo and Inagaki (2007). Ratanapakon and Sharma (2002) studied the short- and long-term relationships in five regional stock indices from the pre-Asian crisis and the crisis period and found that the degree of linkage increased during and after the crisis period. More recently, Phylaktis and Ravazzolo (2005) investigated stock market interactions of Pacific Basin countries over the period 1980–1998 and showed that although linkages have increased in recent years, there is still room for long-term gains when investing in Pacific Asian markets. By contrast, Bilson et al. (2000) show that the regional integration among stock markets in South Korea, Taiwan, Thailand, the Philippines, and Malaysia is faster than their integration within the international market. Roca and Selvanathan (2001) show neither short- nor long-term linkages among the stock markets of Australia, Hong Kong, Singapore, and Taiwan. Phylaktis and Ravazzolo (2000) also identify a lack of co-movements during the 1980s for the free stock markets of Singapore and Hong Kong. More interestingly, other recent studies show that the level to which markets are integrated or segmented is not fixed but changes gradually over time.

To sum up, this literature review shows some interdependencies between emerging and developed stock markets, suggesting further evidence of financial integration. However, it also suggests the difficulty of arbitraging between the two polar cases of strict segmentation and perfect integration. In fact, on the one hand, dynamic approaches show that the financial integration dynamic can be assimilated with a continuous process combining these two extreme cases as well as a continuum of intermediate states. This is even more valid for emerging stock markets which are generally characterized by ongoing liberalization processes.

On the other hand, analysis of the findings of previous studies shows that they define financial integration differently and that they have checked it using different methods. Indeed, they have often used linear modeling tools, even though some of them argue with the fact that the financial integration seems to be time-varying and has tended to increase over the past decade because of the rise in the number of international investors, the increase in new information and communication technologies and market liberalization. However, usual linear techniques yield invariant adjustment and are thus not usually suitable to reproduce dynamic and time-varying financial integration. Consequently, the linear framework used in most previous studies fails to capture certain types of financial integration which are time-varying, asymmetrical and nonlinear, and persistent and irregular.

Thus, in this chapter, we consider markets to be integrated if they share a common trend and move together. Using linear and nonlinear modeling, we study the stock price adjustment dynamic and financial integration of the Philippine market into the world market. In particular, we propose using nonlinear econometric tools given by nonlinear error correction models to investigate the Philippines' emerging stock market integration. Checking the hypothesis of financial integration within nonlinear modeling enables the integration dynamics to be asymmetrical, discontinuous, time-varying, and nonlinear. Moreover, the nonlinear cointegration methodology not only allows integration to be studied in a more general setting, taking into account the asymmetry, persistence, and nonlinearity that characterizes the dynamics of stock price adjustment, but also enables us to check and specify the degree of integration in every regime as well as in the short and long run. This is also interesting for a better understanding of the dynamism of financial markets and decision-making concerning international portfolio diversification in the Asian region.

The article is organized as follows. Section 6.2 will briefly present the econometric methodology. In Section 6.3, we will discuss the empirical results, and Section 6.4 will conclude.

6.2 Checking financial integration within nonlinear modeling

According to several stylized facts (financial crises and stock crashes) and some previous studies, the stock integration dynamics may be nonlinear. This nonlinearity can be differently explained by market imperfections: information and segmentation barriers (Bekaert and Harvey 1995), distinct transaction costs (Anderson 1997), heterogeneous shareholders' expectations (De Grauwe and Grimaldi 2006), etc. This implies an ongoing financial integration process and a nonlinear time-varying correcting mechanism which is adequately reproduced using the class of nonlinear cointegration models. Among the nonlinear cointegration models, we suggest using two nonlinear error correction models (NECM): the exponential switching transition error correction model (ESTECM) and the nonlinear error correction model-rational polynomial (NECM-RP) which we briefly present in the following section.[1]

Formally, let y_t and x_t be respectively the stock prices of the Philippines and the world market, where the long-run relationship corresponds to:

$$y_t = \alpha_0 + \alpha_1 x_t + z_t \tag{6.1}$$

where z_t designates the residuals of the long-run relationship.

The ESTECM is defined as follows:

$$\Delta y_t = \alpha_0 + \rho_1 z_{t-1} + \sum_{i=0}^{q} \beta_i \Delta x_{t-i} + \sum_{j=1}^{p} \delta_j \Delta y_{t-j} + \rho_2 z_{t-1} \times \left[1 - \exp\left\{-\gamma\left(z_{t-1} - c\right)^2\right\}\right] + \varepsilon_t \tag{6.2}$$

Where ρ_1, ρ_2, γ, and c are respectively the adjustment term in the first and the second regimes, the transition speed, and the threshold parameter.

The ESTECM enables the financial integration process to be different per regime, thus defining two extreme regimes when the exponential transition function nears the unity, and a central regime when it is equal to zero, as well as a continuum of intermediate states. The transition between these regimes is assumed to be smooth and gradual, and this specification is recommended to capture temporal paths governed by smooth changing regimes, accounting for a slow adjustment mechanism.

The NECM-RP is defined as follows:

$$\Delta y_t = \alpha_0 + \rho_1 z_{t-1} + \sum_{i=0}^{q} \beta_i \Delta x_{t-i} + \sum_{j=1}^{p} \delta_j \Delta y_{t-j} + \rho_2 \times \frac{\left(z_{t-1} + a\right)^3 + b}{\left(z_{t-1} + c\right)^2 + d} + \mu_t \tag{6.3}$$

Where: a, b, c, and d are the parameters of the rational polynomial function.

As suggested by Chaouachi et al. (2004), the NECM-RP is a more general nonlinear model which can take into account several potential sources of nonlinearities (i.e. abrupt changes in adjustment speeds, the impact of negative and positive shocks on stock price adjustment, multiple long-run attractors, etc.).

In practice, we carry these out to specification to examine the integration process of the Philippines' stock market in several steps. First, we apply the usual unit root tests (augmented Dickey–Fuller [ADF] and Phillips–Perron [PP] tests) to check the integration order of the stock price series. Second, we check the mixing hypothesis, applying KPSS and R/S tests on the residual term ($\hat{z}_t$) to test the nonlinear cointegration hypothesis. Third, accepting the mixing hypothesis suggests that stock prices are nonlinearly mean-reverting and allows our NECM to be estimated through the nonlinear least squares (NLS) method.

6.3 Empirical results

6.3.1 Data and preliminary analysis

Using monthly stock market indices from the Philippines and the world market over the period December 1987 to January 2008, obtained from Morgan Stanley Capital International (MSCI), which we express in US dollars, we first test the order of integration of these series by ADF and PP tests and show that both indices are I (1). Second, based on the matrix of bilateral correlation between the Philippines and the world-market indices, that we compute over two subperiods (January 1988–November 1994 and December 1994–January 2008) and over the period of study,[2] we show that the bilateral correlations between the Philippines and world stock markets are higher after 1994. This finding indicates that the Philippines' stock market has become more integrated in recent years (see Table 6.1).

Third, we test the symmetry, normality hypotheses, and our findings, presented in Table 6.2, suggest further evidence against normality and symmetry for the Philippines' stock returns since the Skewness, and Kurtosis coefficients are statistically significant.

This may be assimilated with a sign of nonlinearity in the dynamics of the Philippine stock market. We then estimate the relation (6.1) and we test for the presence of a unit root in the residuals (z_t) using the ADF test. The hypothesis of linear cointegration is not rejected, suggesting further evidence of integration between the Philippine and the world

Table 6.1 Matrix of bilateral correlations

Series	**RPHI**	**RMSCI**
Subperiod 1: January 1988–November 1994		
RPHI	1.00	0.33
RMSCI	0.33	1.00
Subperiod 2: December 1994, January 2008		
RPHI	1.00	0.44
RMSCI	0.44	1.00
All the period: January 1988–January 2008		
RPHI	1.00	0.40
RMSCI	0.40	1.00

Note: This table shows bilateral correlations between the stock returns of the world and the Philippines before and after the 1990s. RMSCI and RPHI are respectively the stock returns of the world and the Philippines.

Table 6.2 Descriptive statistics and normality test

Series	Mean	Std. Dev.	Maximum	Minimum	Skewness	Kurtosis	Jarque-Bera (Probability)
Philippines	0.0050	0.0927	0.3601	−0.3465	−0.0727	4.8155	33.31(0.0)
MSCI World Index	0.0053	0.0398	0.1055	−0.1444	−0.5733	3.8673	20.75(0.0)

Note: This table presents the descriptive statistics between the stock returns of the world and the Philippine stock markets.

stock markets. However, since several previously cited studies and our correlation analysis indicated that the degree of financial integration is time-varying, we propose testing this hypothesis using nonlinear cointegration tests which are more powerful than linear cointegration tests.

6.3.2 Nonlinear cointegration tests for financial stock market integration

To check for nonlinear cointegration between the Philippine and the world stock market and to test the hypothesis of time-varying financial integration, we apply two "mixing" tests which are more robust than the ADF test: the KPSS and the R/S tests. Both tests check the null hypothesis of "mixing" against its "nonmixing" alternative. For the KPSS, we retain the values suggested by Schwert (1989) for the truncation parameter: $l_4 = \text{int}\left[4\left(\frac{T}{100}\right)^{\frac{1}{4}}\right]$ and $l_{12} = \text{int}\left[12\left(\frac{T}{100}\right)^{\frac{1}{4}}\right]$ where T is the number of observations,[3] while we retain the value of Andrews (1991) concerning the choice of q for the R/S test which corresponds to the following formula: $q_t = [K_T]$, where $K_T = \left(\frac{3T}{2}\right)^{\frac{1}{3}}\left(\frac{2\hat{\rho}}{1-\hat{\rho}^2}\right)^{\frac{2}{3}}$, $[K_T] = \text{int}(K_T)$ and $\hat{\rho}$ is the first-order autocorrelation coefficient. We summarize the results obtained in Table 6.3. The null hypothesis of mixing is retained only at 10 percent according to the R/S test. According to the KPSS test, it is also accepted but only for the second value of the truncation parameter (l_{12}). Accepting the mixing hypothesis confirms the hypothesis of nonlinear cointegration and implies that the Philippine stock price is nonlinearly mean-reverting toward the world market at 10 percent (perhaps over the past decade). In a final step, we estimate both NECM: the ESTECM and the NECM-RP.

6.3.3 Estimation of NECM

We estimate both NECMs following the steps proposed by Escribano and Mira (2002) and Van Dijk *et al.* (2002). Firstly, we specify linear models

Table 6.3 Mixing tests

KPSS		R/S
l_4	l_{12}	Andrews
0.72*	0.29	1.6*

Note: This table presents the results of mixing tests.
* denotes the rejection of the null hypothesis at the 5 percent significance level.

Table 6.4 NECM estimation results for the Philippines

Coefficients	ESTECM (1,1)	NECM-RP
$\alpha 0$	−0.0021 (−0.37)	0.0009 (0.90)
ρ_1	−0.1815 (−1.02)	−0.0086 (−0.81)
ρ_2	0.1624 (0.91)	−0.0053* (−2.329)
β_0	0.2177* (3.55)	0.2086* (3.59)
δ_1	0.9228* (6.82)	0.9224* (6.83)
γ	625.07 (0.49)	–
$\gamma \times \sigma_{z_{t-d}}$	219.53	–
C	−0.3319* (−16.36)	–
ADF^{GLS}	−15.85	−15.85
R/S	2.5*	1.5
$\sigma_{NECM}/\sigma_{LECM}$	0.99	0.95

Note: This table presents the estimation results of NECM for the Philippines. The values in brackets are the t-statistic of nonlinear estimators.
* denotes the significance at 5 percent.

and determine the number of lags (p) for the NECMs on the basis of the information criteria and the autocorrelation functions. The optimal value retained is $p = 1$. We then estimate the NECMs by the NLS method and we report the results in Table 6.4. Our findings show several conclusions regarding the hypothesis of financial integration. Firstly, the current MSCI world index parameter is statistically significant, suggesting the presence of statistical dependence of the Philippine stock market on the world market (external factor). The first AR parameter is also significant, which suggests that its index depends on its past tendencies (local and internal factor). Secondly, our results show that ESTECM is not appropriate to the Philippines. Neither the exponential function parameters nor the nonlinear adjustment terms are statistically significant and

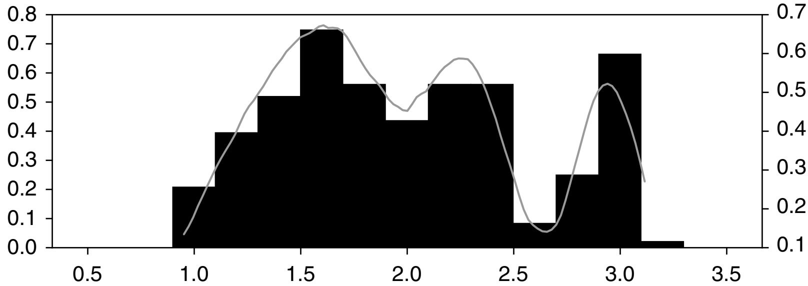


Figure 6.1 Histogram of the rational polynomial function for the Philippines

the residuals of the estimated model are not mixing. We therefore reject this nonlinear representation for the Philippines.

The dynamics of the Philippine stock market is better apprehended using the NECM-RP. This model, estimated under the following restrictions: $a = c = d = 1$ and $b = 0$, as in Chaouachi *et al.*, (2004) in order to simplify the algorithm convergence, seems more suited to capturing the type of asymmetry inherent to the Philippines' stock market. Overall, the estimation results of this model suggest significant correlation between the Philippine and the world stock markets.

More interestingly, from Figure 6.1, the persistence, asymmetry and smoothness associated with the Philippine stock price adjustment dynamics seem to be captured by the NECM-RP. Indeed, while the first adjustment term is negative but statistically non-significant, the second one is statistically significant and the sum of the two adjustment terms is also negative suggestion a nonlinear and asymmetric mean reversion in the stock price of the Philippines. More particularly, this means that in the first regime the Philippine stock price may deviate from the equilibrium and the stock market of the Philippine be segmented but a mean reversion is activated when stock price deviations exceed some threshold.

In order to highlight the pattern of nonlinear and asymmetric behavior characterizing this Asian stock market, we plot the histogram of the rational polynomial function in accordance with the estimated misalignment values ($\hat{z}_{t-1}$). This confirms the rejection of normality and linearity hypotheses. The NECM-RP also captures the asymmetry in the integration process between the emerging and world markets. Indeed, this figure shows a bimodal density and even several modes with two modes of unequal heights. The presence of these unequal modes suggests significant and extreme stock price deviations between the regimes of segmentation and integration. This asymmetry in the distribution of the rational polynomial function also reflects the persistence and smoothness of the integration process of this emerging stock market into the world market. This asymmetry and persistence in the pace of the Philippine stock market integration may perhaps be explained by the different theoretical arguments discussed in the body of this chapter. Finally, our findings are validated via misspecification tests that highlight that the residuals of NECM-RP are mixing and stationary.

6.4 Conclusion

We investigated the hypothesis of time-varying financial integration between the Philippine and the world stock markets over three decades in a nonlinear framework. Our findings suggest further evidence of asymmetrical and nonlinear cointegration between these markets. They confirm the hypothesis of time-varying financial integration for the Philippines and highlight the contribution of nonlinear error correction models in studying this hypothesis. Indeed, these tools enable the extreme cases of financial integration to be reproduced (perfect integration and strict segmentation) as well as a continuum of intermediate states relative to partial integration that characterizes most emerging stock markets. This study may be extended by testing this approach for other emerging and developed stock markets.

Notes

1. The NECM-RP methodology is based on the theorem of Escribano and Mira (2002), whereas that of the ESTECM is developed by Van Dijk et al. (2002), who adapt the methodology to the threshold models, thus defining a particular kind of threshold cointegration model for which the adjustment is relatively smooth and asymmetric. For more details regarding nonlinear cointegration models, see Dufrénot and Mignon (2002).
2. See table1 in the appendices in which we also present all the empirical results.
3. Int [.] denotes the interior part.

References

Adler, M. and Qi, R. (2003) "Mexico's Integration into the North American Capital Market," *Emerging Economic Review*, 4 (2): 91–120.

Anderson, H. M. (1997) "Transaction Costs and Nonlinear Adjustment towards Equilibrium in the US Treasury Bill Markets," *Oxford Bulletin of Economics and Statistics*, 59 (4): 465–484.

Andrews, D. (1991) "Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59 (3): 817–858.

Beakert, G. and Harvey, C. (1997) "Emerging Equity Market Volatility," *Journal of Financial Economics*, 43 (1): 29–77.

Bekaert, G. and Harvey, C. (1995) "Time Varying World Market Integration," *Journal of Finance*, 50 (2): 403–444.

Bekaert, G., Harvey, C., and Ng, A. (2005) "Market Integration and Contagion," *Journal of Business*, 78 (1): 39–69.

Bilson, C., Hooper, V., and Jaugietis, M. (2000) "The Impact of Liberalisation and Regionalism upon Capital Markets in Emerging Asian Economies," *International Finance Review*, 1 (1): 219–255.

Carrieri, F., Errunza, V., and Hogan, K. (2007) "Characterizing World Market Integration through Time," *Journal of Financial and Quantitative Analysis*, 41 (2): 511–540.

Chaouachi, S., Dufrénot, G., and Mignon V. (2004) "Modelling the Misalignments of the Dollar-Sterling Real Exchange Rate: A Nonlinear Cointegration Perspective," *Economics Bulletin*, 3 (19): 1–11.

De Grauwe, P. and Grimaldi, M. (2006) "Heterogeneity of Agents, Transaction Costs and the Exchange Rate," *Journal of Economic Dynamics and Control*, 29 (4): 691–719.

Dufrénot, G. and Mignon, V. (2002) *Recent Developments in Nonlinear Cointegration with Applications in Macroeconomics and Finance,* Boston, Mass.: Kluwer.

Escribano, A. and Mira, S. (2002) "Nonlinear Error Correction Models," *Journal of Time Series Analysis*, 23 (1): 509–522.

Gerard, B., Thanyalakpark, K., and Batten, J. (2003) "Are the East Asian Markets Integrated? Evidence from the ICAPM," *Journal of Economics and Business*, 55 (2): 585–607.

Iwatsubo, K., and Inagaki, K. (2007) "Measuring Financial Market Contagion Using Dually-Traded Stocks of Asian Firms," CEI Working Paper Series No. 14, Institute of Economic Research, Hitotsubashi University, Tokyo, Japan.

Lim, K., Lee, H., and Liew, K. (2003) "International Diversification Benefits in Asian Stock Markets: A Revisit," *Mimeo*, Lebuan School of International Business and Finance, Sabah, Malaysia.

Lo, A. W. (1991) "Long-Term Memory in Stock Market Prices," *Econometrica*, 59 (5): 1279–1313.

Masih, A. and Masih, R. (1997) "A Comparative Analysis of the Propagation of Stock Market Fluctuations in Alternative Models of Dynamic Causal Linkages," *Applied Financial Economics*, 7 (1): 59–74.

Masih, A. and Masih, R. (2001) "Long and Short Term Dynamic Causal Transmission amongst International Stock Markets," *Journal of International Money and Finance*, 20 (4): 563–587.

Phylaktis, K. and Ravazzolo, F. (2000) "Stock Prices and Exchange Rate Dynamics," *Mimeo*, City University Business School, London.

Phylaktis, K. and Ravazzolo, F. (2005) "Stock Market Linkages in Emerging Markets: Implication for International Portfolio Diversification," *Journal of International Markets and Institutions*, 15 (2): 91–106.

Ratanapakon, O. and Sharma, S. (2002) "Interrelationships among Regional Stock Indices," *Review of Financial Economics*, 11 (1): 91–108.

Roca, E. and Selvanathan, E. (2001) "Australian and the Three Little Dragons: Are Their Equity Markets Interdependent?" *Applied Economic Letters*, 8 (3): 203–207.

Sachs, J., Tornell, A., and Velasco, A. (1996) "Financial Crises in Emerging Markets: The Lessons from 1995," *Brookings Papers on Economic Activity*, 1: 147–198.
Schwert, G. W. (1989) "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business and Economic Statistics*, 7 (2): 147–159.
Van Dijk, D., Teräsvirta, T., and Franses, P. H. (2002) "Smooth Transition Autoregressive Models: A Survey of Recent Developments," *Econometric Reviews*, 21 (1): 1–47.
Wang, K. M. and Nguyen Thi, T. B. (2007) "Testing for Contagion under Asymmetric Dynamics: Evidence from the Stock Markets between US and Taiwan," *Physica A*, 376 (15): 422–432.

Part II

Term Structure Models

7
A Macroeconomic Analysis of the Latent Factors of the Yield Curve: Curvature and Real Activity

Matteo Modena

7.1 Introduction

Examining the relation between yields at different maturities is crucial for both macroeconomists and financial economists. From a macroeconomic perspective, the short rate is the policy instrument under the control of the monetary authority; however, from a financial perspective, movements in short-term rates are analyzed to forecast longer yields' dynamics, since yields on long-term bonds are the expected average of risk-adjusted future spot rates. Moreover, the dynamics of the term structure (TS) is influenced both by monetary policy actions and by expectations about policy announcements; while, on the other hand, economists infer the future path of macro variables from different shapes of the yield curve.

Including macro variables in TS models is a quite recent phenomenon; in this chapter we focus on the interpretation of curvature which has been mostly ignored by previous analysis.

We consider the US bond market between March 1987 and December 2007, thus focusing on a sample characterized by price stability and a relatively homogeneous monetary regime: explicit interest-rate targeting. Data evidence suggests that almost all TS movements can be summarized by few underlying factors, namely level, slope, and curvature. The terminology refers to the effect that a shock to these unobservable variables exerts on the shape of the yield curve (Litterman and Scheinkman 1991). When interest rates of all maturities change by the same amount, the yield curve is subject to a level shock; hence, a perturbation of this kind causes a parallel shift of the entire yield curve. A shock to the slope exerts different intensity on the maturity spectrum of interest rates. A positive slope shock decreases short rates more than long rates, enlarging the spread and steepening the yield curve. Finally, a positive curvature shock increases yields at medium-term maturities, impressing a more accentuated hump-shaped form to the yield curve. For its peculiar effect on the maturity field, curvature has been labeled the *butterfly* factor.

An important strand of literature has recently focused on the macroeconomic interpretation of these factors. The existing empirical literature associates the level to some measures of inflation. Rudebusch and Wu (2004), as well as Bekaert et al. (2005), suggest level reflecting the inflation rate targeted by the monetary authority, also known as the long-run equilibrium inflation rate. Dewachter et al. (2006) provide evidence that the level is an indicator of the central tendency of inflation. There is also general consensus about the interpretation of the slope, which is believed to be a monetary policy factor. Rudebusch and Wu (2004) provide evidence that the slope factor tracks a fitted Taylor-type monetary policy rule; Bekaert et al. (2005) relate the slope to monetary policy shocks. A negative slope shock reduces the spread flattening TS, that is, what generally occurs when monetary policy tightens. More controversial seems to be the interpretation of curvature. It has been argued that curvature is either related to monetary policy shocks (Bekaert et al. 2005) or to the real stance of monetary policy[1] (Dewachter et al. 2006), or, eventually, to the expected future path of interest rates (Giese 2008). Finally, Hordahl et al. (2006) emphasize the effect of both inflation and output shocks on medium-term maturities of the yield curve.

The empirical analysis worked out in this chapter finds its inspiration in the diagrams of Figure 7.1, where grey shaded areas highlight National Bureau of Economic Research recessions. The left panel plots the level factor together with the CPI inflation rate and the slope factor with the effective federal funds rate. The constant decline of the level might be due either to the augmented credibility of the monetary regime or to the consolidation of the monetary authority's reputation over time, or to both.

The central and the left panels of Figure 7.1 plot curvature together with different measures of the business cycle. The series appear to display important co-movements; in particular, a visual inspection suggests curvature dropping during slowdown in economic activity. Curvature is positively related to the growth of industrial production (henceforth IP), the Hodrick-Prescott filtered series of *log* IP, the *log* total capacity utilization, the 6-month lagged growth rate in real consumption expenditures. Finally, the dynamics of curvature is inversely related to that of unemployment.

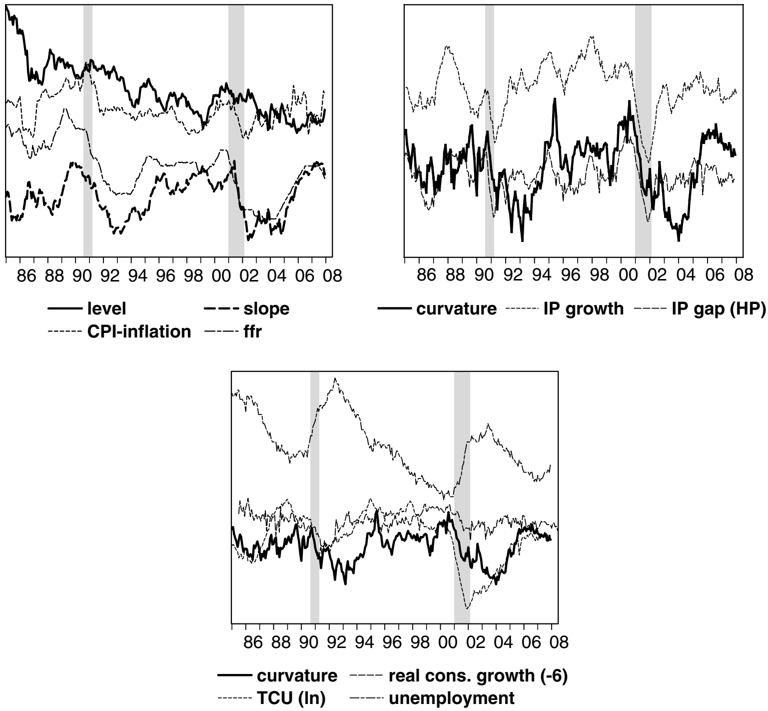


Figure 7.1 Curvature and macroeconomic variables
Notes: Level: level factor of the TS; slope: slope factor of the TS; CPI inflation: consumer price index rate of change; ffr: effective federal funds rate; curvature: curvature factor of the TS; IP growth: industrial production rate of growth; IP gap (HP): Hodrick–Prescott filter gap of IP; real cons. growth: real consumption expenditure rate of growth; TCU (ln): log series of total capacity utilization; unemployment: unemployment rate.

In this chapter we provide evidence supporting the interpretation of curvature as a cyclical indicator. We analyze curvature obtained from both nominal and real TS. It has been argued (Harvey 1988; Chapman 1997) that there exists a significant relationship between real TS of interest rates and consumption growth. Harvey (1988) provides evidence that the expected real TS helps predict consumption growth. We find evidence which is consistent with this story. In particular, despite the fact that curvature from the nominal TS seems unrelated to consumption growth, we find a significant inverse correlation between consumption growth and curvature from real TS.

The rest of this chapter is organized as follows. Section 7.2 presents a brief review of the literature. Data are presented in Section 7.3. In Section 7.4 we outline the Nelson–Siegel latent factor model. The core of the empirical analysis is contained in Sections 7.5, 7.6, and 7.7, where we provide evidence to support the macro-interpretation of curvature. In particular, Section 7.5 shows that curvature reflects the cyclical behavior of the economy. In Section 7.6 we estimate a cyclical model for curvature, while in Section 7.7 we develop and estimate a joint macroeconometric model for curvature and real activity. Section 7.8 concludes.

7.2 Literature review

Arbitrage-free affine TS models have been largely adopted in the literature to examine yield curve dynamics. Affine models are appealing since they summarize TS information in few state variables, or latent factors, given that most of TS movements are due to the effect of few components. Dai and Singleton (2000) show that 99 percent of the variations in the yield curve can be attributed to three factors. A second important group of TS models is the Nelson–Siegel class, where yields are assumed to be a function of factors through Laguerre functions of their maturities.

In a seminal article, Ang and Piazzesi (2003) show that incorporating macro factors into TS models improves the ability to forecast yields' movements both in- and out-of-sample. Approximately 85 percent of bond yields variation is attributable to the impact of macro factors; in particular, macro variables explain movements at short- and medium-term maturities (up to *one* year), while, movements of long-term yields depends upon the effect of financial factors. Hordahl et al. (2006) suggest that inflationary shocks mostly affect yields at medium maturities increasing the TS curvature. Monetary policy shocks, instead, seem to affect the short end of the yield curve; risk premia tend to respond to output gap shocks. Rudebusch and Wu (2004) focus on only two latent factors. The level turns put to be associated with the central bank's long-run inflation target, while the slope reflects the central bank's reaction function à la Taylor (1993).

Dewachter et al. (2004), using a continuous-time affine TS model find that macroeconomic variables are not capable of explaining movements at the long end of the yield curve. The variability of the long-term yields is related to the central tendency of inflation. Medium-term interest rates, from *six* months to *two* years, appear to respond to the real rate central tendency; both observable and unobservable components influence risk premia and bond excess returns.

Diebold and Li (2003), Diebold et al. (2005), and Diebold et al. (2006) employ the Nelson–Siegel interpolant to examine bond pricing. All these studies document a good forecasting performance of the Nelson–Siegel method. The most recent strand of literature has mixed the TS model with modern macroeconomic theory, including latent factor dynamics into New Keynesian general equilibrium frameworks. Bekaert et al. (2005) develop a model combining structural New Keynesian macroeconomics and no-arbitrage TS theory. This line of research has been followed by Wachter (2006) and Garcia and Luger (2007) who consider a consumption-based equilibrium macro-finance model.

7.3 Data

All data have monthly frequency, from March 1987 to December 2007. US yields data between March 1987 and December 1998 are from both the McCulloch–Kown database (three, six, and 120 months) and from the Fama–Bliss dataset (one, two, three, four, and five years). After January 1999, all yields are from Datastream (ZCB yields). The effective federal funds rate is from the Federal Reserve Economic Data (FRED) database. Table 7.1 reports some descriptive statistics of yields. The mean is increasing with maturity suggesting a positive liquidity, or risk, premium. The standard deviation tends to be large at short to medium maturities. Long-term yields are more persistent than short yields. The Jarque and Bera suggest short-term yields being normally distributed around the mean. Autocorrelations decay fast; the partial autocorrelation function suggests the first-order autoregressive structure of yields. AR(1) regressions for each yield return coefficients of approximately 0.98; however,

Table 7.1 Descriptive statistics of yields

	Yields							
maturity	**ffr**	**3**	**6**	**12**	**24**	**36**	**60**	**120**
mean	5.323	5.173	5.324	5.593	5.963	6.232	6.575	7.054
std dev	2.395	2.152	2.189	2.225	2.179	2.102	2.008	1.885
skew	0.088	0.009	0.049	0.129	0.391	0.556	0.825	1.004
kurt	2.598	2.720	2.871	3.030	3.377	3.542	3.875	4.041
JB norm	(0.323)	(0.631)	(0.856)	(0.668)	(0.012)	(0.000)	(0.000)	(0.000)
ADF	(0.141)**	(0.082)**	(0.093)**	(0.064)**	(0.051)**	(0.051)**	(0.040)**	(0.021)**
KPSS	0.096**	0.089**	0.093**	0.095**	0.105**	0.115**	0.140**	0.166**

Notes: Normality and ADF tests: *p*-values in parenthesis. Exogenous included: **Intercept and trend.

the Wald test rejects the null of unity coefficient. Both the augmented Dickey–Fuller (ADF) and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) suggest the series are stationary.[2]

Inflation is the annual change of the seasonally adjusted (SA) Consumer Price Index for all urban consumers (FRED, Bureau of Labour Statistics). M1 is the SA money stock from FRED (Federal Board of Governors). The monthly SA series of IP is from FRED. Different measures of the output gap have been generated: the growth rate of *log* IP; the Hodrick–Prescott filtered *log* IP; the Baxter–King and the Christiano–Fitzgerald cyclical component of *log* IP. All cyclical indicators are highly correlated. The SA civilian unemployment rate series is from FRED (Bureau of Labour Statistics) as well as the SA real personal consumption expenditures (Bureau of Economic Analysis).

7.4 The Nelson–Siegel factor model

The factor model is based on the approach pioneered by Nelson and Siegel (1987). Their method has become increasingly popular among financial economists for its relatively simple tractability and the fairly good fit. The yield on a bond with maturity n is a polynomial function of maturity:

$$y^{(n)} = L_t + S_t \left(\frac{1 - \exp\{-\lambda n\}}{\lambda n} \right) + C_t \left(\frac{1 - \exp\{-\lambda n\}}{\lambda n} - \exp\{-\lambda n\} \right) \quad (7.1)$$

Parameter λ governs the exponential decay.[3] The first loading is a constant, that is the unity coefficient multiplying L_t. The second loading is an exponential function that starts at *one* and decays monotonically toward *zero*. Finally, the third loading starts at *zero*, increases with maturity n and then gradually decays approaching *zero*. The path followed by these loadings allows interpreting them as level, slope, and curvature, respectively.

Factors are stacked in the state vector F_t which is assumed to follow a first-order VAR process. Differently from standard assumptions in canonical affine TS models (Dai and Singleton 2000), we do not restrict the transition matrix (Φ_{NS}) to be lower triangular. Hence, we allow the actual value of each factor to depend on the first lag of all other factors. We consider nine yields with maturities at one, three, six, 12, 24, 36, 48, 60, and 120 months, offering a dense representation of the maturity spectrum domain. The transition equation of the state-space representation is:

$$F_t = \mu_F + \Phi_{NS} \cdot F_{t-1} + \omega_{t,F} \quad (7.2)$$

ω_{t,F_0} is *i.i.d.* Normal with *zero* mean and diagonal covariance matrix (Ω). The initial state vector F_0 is orthogonal to the disturbances ω_{t,F_0} of the transition equation. The observation equation is:

$$y_t = \aleph F_t + \nu_t \tag{7.3}$$

The noise term is *i.i.d.* Normal with *zero* mean and variance σ^2. The white noise transition disturbances ν_t are orthogonal to the initial state vector F_0. The measurement equation is:

$$\begin{bmatrix} y_t^{(1)} \\ y_t^{(3)} \\ \vdots \\ y_t^{(120)} \end{bmatrix} = \begin{bmatrix} 1 & \frac{1-\exp(-\lambda)}{\lambda} & \frac{1-\exp(-\lambda)}{\lambda} - \exp(-\lambda) \\ 1 & \frac{1-\exp(-3\lambda)}{3\lambda} & \frac{1-\exp(-3\lambda)}{3\lambda} - \exp(-3\lambda) \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-\exp(-120\lambda)}{120\lambda} & \frac{1-\exp(-120\lambda)}{120\lambda} - \exp(-120\cdot\lambda) \end{bmatrix} \cdot \begin{bmatrix} L_t \\ S_t \\ C_t \end{bmatrix} + \begin{bmatrix} \nu_{t,1} \\ \nu_{t,3} \\ \vdots \\ \nu_{t,120} \end{bmatrix} \tag{7.4}$$

Estimations suggest that the cross-factor dynamics is weak. The first autoregressive coefficients of the latent factors from the nominal TS are 0.98, 0.92, and 0.92 for L_t, S_t, and C_t respectively. The most persistent factor becomes curvature when estimating the real TS; the autoregressive coefficients of L_t, S_t, and C_t from the real TS are 0.93, 0.92, and 0.96 respectively.

7.5 Curvature and business cycle fluctuations

A better understanding of TS dynamics can be achieved by exploring the macroeconomic underpinning of the yield curve. A certain consensus exists on the interpretation of the level and the slope; however, so far the interpretation of curvature is still controversial. In this section we provide evidence suggesting that curvature is related to the fluctuations of real activity. A curvature shock affects medium-term maturities of the yield curve. A positive shock increases medium-term yields, while a negative shock generates an inverted hump-shaped yield curve, which is sometimes observed in the data. Figure 7.1 shows that a significant negative curvature shock hits the yield curve each quarter preceding NBER recessions. We point out that the relationship between curvature and the economic cycle is not in contrast with previous evidence suggesting that the TS slope is a good predictor of future economic activity (Stock and Watson 1989; Estrella and Hardouvelis 1991).

Intermediate maturities have been largely ignored so far; thus, we try to achieve a more general and comprehensive examination of TS dynamics. Any shock affecting the short end of the yield curve, typically a monetary policy shock, generates only a moderate and delayed reaction of long yields, which are smooth and persistent, so that any shock affecting the short end of TS can be considered a shock to the slope. Our interest in medium maturities arises from the idea that medium-term yields represent an important link between the extremely dynamic short end and the smooth long end of TS. In particular, we argue that the propagation of shocks from the short to the long end of the yield curve reflects the evolution of economic conditions over the business cycle.

The empirical macro-finance literature has proposed different theoretical measures of curvature.[4] The correlation coefficients of these components are positive but not always as high as expected. Curvature is positively correlated to IP growth computed over different horizons, from a quarter to three years. Curvature shares also important co-movements with different measures of the output gap. Curvature is inversely correlated with the variation of unemployment over time (Figure 7.2).

A sharp reduction of curvature occurs immediately before economic slowdowns. Data evidence seems to support the conjecture that curvature is informative beyond the slope about business cycle fluctuations. We speculate that negative shocks to curvature seem either to anticipate or to accompany a decline in economic activity. Moreover, available empirical evidence is consistent with the idea that the curvature effect complements the transition from an upward-sloping to a flat yield curve.

It has been argued that the curvature factor is either related to monetary policy shocks (Cho et al. 2005), or to the real stance of monetary policy (Dewachter et al. 2006), or again to the expected future path of interest rates (Giese 2007). In the following analysis we show that curvature (C_t) is more closely related to the condition of the real economy than to monetary variables. We thus estimate two different equations. The monetary model (M) is:

$$C_t = \delta_0 + \delta_1 \Delta ffr_{t,t-12} + \delta_2 \Delta M1_{t,t-12} + \delta_3 \pi_t + \varepsilon_{t,M} \qquad (7.5)$$

$\Delta ffr_{t,t-12}$ is the annual change in the fed funds; $\Delta M1_{t,t-12}$ is the annual rate of growth of M1; π_t represents CPI inflation. The real-variable model

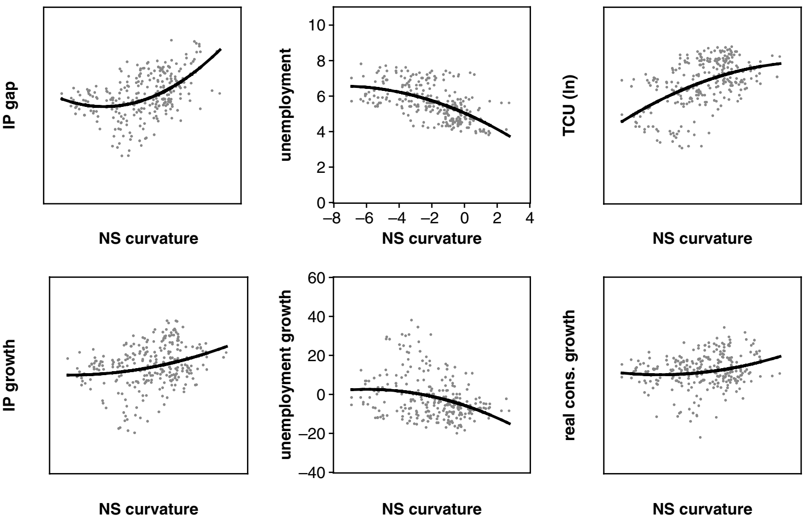


Figure 7.2 Curvature and macroeconomic variables: scatter plots
Notes: NS curvature: curvature obtained with the Nelson–Siegel method; IP gap: industrial production gap; unemployment: unemployment rate; TCU (ln): log series of the total capacity utilization; IP growth: industrial production growth; unemployment growth: rate of growth of the unemployment rate; real cons. growth: real consumption expenditure rate of growth.

(R) relates curvature to some cyclical indicators:

$$C_t = \rho_0 + \rho_1 \Delta IP_{t,t-12} + \rho_2 \Delta rc_{t,t-12} + \rho_3 \Delta un_{t,t-12} + \rho_4 gap_{t-1} + \varepsilon_{t,R} \tag{7.6}$$

$\Delta IP_{t,t-12}$ is the annual change of the seasonally adjusted IP; $\Delta rc_{t,t-12}$ represents the annual change in the real personal consumption expenditures; $\Delta un_{t,t-12}$ is the annual variation in unemployment; gap_t is either the Hodrick–Prescott or the Baxter–King de-trended series of *log* IP. Estimation results are reported in Table 7.2.

To show our results are robust, the above equations have been estimated both by OLS and by IV.[5] OLS estimations have been performed allowing different structures of the variance–covariance matrix of parameter estimates.[6] Real consumption seems weakly related to curvature; however, as we show later, consumption growth is significantly related to curvature from the real TS. Evidence is in line with the idea that the shape of the yield curve changes over the business cycle. We remark that

Table 7.2 Estimates of equations (7.5) and (7.6)

	Curvature – Nominal TS										
	Model R – equation (7.6)						**Model M – equation (7.5)**				
	ρ_0	ρ_1	ρ_2	ρ_3	ρ_4	χ^2	δ_0	δ_1	δ_2	δ_3	χ^2
OLS	−1.6088	0.7693	−0.2423	−0.1015	0.4909		−2.5250	0.8618	−0.1910	0.3713	
WH	(0.236)	(0.040)	(0.091)	(0.010)	(0.106)	55.54	(0.309)	(0.279)	(0.021)	(0.088)	108.1
	[−6.7]	[2.1]	[−2.7]	[−11]	[4.5]		[−8.2]	[3.1]	[−8.7]	[4.2]	
HH (12)	(0.345)	(0.846)	(0.190)	(0.018)	(0.265)	41.32	(0.947)	(0.672)	(0.041)	(0.230)	43.05
	[−4.6]	[0.9]	[−1.3]	[−5.4]	[1.8]		[−2.6]	[1.3]	[−4.6]	[1.6]	
NW (18)	(0.352)	(0.815)	(0.179)	(0.018)	(0.239)	35.05	(0.842)	(0.630)	(0.038)	(0.209)	48.07
	[−4.6]	[0.9]	[−1.3]	[−5.6]	[2.0]		[−2.9]	[1.3]	[−5.0]	[1.7]	
s-HH	(0.827)	(0.132)	(0.266)	(0.026)	(0.268)	24.88	(1.066)	(0.737)	(0.054)	(0.309)	20.31
	[−1.9]	[0.6]	[−0.9]	[−3.8]	[1.8]		[−2.5]	[1.2]	[−3.5]	[1.2]	
	R^2_{adj}	0.48					R^2_{adj}	0.38			
	ρ_0	ρ_1	ρ_2	ρ_3	ρ_4	χ^2	δ_0	δ_1	δ_2	δ_3	χ^2
IV	−1.2650	0.7926	−0.3595	−0.1148	0.5345		−2.4316	0.7526	−0.2570	0.4067	
	(0.458)	(0.816)	(0.194)	(0.017)	(0.221)	38.45	(0.568)	(0.498)	(0.036)	(0.155)	42.17
	[−2.7]	[−1.0]	[−1.8]	[−6.5]	[2.4]		[−4.2]	[1.5]	[−6.9]	[2.6]	
	R^2_{adj}	0.47					R^2_{adj}	0.35			

Notes: Standard error in parenthesis; *t*-statistics in square brackets.

the curvature effect is not incompatible with the fact that also the TS slope varies across the business cycle.

A flat yield curve is usually interpreted as a sign of imminent recession since relative high short- to long-term rates are assumed to reflect a severe monetary policy stance (Bernanke and Blinder 1992). On the contrary, an upward-sloping yield curve reflects expectations about an accommodative monetary policy and, thus, is indicative of a thriving economy. Suppose the economy is growing fast so that strong aggregate demand is likely to generate inflationary pressures. Suppose further the monetary authority raises interest rates to preserve price stability. Two effects may follow. On the one hand, the yield spread shrinks, since short yields are likely to increase by a larger amount than long-term yields; on the other hand, aggregate demand weakens, following the reduction of private investments. The adjustment process of the long end of the yield curve following the shock affecting the short end implies an intermediate step occurring at medium-term maturities, where the corresponding yields rise by more than long-term ones. The propagation along the entire spectrum of TS maturities generates a temporary spike in the medium end of the yield curve. Therefore, both the dynamics of the yield curve and the evolution of the macroeconomic conditions occur at the same time. Expectations may either accelerate or anticipate the process. The contrary happens before a recession. Expectations of accommodative monetary policy exert a negative pressure on TS medium maturities thus causing curvature to drop (Figure 7.1). In this chapter we do not intend to establish any causality relation between curvature and real economy, we simply suggest that curvature reflects the cyclical behavior of the economy. In short, the curvature effect seems to accompany the transition of the yield curve from a positively sloped one, prevailing during booms, to a flat one, which is believed to anticipate recessions.

Estimations of the monetary equation (7.5) suggest curvature being significantly related to the annual change of both the federal funds and M1. Hence, evidence does not entirely reject the hypothesis that curvature is related to monetary policy shocks. However, we provide evidence that the link between curvature and the real economy is stronger than the link between curvature and monetary policy variables.[7] In line with Rudebusch and Wu (2004, 2005), we suggest that the TS slope is more closely related to a Taylor-type monetary policy reaction function.

Since we believe that curvature reflects the state of the economy rather than the monetary conditions, we forecast curvature using both models. According to our results (Figure 7.3) model R, rather than model

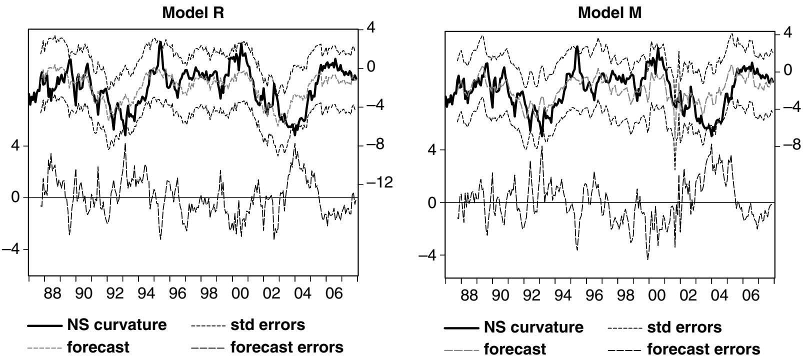


Figure 7.3 Forecasting performance of models (7.5) and (7.6)
Notes: NS curvature: curvature obtained with the Nelson–Siegel method; std errors: standard errors; forecast: forecast of the curvature factor; forecast error: difference between actual and predicted curvature.

M, returns a better fit. Both models return quite accurate forecasts, but standard errors of model R are lower. In addition, forecast errors of model R oscillate closer around the *zero* line. A battery of tests is performed to prove that the predictive accuracy of model (R) is better than that of model (M): Theil inequality, Diebold-Mariano, Morgan-Granger-Newbold, Wilcoxon.

As a further robustness check, we examine whether also curvature obtained from the real TS is related to the real economy. Removing the effect of inflation from TS should not affect the curvature factor, since inflation is mainly reflected in the long end of the yield curve. Ang et al. (2008) point out that the real TS does not exhibit any trend, while the typical upward-sloping shape of the nominal yield curve is due to a positive inflation risk premium which is incorporated in long-term yields. Therefore, the medium part of the yield curve should be unaffected after removing the effect of inflation. Hence, curvature extracted from the real TS should track curvature obtained from the nominal TS.[8] The correlation coefficient between real TS curvature and nominal TS curvature is very high indeed, about 0.95.

After ruling out inflation, we re-estimate both the monetary and the real equations for the real TS curvature. Curvature is still significantly explained by real variables, as reported in Table 7.3. Real TS curvature is also significantly related to consumption growth; such result can be interpreted consistently with previous findings in the literature (Harvey 1988; Chapman 1997).

Table 7.3 Estimates of equations (7.5) and (7.6) for the real TS curvature

	Curvature – Real TS										
	Model R – equation (7.6)						Model M – equation (7.5)				
	ρ_0	ρ_1	ρ_2	ρ_3	ρ_4	χ^2	δ_0	δ_1	δ_2	δ_3	χ^2
OLS	−0.4731	0.0636	−0.0937	−0.3403	0.1075		−0.8308	0.5617	−0.0419	0.1326	
WH	(0.074)	(0.011)	(0.024)	(0.022)	(0.025)	94.47	(0.092)	(0.069)	(0.005)	(0.024)	85.50
	[−6.3]	[5.6]	[−3.8]	[−15]	[4.2]		[−9.1]	[8.0]	[−7.1]	[5.4]	
HH (12)	(0.166)	(0.018)	(0.056)	(0.048)	(0.048)	83.95	(0.314)	(0.201)	(0.012)	(0.068)	40.00
	[−2.8]	[3.4]	[−1.7]	[−7.0]	[2.2]		[−2.6]	[2.7]	[−3.7]	[1.9]	
NW (18)	(0.158)	(0.018)	(0.056)	(0.045)	(0.048)	68.30	(0.283)	(0.187)	(0.010)	(0.064)	51.61
	[−2.9]	[3.4]	[−1.7]	[−7.4]	[2.2]		[−2.9]	[3.1]	[−4.2]	[2.1]	
s-HH	(0.204)	(0.033)	(0.060)	(0.063)	(0.067)	48.91	(0.262)	(0.193)	(0.014)	(0.080)	30.17
	[−2.3]	[1.9]	[−1.6]	[−5.3]	[1.6]		[−3.2]	[2.9]	[−2.9]	[1.6]	
	R^2_{adj}	0.63					R^2_{adj}	0.51			
	ρ_0	ρ_1	ρ_2	ρ_3	ρ_4	χ^2	δ_0	δ_1	δ_2	δ_3	χ^2
IV	−0.4303	0.0649	−0.1097	−0.3779	0.0935		−0.8312	0.5390	−0.0474	0.1397	
	(0.172)	(0.019)	(0.068)	(0.040)	(0.049)	70.41	(0.193)	(0.129)	(0.009)	(0.049)	42.50
	[−2.5]	[3.3]	[−1.6]	[−9.2]	[1.9]		[−4.3]	[4.1]	[−5.1]	[2.8]	
	R^2_{adj}	0.63					R^2_{adj}	0.51			

Notes: Standard error in parenthesis; *t*-statistics in square brackets.

We thus use both models (7.5) and (7.6) to predict curvature obtained from the real TS. Results support the thesis that model R performs better than model M. During periods of economic slowdown, both models tend to overestimate curvature. As before, a further robustness check is provided by predictive accuracy tests.

Finally, we wish to verify whether the curvature factor is related to the aggregate demand (AD) curve, which is usually assumed to describe the state of the economy.

$$\begin{cases} gap_t = \psi_{0,g} + \psi_{1,g} E_t\left(gap_{t+1}\right) + \left(1-\psi_{1,g}\right)\cdot\left[\psi_{2,g}\, gap_{t-1}\right] \\ \qquad\quad +\psi_{3,g}\left[ffr_t - E_t\left(\pi_{t+1}\right)\right] + \varepsilon_{t,g} \\ cur_t = \psi_{0,c} + \psi_{1,c} E_t\left(cur_{t+1}\right) + \left(1-\psi_{1,c}\right)\cdot\left[\psi_{2,c}\, cur_{t-1}\right] \\ \qquad\quad +\psi_{3,c}\left[ffr_t - E_t\left(\pi_{t+1}\right)\right] + \varepsilon_{t,c} \end{cases} \tag{7.7}$$

A traditional AD (IS) curve implies that the output gap is a function of its forward-looking component, its lagged realizations, and the expected real interest rate. We jointly estimate equations (7.7) in order to compare the coefficients obtained from the actual AD curve with those coming from the curvature equation. The forward-looking real component in the AD equation captures both consumption-smoothing behavior, which is an empirical regularity, and the sentiment about the future state of the economy. The system is GMM estimated, thus matching a twofold objective. Instruments are necessary since expected future (unobserved) variables appear in both equations; they are also required to back generated regressors in the second equation. In both cases, variables may be eventually measured with errors, so that the GMM allows obtaining robust estimates.[9] In addition, GMM estimation handles with heteroscedasticity and serial correlation of unknown forms.

Result of estimating system (7.7) is reported in Table 7.4. If estimated coefficients of the first equation are similar to the respective coefficients

Table 7.4 Estimates of equations (7.7)

	AD equation							
	IP gap				**Curvature**			
	$\psi_{1,g}$	$\psi_{1,g}$	$\psi_{2,g}$	$\psi_{3,g}$	$\psi_{0,c}$	$\psi_{1,c}$	$\psi_{2,c}$	$\psi_{3,c}$
GMM	−0.0494	0.8931	0.8292	0.0246	−0.0608	0.7366	0.9711	0.0197
	(0.035)	(0.089)	(0.339)	(0.013)	(0.183)	(0.292)	(0.209)	(0.042)
	[−1.4]	[10]	[2.4]	[1.9]	[−0.3]	[2.5]	[4.6]	[0.5]
	R^2_{adj}	0.88			R^2_{adj}	0.95		

Notes: Standard errors in parenthesis; *t*-statistics in square brackets.

of the second equation, we may presume that curvature can proxy the IP gap, that is the curvature reflects the cyclical fluctuations of real activity. The magnitude of coefficients are comparable. The better fit of the curvature equation might be due to the higher persistence of the financial factor. Residuals from both equations are serially uncorrelated and sufficiently homoscedastic. Residuals' correlograms is regularly distributed around *zero*. The Wald test, used to check whether the estimated parameters in the IP gap equation are equal to the respective counterparts in the curvature equation, supports both individual and joint coefficient equality.

As a final check, we forecast the IP gap after replacing the dependent variable of the curvature equation with the aforementioned variable, thus imposing the IP gap to be a function of curvature and the real interest rate. The forecast tracks quite well the actual series of the IP gap reproducing the real pattern of business cycle fluctuations.

7.6 A cyclical model for curvature

In this section, we provide some more evidence relating curvature to the cyclical component extracted from a structural time series model for IP (Harvey 1989). IP is assumed to have a stochastic trend and a cyclical component; the former represents the long-term movement in the time series while the latter determines the entity of economic fluctuation, that is the cycle dynamics. Two random walk processes underlie the stochastic trend μ_t:

$$\mu_t = \mu_{t-1} + \beta_t + \eta_t \tag{7.8}$$

$$\beta_t = \beta_{t-1} + \varsigma_t \tag{7.9}$$

η_t and ς_t are white noise mutually uncorrelated disturbances with *zero* mean and standard deviation σ_n and σ_ς respectively. Both the upward and downward movements of the trend are driven by the η_t component; while the steepness of the trend depends on ς_t. Whenever the variance of the disturbances collapses to *zero* the stochastic trend turns into a deterministic one; on the other hand, the larger the variances the greater the stochastic movements of the trend.

The cycle ψ_t is technically constructed using the sine/cosine wave functions. The length of a cycle is called the *period*, which represents the time taken to go through its complete range of values ($2\pi/\lambda$); while the *frequency* (λ) measures how often the cycle is repeated in the unit of time. The cycle is then characterized by few other parameters, the amplitude (A) and the phase shift (θ). The cyclical component is thus expressed as follows:

$$\psi_t = A\cos(\lambda t - \theta) \tag{7.10}$$

A complete formulation representing the cycle combines both the sine and the cosine waves:

$$\psi_t = a\cos(\lambda t) + b\sin(\lambda t) \tag{7.11}$$

The time series of the IP cycle thus can be seen as the summation of the above cyclical component plus a white noise error term with *zero* mean. A stochastic pattern for the cycle requires parameters a and b to evolve over time; to preserve time series continuity we adopt the following recursion:

$$\left[\frac{\psi_t}{\psi_t^*}\right] = \left[\frac{\cos\lambda \quad \sin\lambda}{-\sin\lambda \quad \cos\lambda}\right]\left[\frac{\psi_{t-1}}{\psi_{t-1}^*}\right] + \left[\frac{\kappa_t}{\kappa_t^*}\right] \tag{7.12}$$

with initial states $\psi_0 = a$ and $\psi_0^* = b$; and, where κ_t and κ_t^* are white noise disturbances. The model is identified if either we assume that two disturbances have the same variance or if they are uncorrelated. Finally, we introduce a dumping factor (ρ) affecting the amplitude of the cycle in order to allow for more flexibility. System (7.13) summarizes the entire structural model for IP:

$$\begin{bmatrix} \mu_t \\ \beta_t \\ \psi_t \\ \psi_t^* \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \rho\cos\lambda & \rho\sin\lambda \\ 0 & 0 & -\rho\sin\lambda & \rho\cos\lambda \end{bmatrix}\begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \\ \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} \eta_t \\ \varsigma_t \\ \kappa_t \\ \kappa_t^* \end{bmatrix} \tag{7.13}$$

Equation (7.13) is the state, or transition, equation of the state-space representation. The transition matrix on RHS describes the evolution of the unobservable components so that it captures the stochastic behavior of both the trend and the cycle.

The seasonal component is excluded from the model since we use the seasonally adjusted IP series. IP is thus decomposed into a trend and a cycle (plus a residual component). The model is estimated using the Kalman filter for both the level of the seasonally adjusted IP series and for its *log* transformation. The estimation of the amplitude is 0.9357 (p-value: 0); it is a stable solution ($|\rho| < 1$) that denotes a cycle with decreasing amplitude (i.e., convergence).[10]

Diagrams in Figure 7.4 plot the cyclical component together with different indicators of the business cycle. There is a positive and significant

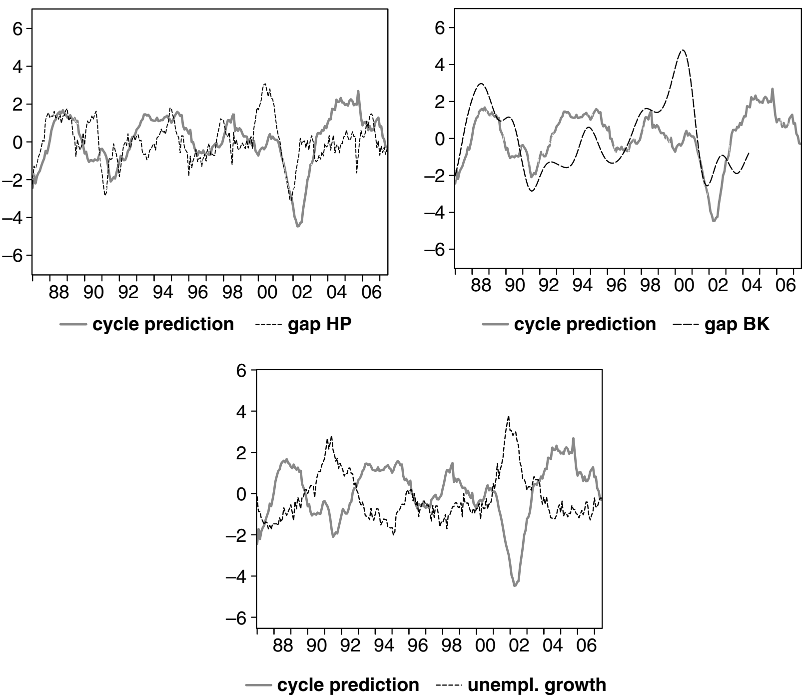


Figure 7.4 Cycle prediction and cyclical variables
Notes: Cycle prediction: cyclical component of the structural model; gap HP: IP gap obtained with the Hodrick–Prescott filter; gap BK: IP gap obtained with the Baxter King filter; unempl. growth: rate of growth of the unemployment rate.

relationship with the output gap computed applying both the Hodrick-Prescott and the Baxter-King frequency filters; while, there is an evident inverse relation with the annual5, instead, shows that the IP cyclical component is highly correlated with the curvature factor *one*-year ahead. The correlation coefficient is about 0.71. Moreover, curvature lies within the forecast standard error bands only except for periods of economic downturn, when it slightly crosses the bands downward. The central and the right panels of Figure 7.5 plot respectively the deviations between curvature and the cycle prediction and the associated correlogram. The forecast errors series is stationary, as suggested by both the autocorrelation function. Stationarity of the forecast errors series is also supported by unit root tests; both the ADF and the Phillips–Perron tests reject the null hypothesis of unit root. In addition, the KPSS test confirms stationarity.

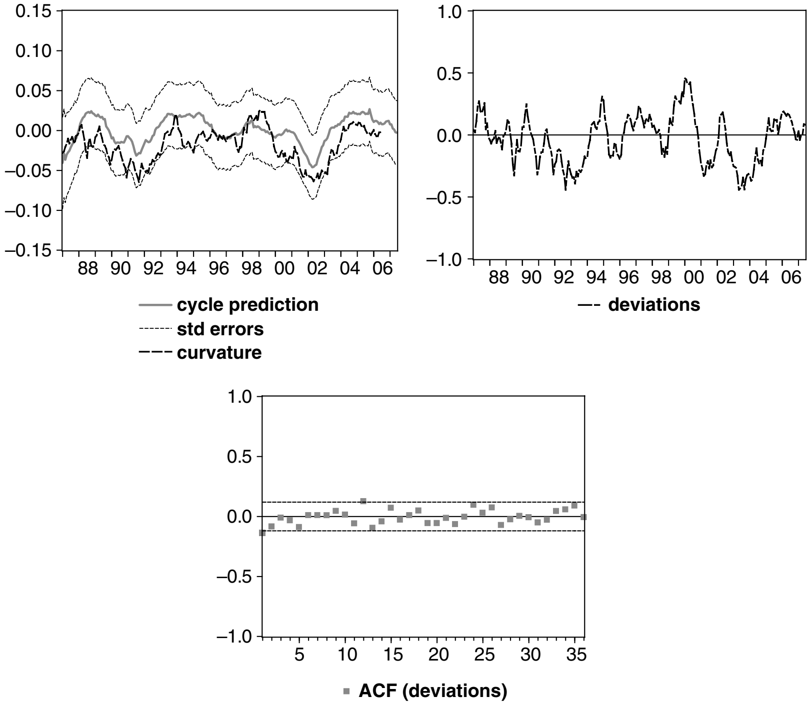


Figure 7.5 Cycle prediction and curvature (nominal TS); autocorrelogram of the deviations

Notes: Cycle prediction: cyclical component of the structural model; std errors: curvature: standard errors of the prediction; curvature: curvature factor; deviations: difference between curvature and its prediction; ACF(deviations): autocorrelogram of the deviation series.

We also compare the cyclical component with curvature from the real TS. The correlation coefficient between the series is almost 0.70. The left panel of the diagram below highlights how similar the path of both series are. The central diagram shows the discrepancies between the curvature and the predicted cycle and the right diagram plots the associated autocorrelogram (Figure 7.6).

7.7 A joint macroeconometric model for curvature and industrial production

To provide more evidence about the economic relationship between curvature and the business cycle, in this section we develop and estimate

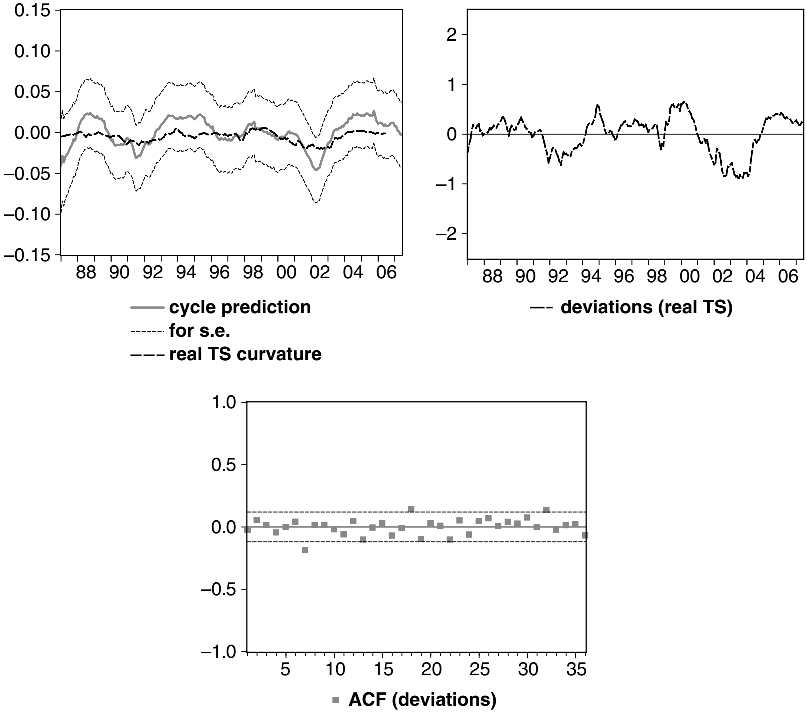


Figure 7.6 Cycle prediction and curvature (real TS); autocorrelogram of the deviations

Notes: Cycle prediction: cyclical component of the structural model; std errors: curvature: standard errors of the prediction; curvature: curvature factor; deviations: difference between curvature and its prediction; ACF(deviations): autocorrelogram of the deviation series.

a joint structural macroeconometric model for both curvature and IP.[11] The measurement equations are summarized by the following system:

$$\begin{bmatrix} \log\,(ip_t) \\ \widehat{cur}_t \end{bmatrix} - \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_t \\ \psi_t \\ \psi_t^* \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \varepsilon_{c,t} \end{bmatrix} \tag{7.14}$$

where $\widehat{cur}_t$ is the simulated trended curvature series. In the model above we assume that both trends follow first-order integrated stochastic processes, while the cyclical components are a combination of sine and cosine waves. The model structure is:

$$\begin{bmatrix} \mu_t \\ \beta_t \\ \psi_t \\ \psi_t^* \\ \mu_{c,t} \\ \beta_{c,t} \\ \psi_{c,t} \\ \psi_{c,t}^* \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \rho\cos\lambda & \rho\sin\lambda & 0 & 0 & 0 & 0 \\ 0 & 0 & -\rho\sin\lambda & \rho\cos\lambda & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \rho_c\cos\lambda & \rho_c\sin\lambda \\ 0 & 0 & 0 & 0 & 0 & 0 & -\rho_c\sin\lambda & \rho_c\cos\lambda \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \\ \psi_{t-1} \\ \psi_{t-1}^* \\ \mu_{c,t-1} \\ \beta_{c,t-1} \\ \psi_{c,t-1} \\ \psi_{c,t-1}^* \end{bmatrix} + \begin{bmatrix} \eta_t \\ \varsigma_t \\ \kappa_t \\ \kappa_t^* \\ \eta_{c,t} \\ \varsigma_{c,t} \\ \kappa_{c,t} \\ \kappa_{c,t}^* \end{bmatrix} \tag{7.15}$$

The model has been estimated with data from 1987 to 2007. The estimated amplitude of the cycle is 0.9311 for IP (ρ) and 0.7741 for the simulated curvature factor (ρ_c). These results are coherent with a decreasing amplitude of the cycle, that is, stability. The covariance between the cycles has been imposed to be approximately *zero*. The left diagram in Figure 7.7 shows the evolution over time of both the predicted

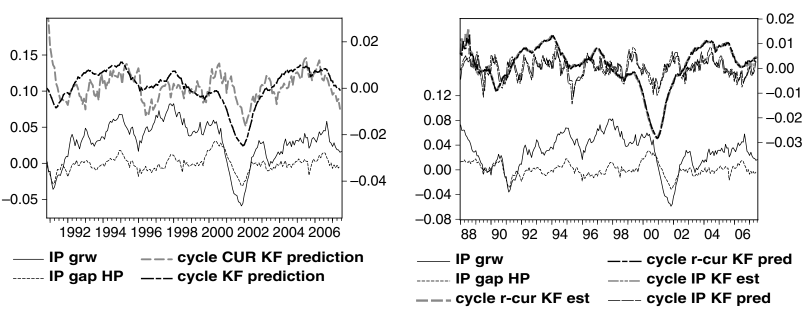


Figure 7.7 Cyclical indicators and curvature

Notes: IP grw: rate of growth of the industrial production index; IP gap HP: industrial production gap obtained with the Hodrick–Prescott filter; cycle CUR KF prediction: curvature prediction obtained by Kalman filtering the structural model; cycle KF prediction: cyclical component obtained by Kalman filtering the structural model.

states of the cycle (right scale) and the cyclical indicators (left scale). Co-movements with the IP growth and the IP gap (HP filtered) are important.

As far as the trend component is concerned, Figure 7.8 plots the estimated deterministic trend of *log* IP and the trend series obtained after Kalman filtering the joint model. The predicted series displays a slightly larger variance than the estimated one; both series fluctuate regularly around the deterministic time trend though. We now show how the decomposition of the SA series of *log* IP into a cyclical component and

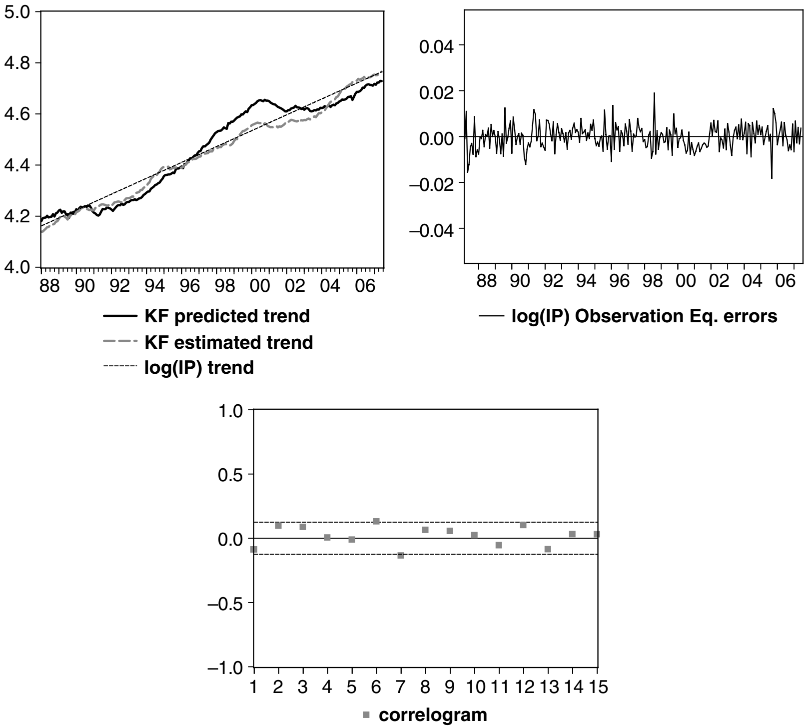


Figure 7.8 Industrial production trend

Notes: KF predicted trend: Kalman filtered trend prediction; KF estimated trend: Kalman filtered estimation of the trend; log (IP) trend: log series of the IP trend; log (IP) Observation Eq. errors: residual of the observation equation of model (7.14) and its correlogram.

a trend reliable is. We thus consider the error term of the measurement equation for *log* IP, that is ε_Z in equation (7.15).

Residuals are covariance stationary (ADF, Phillips–Perron, KPSS). The correlogram confirms the noise series is stationary. The Jarque and Bera test suggests the normal distribution of residuals.[12] Before concluding, we repeat the same experiment using the curvature series obtained from the real TS. The simulated trended curvature series is obtained as described above. The estimated dumping factor that affects the amplitude of the cycle is 0.9365 for IP (ρ) and 0.6926 for simulated curvature (ρ_C). The right panel of Figure 7.8 shows both the predicted and estimated states of the cycle together with the annual IP growth rate and the output gap (constructed by removing the HP filtered *log* IP from the actual series). There is an evident relationship between the series extracted by Kalman filtering and the real economic indicators.

7.8 Conclusion

Both macroeconomists and financial economists have always paid scrupulous attention to the bidirectional relation linking macroeconomics and finance. The yield curve certainly represents an appealing bridge to explore the aforementioned relation. In this vein, TS models provide an effective framework to summarize in few factors all the information contained in the yield curve, which is regarded to be a leading economic indicator. So far, the empirical literature has expressed a certain consensus about the macroeconomic interpretation of only two components underlying TS, namely the level and the slope. The former is associated to the rate of inflation targeted by the monetary authority, while the latter is considered a sign of the monetary policy stance. This study offers a refinement of traditional TS factor-models since we mainly focus on the third latent factor.

Working with US data we provide significant evidence that curvature reflects the cyclical behavior of the economy, as represented by the dynamics of unemployment and IP. We find evidence, in fact, that a negative shock to curvature seems either to anticipate or to accompany a slowdown in economic activity. The curvature effect thus appears to complement the transition from an upward-sloping to a flat yield curve. Interestingly, our main result holds despite the fact that the curvature factor is extracted from the real or the nominal TS of interest rates. In particular, US data also suggest that curvature from the real TS is related to consumption growth.

Notes

1. Dewachter et al. (2006) find evidence that a shock to the central tendency of real interest rates exerts a significant effect on intermediate maturities of the yield curve. Dewachter and Lyrio (2002) suggest that curvature represents a clear independent monetary policy factor; in particular, curvature reflects movements of the real interest rates that are orthogonal to any other macroeconomic variables. They also argue that the slope reflects business cycle conditions.
2. To match the monthly frequency of data 12 is the selected number of lags in the auxiliary regression. Automatic lag selection (Akaike, Schwarz, Hannan-Quinn criteria) leads to similar results. KPSS critical values (including intercept and trend) are 0.216, 0.146, and 0.119 at 1%, 5%, and 10% significance levels respectively. We cannot reject the null of stationarity when the empirical KPSS statistics (reported in the Table) is below the critical values.
3. Nelson and Siegel suggest fixing it equal to 0.06, the value that maximizes the third loading. Following Diebold, Rudebusch and Aruoba (2006) we fix it equal to 0.077.
4. The theoretical measure of curvature proposed by Ang and Piazzesi (2003) is computed as y(1*m*)+y(60*m*)-2*y(12*m*), where *m* indicate the maturity in months. Bekaert, Cho and Moreno (2005) propose y(3*m*)+y(60*m*)-2*y(12*m*). Finally, Nelson and Siegel (1987), as well as Diebold and Li (2006), and Diebold, Rudebusch and Aruoba (2006), compute the curvature as 2*(y24*m*)-y(120*m*)-y(3*m*).
5. The first lag of explanatory variables has been used as instruments.
6. Since we cannot assume that residuals are serially uncorrelated and *iid* normal, the asymptotic *chi-square* test rather than the small sample *F*-test is used to assess the joint significance of estimated coefficients. The White correction is used in presence of heteroscedasticity of unknown form. The Hansen-Hodrick correction is a standard way to deal with overlapping data and serially correlated residuals in forecasting models. The Hansen-Hodrick procedure does not guarantee a positive definite covariance matrix. The Newey-West correction returns a positive definite matrix. The *chi-square* statistics to test for joint significant is suspiciously large, so that we carry out estimates with the simplified HH. Standard errors are built ignoring conditional heteroscedasticity and assuming that serial correlation is simply due to overlapping observations of homoscedastic forecast errors.
7. We have estimated the monetary equation (7.5) for the slope factor; all coefficients are statistically significant and robust (White, Hansen-Hodrick, Newey-West, simplified HH). The goodness of fit is about 60%. Monetary variables predict the slope more accurately than curvature. We have jointly estimated two simple Taylor-type reaction functions by maximum likelihood. One equation based on the federal funds rate, the other on the slope factor, which (as in Rudebusch and Wu, 2004) is considered as a proxy of the policy rate. The first lag of the dependent variable has been included among regressors to capture monetary policy inertia. The Wald test confirms that respective coefficients are the same in the two Taylor-type equations. GMM results are similar if we include future (expected) inflation.

8. Our analysis is theory-consistent since the level obtained from the nominal TS dominates in magnitude the level factor extracted from the real TS, which is flat.
9. GMM estimation does not require distributional assumptions. GMM is a large sample estimator; each equation is estimated using more than 250 observations. Instruments: the annual rate of growth of industrial production and its first lag, both of which are highly correlated with both the industrial production gap and the curvature factor; the realized real interest rate, computed as the difference between the fist lag of the federal funds rate and actual inflation.
10. The estimation of the of the variance of the disturbances are the following. Measurement equation: $(0.1588)^2$ (p-value: 0); trend component: $(0.0564)^2$ (p-value: 0.0040); and cyclical component: $(0.3358)^2$ (p-value: 0). The estimate of the amplitude of the cycle extracted from the log series of the IP is 0.9461; the variance of disturbances are the following. Measurement equation: $(0.0016)^2$ (p-value: 0); trend component: $(0.006)^2$ (p-value: 0.0030); and cyclical component $(0.004)^2$ (p-value: 0). In both cases convergence is achieved after quite a few iterations.
11. Since curvature is stationary we need to add a stochastic trend in order to make it comparable with the *log* IP series. We estimate the trend and the intercept of the IP series; then we run a stochastic simulation (with 1000, 5000, and 10000, repetitions achieving similar results) in order to get the trended series for curvature. The OLS regression of *log* IP onto the constant and the trend returns an estimate of 3.41 and 0.0026 respectively; both coefficients are statistically significant with null p-values. Statistical significance is confirmed by both the White and the Newey-West corrections.
12. We have also run an auxiliary OLS estimation of *log* IP onto the trend and the cycle. Residuals obtained from this regression turns out to be homoscedastic and serially uncorrelated. The statistical properties of both the error series from the measurement equation and the residuals from this auxiliary regression are almost identical. Moreover, since regressors of the aforementioned auxiliary regression are generated series (KF estimated trend and cycle), we have also employed the IV method, using as instruments the lagged values of IP. The IV estimated coefficients are actually the same; the estimation returns a suitable pattern for residuals. Trivially, as largely expected, the goodness of fit of the auxiliary regression is almost 1.

References

Ang, A., Bekaert, G., and Wei, M. (2008) "The Term Structure of Real Rates and Expected Inflation," *Journal of Finance*, 63 (2): 797–849.

Ang, A. and Piazzesi, M. (2003) "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables," *Journal of Monetary Economics*, 50 (4): 745–787.

Ang, A., Piazzesi, M., and M. Wei. (2006) "What Does the Yield Curve Tell Us about GDP Growth?" *Journal of Econometrics*, 131 (1): 359–403.

Bekaert, G., Cho S., and Moreno, A. (2005) "New-Keynesian Macroeconomics and the Term Structure," NBER Working Paper 11340, Cambridge, Massachusetts.

Bernanke, B. S. and Blinder, A. S. (1992) "The Federal Funds Rate and the Channel of Monetary Transmission," *American Economic Review*, 82 (4): 901–921.

Bernanke, B. S. and Mihov, I. (1998) "Measuring Monetary Policy," *The Quarterly Journal of Economics*, 113 (3): 869–902.

Campbell, J. Y. and Cochrane, J. H. (1999) "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behaviour," *Journal of Political Economy*, 107 (2): 205–251.

Chapman, D. A. (1997) "The Cyclical Properties of Consumption Growth and the Real Term Structure," *Journal of Monetary Economics*, 39 (2): 145–172.

Cochrane, J. H. (2005) *Asset Pricing,* Princeton, NJ: Princeton University Press.

Cochrane, J. H. and Piazzesi, M. (2005) "Bond Risk Premia," *American Economic Review*, 95 (1): 138–160.

Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985) "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53 (2): 385–408.

Dai, Q. and Singleton, K. J. (2000) "Specification Analysis of Affine Term Structure Models," *Journal of Finance*, 55 (5): 1943–1978.

Dai, Q. and Singleton, K. J. (2001) "Expectations Puzzles, Time-Varying Risk Premia and Dynamic Models of the Term Structure," *Journal of Financial Economics*, 63 (3): 415–441.

Dewachter, H. and Lyrio, M. (2002) "Macro Factors and the Term Structure of Interest Rates," Working Paper, Catholic University of Leuven, Leuven.

Dewachter, H., Lyrio, M., and Maes, K. (2006) "A Joint Model for the Term Structure of Interest Rates and the Macroeconomy," *Journal of Applied Econometrics*, 21 (4): 439–463.

Diebold, F. X. and Li, C. (2006) "Forecasting the Term Structure of Government Bond Yields," *Journal of Econometrics*, 130 (2): 337–364.

Diebold, F. X., Piazzesi, M., and Rudebusch, G. D. (2005) "Modeling Bond Yields in Finance and Macroeconomics," *American Economic Review*, 95 (2): 415–420.

Diebold, F. X., Rudebusch, G. D., and Aruoba, S. B. (2006) "The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach," *Journal of Econometrics*, 131 (1): 309–338.

Duffee, G. R. (2002) "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance*, 57 (1): 405–433.

Duffie, D. and Kan, R. (1996) "A Yield-Factor Model of Interest Rates," *Mathematical Finance*, 6 (4): 379–406.

Durbin, J. and Koopman, S. J. (2001) *Time Series Analysis by State Space Methods,* Oxford: Oxford University Press.

Estrella, A. and Hardouvelis, G. A. (1991) "The Term Structure as a Predictor of Real Economic Activity," *Journal of Finance*, 46 (2): 555–576.

Estrella, A. and Mishkin, F. S. (1997) "The Predictive Power of the Term Structure of Interest Rates in Europe and the United States: Implications for the European Central Bank," *European Economic Review*, 41 (7): 1375–1401.

Evans, C. L. and Marshall, D. A. (2007) "Economic Determinants of the Nominal Treasury Yield Curve," *Journal of Monetary Economics,* 54 (7): 1986–2003.

Fama, E. F. and Bliss, R. R. (1987) "The Information in Long Maturity Forward Rates," *American Economic Review*, 77 (4): 680–692.

Favero, C. A. (2006) "Taylor Rules and the Term Structure," *Journal of Monetary Economics*, 53 (7): 1377–1393.

Gallmeyer, M. F., Hollifield B., and Zin, S. E. (2005) “Taylor Rules, McCallum Rules and the Term Structure of Interest Rates,” *Journal of Monetary Economics*, 52 (5): 921–950.

Garcia, R. and Luger, R. (2007) “The Canadian Macroeconomy and the Yield Curve: An Equilibrium-Based Approach,” *Canadian Journal of Economics*, 40 (2): 561–583.

Hamilton, J. D. and Kim, D. H. (2002) “A Reexamination of the Predictability of Economic Activity Using the Yield Spread,” *Journal of Money, Credit, and Banking*, 34 (2): 340–360.

Harvey, C. R. (1988) “The Real Term Structure and Consumption Growth,” *Journal of Financial Economics*, 22 (2): 305–333.

Hordahl, P., Tristani, O., and D. Vestin. (2006) “A Joint Econometric Model of Macroeconomic and Term Structure Dynamics,” *Journal of Econometrics*, 131 (1): 405–444.

Kim, C. J. and Nelson, C. R. (1999) *State Space Models with Regime Switching,* Cambridge, Mass.: MIT Press.

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

Nelson, C. R. and Siegel, A. F. (1987) “Parsimonious Modelling of Yield Curves,” *Journal of Business*, 60 (4): 473–489.

Nelson, C. R. and Siegel, A. F. (1988) “Long-Term Behaviour of Yield Curves,” *Journal of Financial and Quantitative Analysis*, 23 (1): 105–110.

Piazzesi, M. (2003) “Affine Term Structure Models,” in Y. Ait-Sahalia and L. P. Hansen (eds.), *Handbook of Financial Econometrics*, Princeton, NJ.

Rudebusch, G. D. (1995) “Federal Reserve Interest Rate Targeting, Rational Expectations, and the Term Structure,” *Journal of Monetary Economics*, 35 (2): 245–274.

Rudebusch, G. D. and Wu, T. (2007) “Accounting for a Shift in Term Structure Behavior with No-Arbitrage and Macro-Finance Models,” *Journal of Money, Credit and Banking*, 39 (2–3): 395–422.

Rudebusch, G. D. and Wu, T. (2008) “A Macro-Finance Model of the Term Structure, Monetary Policy, and the Economy,” *Economic Journal*, 118 (530): 906–926.

Singleton, K. J. (2006). *Empirical Dynamic Asset Pricing. Model Specification and Econometric Assessment,* Princeton, NJ: Princeton University Press.

Stock, J. H. and Watson, M. W. (1989) “New Indexes of Coincident and Leading Indicators,” in O. Blanchard and S. Fischer (eds.), *NBER Macroeconomics Annual,* vol. IV, Cambridge, Mass.: MIT Press.

Wachter, J. (2006) “A Consumption Based Model of the Term Structure of Interest Rates,” *Journal of Financial Economics*, 79 (2): 365–399.

8
On the Efficiency of Capital Markets: An Analysis of the Short End of the UK Term Structure

Andrew Hughes Hallett and Christian Richter

8.1 Introduction

In this chapter, we analyze the term structure of interest rates in a novel way. We test to what extent the UK short-term interest rate is determined by the short-term US interest rate, *and* how much by the UK monetary instrument. In other words, we test jointly whether and to what extent the uncovered interest parity (UIP) and/or the expectations hypothesis (EH) of the term structure of interest rates holds.

The EH of the term structure was prominently formulated by Fisher (1930), Keynes (1930), and Hicks (1953) and states that long-term interest rates are determined by expectations of future short-term interest rates. UIP, in turn, postulates that the interest differential between two countries should equal the expected rate of depreciation or appreciation of the corresponding exchange rate. UIP received prominence from expositions by Keynes (1923), whose attention had been captured by the rapid expansion of organized trading in the forward exchange markets following World War I.

Both hypotheses have in common that they use expectations. The UIP uses expectations concerning the (spot) exchange rate; and the EH uses expectations of the monetary instrument in our case, but of shorter term interest rates in the general case. Both hypotheses have also in common that the expectations are usually modeled assuming rational expectations, although both hypotheses were actually formulated well before the concept of rational expectations had been developed. Hence, expectations in UIP and EH do not have to be "rational." In fact, in the original literature, expectations were given exogenously. As a result, when combining UIP and EH with rational expectations, one has to be aware that an empirical rejection is not necessarily a rejection of UIP and EH per se but may simply be a rejection of rational expectations. This outcome should be of particular interest in the light of the financial crisis that started in 2007. Rational expectations imply that agents have a complete knowledge of the economy and the economics involved. Therefore, it is not surprising that UIP and EH are usually rejected when rational expectations are included.[1] At the same time, experimental and survey evidence on exchange-rate expectations often rejects rational expectations and also static expectations for that matter but tends to support extrapolative, adaptive, or regressive expectations instead (Marey 2004).[2]

In this chapter, we test UIP and EH jointly using extrapolative expectations for the short end of the UK term structure. We can show that both hypotheses affect the short-term interest rate albeit with unequal weights. In our sample, the short-term interest rate is more affected by the US interest rate than by the UK monetary instrument. However, that does not mean that UK monetary policy has no impact. Indeed, we find that extrapolative expectations serve well as a proxy for the formation of expectations. Finally, we are also able to show how the current financial crisis has affected the link between UK and US interest rates.

This chapter is organized as follows: Section 8.2 introduces our model to be tested and how it is estimated. Section 8.3 presents the results, and Section 8.4 concludes.

8.2 Empirical techniques

8.2.1 The hypotheses tested in this chapter

For UIP we use the common notation:

$$i_{t,1} = i^{f}_{t,1} + \Delta s^{e}_{t} + \lambda_t \tag{8.1}$$

where "λ_t" is the time-varying risk premium, "s_t" is the spot exchange rate, "i_t" is the interest rate (maturity of three months), superscript "f" is foreign, and "e" means expectation.

The EH of the term structure in turn implies:

$$i_{t,1} = \alpha CB^{e}_{t} \tag{8.2}$$

where "CB_t" is the central-bank interest rate. We use extrapolative expectations of the form:

$$x_t^e = (1-\beta)\,x_t - \beta x_{t-1} \tag{8.3}$$

The reason for using extrapolative expectations is that they represent, together with the Kalman filter, an optimal learning and updating algorithm (Garratt and Hall 1997a). It therefore has a sound theoretical foundation in markets where agents may not be perfectly informed all the time. Moreover, using extrapolative expectations results in a lag structure, which is important for calculating phase shifts between monetary policy changes and avoids the poor fits often found with rational expectations.

If we substitute Equation (8.3) into (8.4), we get:

$$\begin{aligned} i_{t,1} &= i^f_{t,1} + (1-\gamma)\,s_t - \gamma s_{t-1} - s_{t-1} + \lambda_t \\ &= i^f_{t,1} + (1-\gamma)\,s_t - (1+\gamma)\,s_{t-1} + \lambda_t \end{aligned} \tag{8.4}$$

Using extrapolative expectations in EH results in:

$$i_{t,1} = \alpha\,(1-\delta)\,CB_t - \alpha\delta CB_{t-1} \tag{8.5}$$

Take a time-varying weighted average of both equations (i.e., sometimes domestic influences are more important but sometimes foreign pressures are more important):

$$\begin{aligned} w_t i_{t,1} + (1-w_t) i_{t,1} &= w_t[\alpha\,(1-\delta)\,CB_t - \alpha\delta CB_{t-1}] \\ &\quad + (1-w_t)[i^f_{t,1} + (1-\gamma)\,s_t - (1+\gamma)\,s_{t-1} + \lambda_t] \\ \Leftrightarrow i_{t,1} &= w_t\alpha(1-\delta)CB_t - w_t\alpha\delta CB_{t-1} + (1-w_t) \\ &\quad [i^f_{t,1} + (1-\gamma)s_t - (1+\gamma)s_{t-1} + \lambda_t] \end{aligned} \tag{8.6}$$

Equation (8.6) is the equation we estimate in this chapter, although we added one extra lag for the foreign interest rate. If UIP and EH had the same impact on the UK interest rate, then w_t would be equal to 0.5. Notice that in equation (8.6), not only do the variables vary over time but so also do the parameters. Our approach therefore allows for time-varying risk premiums as well as for a changing relationship between foreign and domestic interest rates (market conditions at home and abroad).

In order to estimate the parameters in Equation (8.6) we use the Kalman filter. That is, we estimate the following state space model:

$$i_t = D_t X_t + \varepsilon_{1,t} \tag{8.7}$$

where equation (2.7) is the measurement equation, and with

$$D_t = D_{t-1} + \varepsilon_{2,t} \tag{8.8}$$

where $\varepsilon_{a,t} \sim$ i.i.d. $(0, \sigma^2_{\varepsilon_a})$ for a = 1, 2, in the state equation.

In this formulation, i_t is the British three-month T-Bill rate; X_t is a set of determining variables such as the British base rate (the monetary instrument in the British case) and the US three-month T-Bill rate; and D_t is a matrix of estimated parameters, including any time-varying risk premium. In either case, the rationale of equation (2.8) is that agents only update the parameters of the model once an unforeseen shock has occurred (Lucas 1976). Moreover, an attractive advantage of the Kalman filter algorithm is that it assumes that agents form one-period ahead forecasts. These forecasts are then compared with the corresponding (new) observation for the same variable. According to the Kalman gain, the coefficients are systematically updated in order to minimize the one-period ahead forecast error. That property makes the Kalman filter convenient for modeling the process of learning[3] and the acquisition of new information It incorporates rational learning behavior by market participants, defined as the ability to minimize their short-run forecasting errors.

The question now is how the parameters are updated to reflect learning. Wells (1996) shows that, in the case of an exogenous shock, the parameters are optimally updated as follows:

$$d_{t|t} = d_{t|t-1} + K_t \left(i_t - X_t d_{t|t-1}\right) \tag{8.9}$$

where $d_{t|s}$ denotes the estimate of the state "d" at time t conditional on the information available at time s. The interesting part of equation (8.9) is the term in brackets. It shows the forecast error. Hence, the current parameters are updated according to the forecast error resulting from an estimated parameter, which did not contain the additional information revealed in the current period. This forecast error in turn affects the Kalman gain. Thus, the Kalman gain may be calculated as:

$$K_t = P_{t|t-1} X_t' \left(X_t P_{t|t-1} X_t' + \Xi\right)^{-1} \tag{8.10}$$

where $P_{t|s}$ is the variance of the forecast error at time t conditioned on the system at time s and Ξ is the covariance matrix of $\varepsilon_{2,t}$. In other words, the updating process depends on the one period forecast error *and* its distribution in the past.

8.2.2 Significance tests and diagnostic test

Using the procedure described so far implies that we get a set of parameter values for each point in time. Hence, a particular parameter could be significant for all points in time, or at some periods but not others; or it might never be significant. These parameter changes are at the heart of this chapter since they imply changes in the lag structure and hence in our frequency and dependency analysis. We therefore employed the following testing strategy. We start with a general lag structure of order q. The value of q is determined by the Akaike information criterion (AIC) test. If a particular lag was never significant (across successive time periods) then this lag was dropped from the equation and the model estimated again. If the AIC criterion was less than before, then that lag was excluded altogether. If a parameter was significant for some periods but not others, it was kept in the equation with a parameter value of zero for those periods in which it was insignificant. This strategy minimizes the AIC criterion and leads to a parsimonious specification. Finally, we tested the residuals in each regression for the absence of serial correlation and heteroscedasticity.

The final specification, equations (8.8)–(8.9), was then *validated* using two different stability tests. Both tests check the same null hypothesis against differing temporal instabilities. The first is the fluctuations test of Ploberger et al. (1989), which detects *discrete* breaks at any point in time in the coefficients of a (possibly dynamic) regression. The second test is due to LaMotte and McWorther (1978) and is designed specifically to detect *random* parameter variation of a specific unit root form (our specification). We found that the random walk hypothesis for the parameters was justified for each country (results available on request). Finally, we chose the fluctuations test for detecting structural breaks because the Kalman filter allows structural breaks at any point and the fluctuations test is able to accommodate this.[4] Thus, and in contrast to other tests, the fluctuations test is not restricted to any prespecified (and hence untested) number of breaks.[5]

Once this regression is done, it gives us a time-varying model. From this model, we can then *calculate* the short-time Fourier transform as outlined below in order to *calculate* the associated time-varying spectra and cross-spectra.

8.2.3 Spectral analysis

The spectral density function shows the strength of the variations of a time series at each frequency of oscillation. It decomposes the variance of a time series into the component that occurs at each frequency or cycle length. Put in a diagram, it shows at which frequencies the variance or fluctuations are strong or powerful and at which frequencies the variations are weak.

In order to calculate the spectrum from the estimated version of equation (8.6), it is convenient to use the fast Fourier transform. The fast Fourier transform creates a *frequency domain* representation of the original *time domain* representation of the data. Thus, the spectra, cross-spectra, and phase shifts are based on regressions done in the time domain but then transformed into a frequency domain representation by the Fourier transform. However, we have allowed the coefficients in our regressions to vary over time. We therefore have to derive one Fourier transform for each point in time. These calculations define a sequence of short-time Fourier transformations (STFTs). In discrete time, this means the data to be transformed has been broken up into frames (which usually overlap each other). Each frame is then transformed as described, and the result added to a matrix, which records its magnitude, phase, and frequency at each time point. These steps may be expressed as:

$$STFT\{x[n]\} \equiv X(m,\omega) = \sum_{n=-\infty}^{\infty} x[n]\,w[n-m]\,e^{-j\omega n} \tag{8.11}$$

In this case, m and n are different points in time; ω is the frequency and is continuous; $\mathrm{j} = \sqrt{-1}$; and "$n-m$" is the estimation period of the regression currently in play. In our application, the estimation period is not constant but is increasing with each new observation. The squared magnitude of the STFT then yields the spectrogram of the function:

$$spectogram\left\{x_t\right\} \equiv |X(\tau,\omega)|^2 \tag{8.12}$$

In this chapter, the specific algorithm used to calculate the various Fourier transforms is the Bluestein algorithm (Bluestein 1968). This is a well established algorithm, widely used in engineering (Boashash 2003; Boashash and Reilly 1992) but not commonly used in economics.

Finally, Boashash and Reilly (1992) have shown that, once equation (8.2) has been estimated, its coefficients $\alpha_{i,t}$ can be used to calculate the STFT and the power spectra directly. That has the convenient property that the traditional formulae are still valid and may still be used, but they have to be recalculated at each point in time. The time-varying spectrum of the growth rate series can therefore be calculated as follows (see also Lin 1997):

$$P_t(\omega) = \frac{\sigma^2}{\left|1 + \sum_{i=1}^{9} \alpha_{i,t} \exp(-j\omega i)\right|_t^2} \tag{8.13}$$

Hence, at any point in time, the power spectrum can be calculated instantaneously from the updated parameters of the model. In addition, we are able to generate a power spectrum even if we have a short time series, and even if that time series contains structural breaks.

8.2.4 Cross-spectral analysis

Let us assume that we estimated the following model:

$$i_t = A(L)_t x_t + u_t,\ u_t \sim i.i.d.\left(0, \sigma^2\right) \tag{8.14}$$

where $A(L)_t$ is a filter, and L is the lag operator such that $Lz_t = z_{t-1}$. Notice that the lag structure, $A(L)_t$, is time-varying. That means we need to use a time-varying model (we use the Kalman filter again) to estimate the implied lag structure. That is:

$$a_{j,t} = a_{j,t-1} + \eta_{j,t},\ \text{for j = 0, …, q and } \eta_t \sim \left(0, \sigma_\eta^2\right) \tag{8.15}$$

What we are interested in is to find a lead-lag relationship between the different variables. For example, for UIP we would like to know how fast the adjustment of the UK interest rate is, once the US rate changes. That is, by how much is the US rate leading? With respect to the central-bank rate, we would also like to know how much the central-bank rate is reflected in movements in the three-month market interest rate. In other words, by how much does the three-month interest rate lead the central bank's rate? A convenient tool to measure these lead-lag relationships is the phase shift. The phase shift is widely used in frequency or time-frequency analysis[6]. Given that we have already estimated the model (8.13), all we have to do is to use the coefficient to calculate the phase shift from it.

In what follows, we briefly explain the concept of the phase shift. In order to calculate the phase shift, we need the *phase angle*. The phase angle measures the lead or lag relationship between two variables at each cyclical frequency. Formally:

$$\varphi(\omega) = \tan^{-1} \frac{-Q_{YX}(\omega)}{C_{YX}(\omega)} \tag{8.16}$$

where

$$C_{YX}(\omega) = f_{XX}(\omega)\sum_{j=0}^{\infty} a_j \cos\omega j, \text{ and } Q_{YX}(\omega) = f_{XX}(\omega)\sum_{j=0}^{\infty} a_j \sin\omega j \tag{8.17}$$

The phase angle can therefore be written as

$$\varphi(\omega) = \tan^{-1}\left(\frac{\sum_{j=0}^{\infty} a_j \sin\omega j}{\sum_{j=0}^{\infty} a_j \cos\omega j}\right) \tag{8.18}$$

Hence, to calculate the phase angle, all we need to know are the coefficients a_j from equation (8.14). However, in this chapter we will actually analyze a "standardized" phase angle, or *phase shift:*

$$\tau(\omega) = \frac{\varphi(\omega)}{\omega} \tag{8.19}$$

To see how to interpret the phase shift statistic, consider Figure 8.1, which shows one variable is following the other at long cycles, with a delay of one month – peak to peak, say. However, for smaller cycles, the delay is shorter. If the markets are efficient in the conventional sense, the two processes should follow each other very closely since agents are able to process new information relatively quickly. Nevertheless, in other cases, there will be natural leads or lags depending on the structure of the markets, the institutional arrangements, and the degree of financial integration.

The formulae given above are for the time-invariant case. Since we get new values for a_j for each point of observation t, we can apply the above

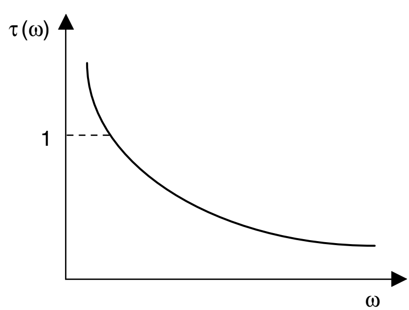


Figure 8.1 Assumed shape of a phase shift

formulae for every point in time t. That means the time-varying phase shift changes to:

$$\tau(\omega)_t = \frac{\varphi(\omega)_t}{\omega} \tag{8.20}$$

8.3 Empirical results for the UK and US money markets

8.3.1 Data set

The UK and US three-month T-Bill rates and the dollar–pound exchange rate are taken from the IFS database. We used monthly data from 1972M1 to 2008M12. The UK central-bank rate is from the Bank of England and is available from its webpage. The sample for the bank rate is also 1972M1–2008M12.

8.3.2 Kalman filter results

To economize on space, we show only the final regressions here. All other results, for the earlier periods, are available from the authors upon request. The time series estimates of equation (8.6) are shown in Table 8.1.

Table 8.1 Regression results

VAR/System (estimation by Kalman filter)			
Dependent variable	UKTBILL	Monthly data from	January 1976 to December 2008
Usable observations	396	Std error of dependent variable	3.4868193831
R^2	0.996184	Standard error of estimate	0.5527347969
Mean of dependent variable	8.1563131313	Sum of squared residuals	118.84562898
Akaike (AIC) criterion	0.54873	Ljung-box test: Q* (40)	44.3872
Variable	Coefficient	Std error	T-Stat
Constant	0.178406348	0.035619453216	5.008677325101
UKDISC	1.121130944	0.100796310081	11.12273795870
UKDISC{1}	−0.302678222	0.084186905184	−3.59531237803
USTBILL	0.244651619	0.041170474386	5.94240467455
GBPDOL	−1.261248657	0.169144089998	−7.4566522360
GBPDOL{1}	0.764879193	0.103076219570	7.42052043125
USTBILL{4}	−0.001187483	0.005018534745	−0.2366194133

In Table 8.1, UKTBILL is the UK three-months T-bill rate, UKDISC is the central-bank rate; USTBILL is the US three-months T-bill rate, and GBPDOL is the pound–dollar rate. We included the fourth lag of the US T-bill rate in order to generate non-autocorrelated errors. If we add the two coefficients of USTBILL and use that value to calculate w_t from equation (8.6), then we get the long-run impact of the US T-bill rate on UK rate, that is it is the impact if the system reaches a steady state. As figure 8.2 shows, the impact of the US rate varies over time but is always the most important determinant, although its influence is shrinking sharply towards the end of the sample. That may be an indication that the UK rate is decoupling itself from the US rate. However, that is a relatively recent phenomenon and follows a period when the US influence was unusually high (2003–2008). The history before that shows that US rates had a steadily declining influence, although it was always high, through the EMS period (1979–1990); but then a slowly increasing influence from the breakdown of the EMS (1993) until the period of recent growth (2006).

The obvious next question that we need to answer is by how much does the US Treasury bill rate lead the UK rate? For that, we look at the phase shift.

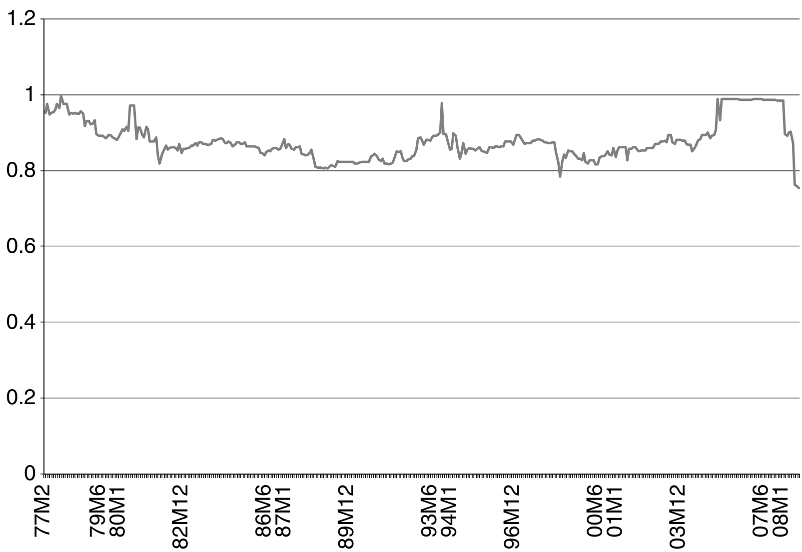


Figure 8.2 Impact of the US rate on the UK rate (w_t)

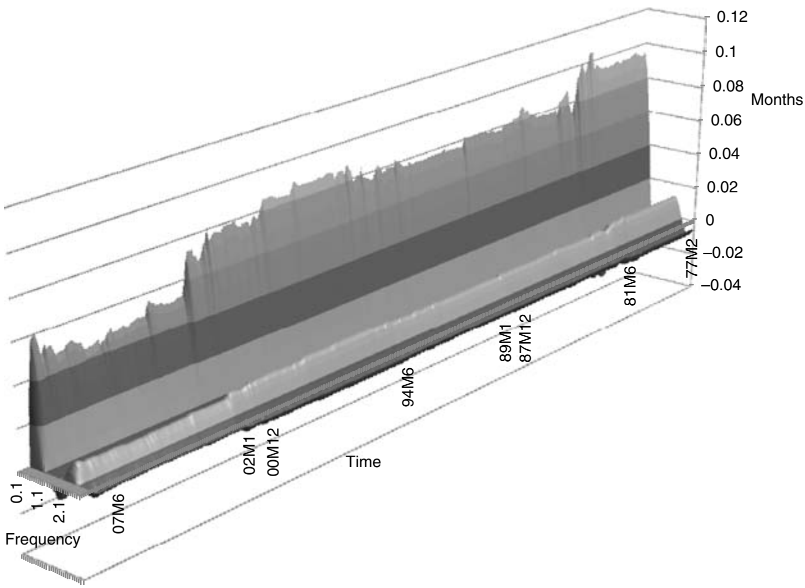


Figure 8.3 Phase shift between the UK and the US rate

From Figure 8.3, we can see that the lead of the US rate at long cycles is small and has decreased over time, essentially since 2002. At the beginning of the sample it was about 0.08 months and that was reduced to 0.06 months by the start of the financial crisis in mid-2007. The financial crisis itself has led to immediate reactions of the UK rate and a small increase in the lead of US interest rates. However, the interesting result is that, according to UIP, the UK rate does not necessarily have to follow the US rate. Instead, the exchange rate (expectation) could change and that could lead to a change of the UK rate. The above figure shows that this is not the case. US rate changes have had a direct impact on the UK rate, implying fairly fixed exchange rates and exchange rate expectations. This appears to have held, even into the current financial crisis. Whilst previous changes in the lead-lag relationship could be attributed to changes in technology (early 70s), the recent immediate incorporation of US rate into UK rates reflects much more markets sentiments towards the US.

The fact that UK agents now incorporate changes in US rates into UK rates more rapidly than they used to, does not contradict the fact that the importance of the US rate decreased. What the above diagram shows is the speed of adjustments and not the extent.

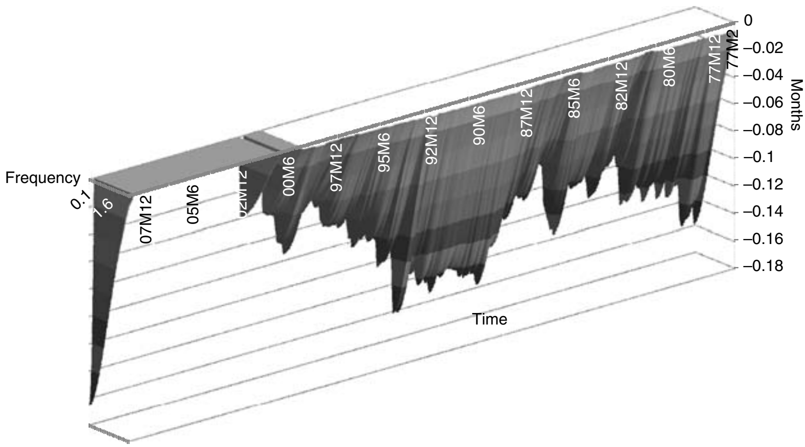


Figure 8.4 Phase shift between the central-bank rate and the T-Bill rate

We now turn our attention to the central bank's rate. Here we asked how well extrapolative expectations work in this model. If extrapolative expectations work well, then they should enable agents to incorporate their expectations into the interest rate and hence the T-bill rate should lead the central bank rate.

Figure 8.4 shows the phase shift between the T-bill rate and the central bank rate in the UK.

From the above figure we can see, that extrapolative expectations work well on average, because the T-bill rate is leading the monetary instrument except in the period 2002 – 2007. Hence, expectations up to 2007 were formed in terms of what they could anticipate of monetary policy and incorporate into the current T-bill rate, otherwise the T-bill rate could not lead the monetary instrument. This example shows that forming extrapolative expectations does not imply that agents are irrational. Instead, they may just be learning how to anticipate the behavior of the monetary authorities in the sense we defined earlier. Thus, if we allow for learning, extrapolative expectations can serve as a tool to anticipate the behavior of other variables.

Finally, we look at the lead-lag relationship between the T-bill rate and the spot exchange rate. From the UIP relationship (8.1), the interest rate will depend on the expected spot rate. As in the previous example, we have assumed extrapolative expectations are at work here. If extrapolative expectations are at work, then they should help to incorporate future developments of the exchange rate into the interest rate and imply

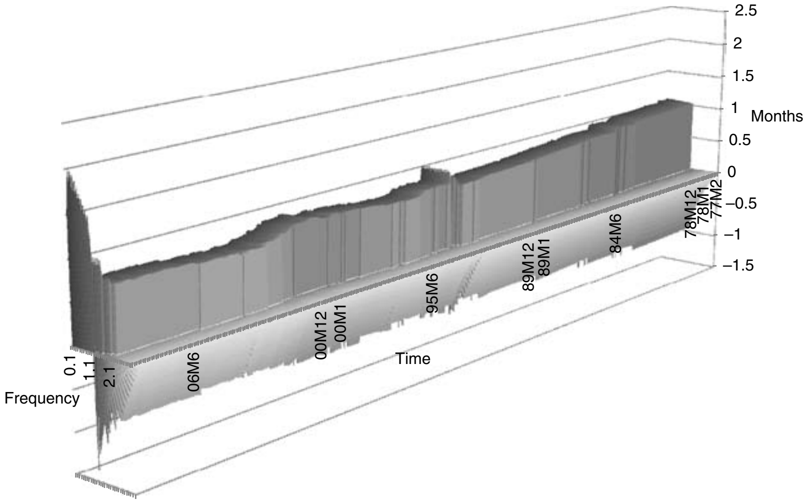


Figure 8.5 The phase shift between the T-bill rate and the exchange rate

that the T-bill rate leads the exchange rate. Figure 8.5 now shows the relationship between the two variables in practice.

Figure 8.5 shows that, in qualitative terms, the lead-lag relationship is relatively stable. Up to a frequency of 1.2, or 5.2 months, the exchange rate is leading the T-bill rate. Nevertheless, for shorter cycles the T-bill rate leads. The interesting thing about these results is that the current crisis has increased the lead of the exchange rate in the long term, and the lead of the T-bill rate in the short term. This says that agents have been able to anticipate short-term fluctuations in the exchange rate relatively well, but have difficulties concerning long run behavior.

Moreover, this diagram also highlights why it is sometimes difficult to reach results concerning the validity of an expectation formation. The advantage of the Fourier transform is that it shows for *all* frequencies how one variable affects another one. It does not average an effect across some or all frequencies. If we had focused solely on a time-series approach, the property that extrapolative expectations work for some cycles but not for all would have been hidden. The reason is that a time-series approach, if not filtered, calculates averages over all the different frequencies.

8.4 Conclusion

In this chapter, we test the EH for the term structure of interest rates jointly with the UIP condition for the short end of the UK term structure.

We find that the US interest rate, the UK monetary instrument, and the (spot) exchange rate all affect the short-term interest rate. However, the impact of the US rate is the biggest effect, although that has decreased a little during the recent financial crisis.

We also tested a bounded rationality approach. We deviated from the contemporary literature by refusing to impose rational expectations. Instead, we have assumed extrapolative expectations as an obvious behavioral alternative. We have shown that incorporating extrapolative expectations in both hypotheses turned out to be a significant improvement. Hence, and in contrast to previous work which has assumed rational expectations, we find the UIP and the EHs are not rejected. Thus, the problem seems to have been violations of the rational expectations paradigm, not violations of UIP or the EH of behavior in the financial markets. Second, we also show that extrapolative expectations formation can help us anticipate the central bank's impact on short-term interest rates. Finally, concerning the exchange rate, we were able to show that extrapolative expectations can help to anticipate short-term movements of the exchange rate but not long-term movements of the exchange rate. Hence, in the short-run, interest rates lead movements in the exchange rate. However, in the longer term they do not.

Notes

1. Many comprehensive surveys exist: see, for example, Froot and Thaler (1990), Lewis (1995) and Engel (1996) for UIP, and Cook and Hahn (1990) and Campbell and Shiller (1991) for EH.
2. See also Hughes Hallett and Richter (2002; 2003a; 2003b; 2004).
3. The Kalman filter is widely used in finance and macroeconomics as a learning algorithm: see for example, Lucas (1976), Garatt and Hall (1997a; 1997b), Whitley (1994). Salmon (1995) shows that the Kalman filter is a special case of a neural network. Hence the Kalman filter can be regarded as an optimal procedural learning algorithm.
4. Note that all our tests of significance, and significant differences in parameters, are being conducted in the time domain, *before* transferring to the frequency domain. This is because no statistical tests exist for calculated spectra (the data transformations are nonlinear and involve complex arithmetic). Stability tests are important here because our spectra are sensitive to changes in the underlying parameters. But, given the extensive stability and specification tests conducted, we know there is no reason to switch to another model that fails to pass those tests.
5. The fluctuations test works as follows: one parameter value is taken as the reference value, for example, the last value of the sample. All other observations are now tested whether they significantly differ from that value. In order to do so, Ploberger et al. (1989) have provided critical values which we have used in

the figures. If the test value is above the critical value then we have a structural break, i.e. the parameters differ significantly from their reference values and vice versa. For reasons of limited space we have excluded the test diagrams from this chapter but report on the results. The diagrams are available from the authors upon request.

6. See, for example, Boashash (2003), Boashash and Reilly (1992) and Hughes Hallett and Richter (2009).

References

Bluestein, L. I. (1968) “A Linear Filtering Approach to the Computation of the Discrete Fourier Transform,” *Northeast Electronics Research and Engineering Meeting Record*, 10 (3): 218–219.

Boashash, B. (2003) *Time Frequency Signal Analysis and Processing,* Oxford: Elsevier.

Boashash, B. and Reilly, A. (1992) “Algorithms for Time-Frequency Signal Analysis,” in B. Boashash (ed.), *Time-Frequency Signal Analysis: Methods and Applications,* Melbourne: Longman-Cheshire, pp. 163–181.

Campbell, J. Y. and Shiller, R. (1991) “Yield Spreads and Interest Rate Movements: A Bird’s Eye View,” *The Review of Economic Studies,* 58 (5): 496–514.

Cook, T., and Hahn, T. (1990) “Interest Rate Expectations and the Slope of the Money Market Yield Curve,” *Federal Reserve Bank of Richmond Economic Review,* 76 (1): 3–26.

Engel, C. (1996) “The Forward Discount Anomaly and the Risk Premium: A Survey of Recent Evidence,” *Journal of Empirical Finance,* 3 (2): 123–192.

Fisher, I. (1930) *The Theory of Interest,* New York: Macmillan Press.

Froot, K. A. and Thaler, R. H. (1990) “Anomalies: Foreign Exchange,” *Journal of Economic Perspectives,* 4 (3): 179–192.

Garratt, A. and Hall, S. (1997a) “E-equilibria and Adaptive Expectations: Output and Inflation in the LBS Model,” *Journal of Economic Dynamics and Control,* 21 (7): 1149–1171.

Garratt, A. and Hall, S. G. (1997b) “The Stability of Expectational Equilibria in the LBS Model,” in C. Allen and S. G. Hall (eds), *Macroeconomic Modelling in a Changing World,* New York: Wiley, pp. 217–246.

Hicks, J. (1953) *Value and Capital,* London: Oxford University Press.

Hughes Hallett, A. and Richter, C. (2002) “Are Capital Markets Efficient? Evidence from the Term Structure of Interest Rates in Europe,” *Economic and Social Review,* 33 (3): 333–356.

Hughes Hallett, A. and Richter, C. (2003a) “Learning and Monetary Policy in a Spectral Analysis Representation,” in P. Wang and S.-H. Chen (eds.), *Computational Intelligence in Economics and Finance,* Berlin: Springer, pp. 91–103.

Hughes Hallett, A. and Richter, C. (2003b) “A Spectral Analysis of the Short-End of the British Term Structure,” in R. Neck (ed.), *Modelling and Control of Economic Systems,* Amsterdam: Elsevier, pp. 123–128.

Hughes Hallett, A. and Richter, C. (2004) “Spectral Analysis as a Tool for Financial Policy: An Analysis of the Short End of the British Term Structure,” *Computational Economics*, 23 (3): 271–288.

Hughes Hallett, A. and Richter, C. (2009): "Is the US No Longer the Economy of First Resort? Changing Economic Relationships in the Asia-Pacific Region," *International Economics and Economic Policy*, 6 (2): 207–234.

Keynes, J. M. (1923) *A Tract on Monetary Reform,* London: Macmillan Press.

Keynes, J. M. (1930) *A Treatise on Money,* New York: Macmillan Press.

LaMotte, L. R. and McWorther, A. J. (1978) "An Exact Test for the Presence of Random Walk Coefficients in a Linear Regression," *Journal of the American Statistical Association,* 73 (364): 816–820.

Lewis, K. K. (1995) "Puzzles in International Financial Markets," in G. M. Grossman and K. Rogoff (eds.), *Handbook of International Economics,* New York: Elsevier.

Lin, Z. (1997) "An Introduction to Time-Frequency Signal Analysis," *Sensor Review,* 17 (1): 46–53.

Lucas, R. E. (1976) "Econometric Policy Evaluation: A Critique," in K. Brunner, and A. Meltzer (eds.), *The Phillips Curve and Labor Markets,* Amsterdam: North-Holland.

Marey, P. S. (2004) "Uncovered Interest Parity Tests and Exchange Eate Expectations," Working Paper, Maastricht University, Maastricht.

Ploberger, W., Krämer, W., and Kontrus, K. (1989) "A New Test For Structural Stability in the Linear Regression Model," *Journal of Econometrics,* 40 (2): 307–318.

Salmon, M. (1995) "Bounded Rationality and Learning: Procedural Learning," in A. Kirman and M. Salmon (eds.), *Learning and Rationality in Economics,* Oxford: Basil Blackwell.

Wells, C. (1996) *The Kalman Filter in Finance,* Dordrecht: Kluwer Academic Publishers.

Whitley, J. D. (1994) *A Course in Macroeconomic Modelling and Forecasting,* London: Harvester Wheatsheaf.

9
Continuous and Discrete Time Modeling of Short-Term Interest Rates

Chih-Ying Hsiao and Willi Semmler

9.1 Introduction

In modern finance theory, the short-term interest rate is important in characterizing the term structure of interest rates and in pricing interest-rate-contingent-claims. There is some pioneering work in the continuous-time framework, for example by Vasicek (1997) and Cox et al. (1985). A survey of is provided by Chan et al. (1992). Chan et al. (1992) show that a wide variety of well-known one-factor models for short rates can be nested within the following stochastic different equation (SDE):

$$dX_t = (c - \beta X_t)dt + \sigma X_t^{\gamma} dW_t. \tag{9.1}$$

The unpredictable residual of the Chan, Karolyi, Longstaff and Sanders (CKLS) model is modeled as a Brownian motion W_t. The features of this model include a mean-reverting drift coefficient[1] and a level-dependent diffusion coefficient. This *continuous-time* framework can provide elegant expressions in theory, but it entails some difficulty in the empirical research (see Lo 1988). Many methods have been developed to implement the empirical estimations. For example, among others, one can mention the indirect inference method of Gouriéroux et al. (1993), the approximate likelihood method of Perdersen (1995), the general method of moments with respect to diffusion generators by Hansen and Scheinkman (1995) and Duffie and Glynn (2001), the efficient method of moments of Gallant and Tauchen (1996), the nonparametric method of Ait-Sahalia (1996) and Ait-Sahalia (1997), the density-approximation method by Dacunha-Castelle and Florens-Zmirou (1986) and Ait-Sahalia (1999), and the Milstein method by Elerian (1998). Finally, this chapter[2] considers the new local linearization (NLL) method developed by Shoji and Ozaki (1997, 1998).

In order to take continuous-time models to the data, one first has to discretize those models. We employ here three discretization methods: the Euler method, the NLL method, and the Milstein method. The three methods deliver discrete-time approximate models for discrete-time-observed data of a continuous-time diffusion process. In this way we can implement the maximum likelihood estimation (ML estimation) and provide predictions. In the literature, Lo (1988) pointed out that the Euler estimator[3] is not consistent. The Milstein and NLL approximations are shown to improve the Euler approximation (see Elerian 1998: 11, Table 1; and Shoji and Ozaki 1997: 494–501). The improvement in their papers is represented by smaller errors of the *parameter estimations* in the numerical experiments. Our chapter takes another view to assess those models. Besides the accuracy of parameter estimation, we also consider the accuracy of prediction. For the SDE (1) where the drift coefficient is linear, we show that the Euler and the NLL methods provide the same prediction. Thus, these two methods are actually statistically equivalent. For comparing the Euler and the Milstein methods we do not verify the superiority of the Milstein method, in contrast to Elerian (1998). The parameter estimations and the one-step-ahead predictions of the two models are very similar. We argue it is because of the small scales of the parameters, which lead to a relatively small effect on the discretization bias. The scales of the parameters are chosen from the results of our empirical study. In other words, the advantage of the Milstein method is not very significant in the current short-term interest-rate case.

The Euler and Milstein approximate models are applied to the short-term interest-rate data of Germany, the UK, and the USA. The two approximate models perform quite similarly in both estimation and prediction. We find none of the country short-rate data can pass our specification tests in a satisfactory way where the estimated residuals of all the three countries have too high autocorrelation and too thick tails. So we look for new models which can explain these stylized facts. In the continuous-time framework there is some work pointing out the shortcomings of the CKLS model (9.1) (see, for example, Ait-Sahalia 1996 and Andersen and Lund 1997). However, the data simulated by these two continuous-time models still cannot generate the high autocorrelation of the residuals either.

Since we cannot find a suitable model in the continuous-time framework we turn to the discrete-time framework. The autoregressive moving-average (ARMA) structure is a candidate for fitting high autocorrelations of the estimated white noise. We will see, in Section 9.4, that we can model the autocorrelation of the estimated noise by taking more lags in the models. To model the thick tails in the estimated white noise we follow the work of Brenner et al. (1996) and Koediji et al. (1997). They employ the autoregressive conditional heteroscedastic (ARCH) model suggested by Engle (1982) and Bollerslev (1986) to model the thick tails. In addition, their conditional variance depends on the level of the short-time interest rates. Combining the modeling strategies above, we obtain an ARMA-ARCH structure with level-dependent volatility. Our model generalizes the model of Brenner et al. (1996) by the ARMA structure.

The remainder of this chapter is organized as follows. Section 9.2 introduces the three discretization methods. We show that the Euler and NLL approximate models for the SDE (9.1) are equivalent under reparametrization. In Section 9.3, the Euler and the Milstein approximations will be applied to the empirical short-rate data. We find evidence which cannot be represented by the model (9.1). We thus look for new models in Section 9.4, where we find the discrete-time ARMA-ARCH model with level-dependent volatility is a better candidate for the short rates. Section 9.5 concludes this chapter.

9.2 Discrete-time approximation

Here we introduce briefly the three methods of discrete-time approximation: the Euler, the Milstein, and the NLL method.

9.2.1 Euler method

The idea of the Euler method is to replace dt in the equation (1) by a time interval Δt and we have a discrete-time approximation for the diffusion process X:

$$X_{t_{i+1}} - X_{t_i} = b(X_{t_i}, \theta)\Delta t_i + a(X_{t_i}, \theta)\Delta W_{t_i}. \tag{9.2}$$

9.2.2 Milstein method

The Milstein method approximates the SDE by the following scheme:

$$X_{t_{i+1}} - X_{t_i} = b(X_{t_i}, \theta)\Delta t_i + a(X_{t_i}, \theta)\Delta W_{t_i} + \frac{1}{2}a(X_{t_i})a'(X_{t_i})((\Delta W_{t_i})^2 - \Delta t_i) \tag{9.3}$$

where $\Delta t_i = (t_{i+1} - t_i)$ and $\Delta W_{t_i} = W_{t_{i+1}} - W_{t_i}$.[4] It has one more term then the Euler method of the equation (9.2) and better convergence.[5]

The likelihood based on the Milstein method is calculated as follows. Following (9.3), the dynamic of the SDE (9.1) is approximated by

$$X_{t_{i+1}} - X_{t_i} = (c - \beta X_{t_i})\Delta t_i + \sigma X_{t_i}^{\gamma} \Delta W_{t_i} + \frac{1}{2}\sigma^2 \gamma X_{t_i}^{2\gamma-1}(\Delta W_{t_i}^2 - \Delta t_i),$$

where $\Delta t_i = t_{i+1} - t_i$, $\Delta W_{t_i} = W_{t_{i+1}} - W_{t_i}$.

Let

$$Y_{t_{i+1}} = X_{t_{i+1}} - X_{t_i} - (c - \beta X_{t_i})\Delta t_i + \frac{1}{2}\sigma^2 \gamma X_{t_i}^{2\gamma-1} \Delta t_i.$$

The above equation becomes

$$\frac{1}{2}\sigma^2 \gamma X_{t_i}^{2\gamma-1}(\Delta W_{t_i})^2 + \sigma X_{t_i}^{\gamma} \Delta W_{t_i} = Y_{t_{i+1}}.$$

Let $x_i \in \mathrm{R}$ still be the realizations of X_{t_i} and y_i be the realizations of Y_{t_i} for $i = 0, \ldots, N$ correspondingly. We solve the equation (16) to obtain the realizations of $\Delta W_{t_i} = u_{i+1}^+, u_{i+1}^-$, where

$$u_{i+1}^+ = \frac{-1 + \sqrt{1 + \frac{2\gamma y_{i+1}}{x_i}}}{\sigma \gamma x_i^{\gamma-1}}$$

$$u_{i+1}^- = \frac{-1 - \sqrt{1 + \frac{2\gamma y_{i+1}}{x_i}}}{\sigma \gamma x_i^{\gamma-1}}.$$

Then the conditional density is given by

$$p(X_{t_{i+1}} = x_{i+1} | X_{t_i} = x_i) = \frac{dP(\{\Delta W_{t_i} = du_{i+1}^+\} \cup \{\Delta W_{t_i} = du_{i+1}^-\})}{dy_{i+1}}$$

$$= \frac{dP(\Delta W_{t_i} = du_{i+1}^+)}{du_{i+1}^+}\left|\frac{du_{i+1}^+}{dy_{i+1}}\right| + \frac{dP(\Delta W_{t_i} = du_{i+1}^-)}{du_{i+1}^-}\left|\frac{du_{i+1}^-}{dy_{i+1}}\right|$$

$$= \frac{1}{\sqrt{2\pi \Delta t_i}}\left(\exp\left(-\frac{(u_{i+1}^+)^2}{2\Delta t_i}\right) + \exp\left(-\frac{\left(u_{i+1}^-\right)^2}{2\Delta t_i}\right)\right)\left|\frac{1}{\sigma x_i^{\gamma}\sqrt{1 + \frac{2\gamma y_{i+1}}{x_i}}}\right|,$$

as $1 + \frac{2\gamma y_{i+1}}{x_i} > 0$. If $1 + \frac{2\gamma y_{i+1}}{x_i} < 0$, then the density above is infinity. If $1 + \frac{2\gamma y_{i+1}}{x_i} < 0$, which means there is no real solution of ΔW_{t_i} in (16) for such y_{i+1}, therefore the density is equal to zero

$$p(X_{t_{i+1}} = dx_{i+1} | X_{t_i} = x_i) = 0.$$

Comparing this density function and the density function in Equation (2.5) in Elerian (1998: 7), it is not difficult to show the identity of these two functions by some calculation.

By numerical operations of the ML estimations we must modify the density function, because when $1+\frac{2\gamma y_{i+1}}{x_i}=0$, the value of the density function is infinity. Therefore, we apply the following density function for the ML estimations:

$$g_{mil}(x_i,x_{i+1},\theta,\Delta t_i)=\frac{dP(X_{t_{i+1}}=dx_{i+1}|X_{t_i}=x_i)}{dx_{i+1}}$$

$$=\frac{1}{\sqrt{2\pi\Delta t_i}}\left(\exp\left(-\frac{\left(u_{i+1}^{+}\right)^2}{2\Delta t_i}\right)+\exp\left(-\frac{\left(u_{i+1}^{-}\right)^2}{2\Delta t_i}\right)\right)\left|\frac{1}{\sigma x_i^{\gamma}\sqrt{1+\frac{2\gamma y_{i+1}}{x_i}}}\right|,$$

$$\textit{for } 1+\frac{2\gamma y_{i+1}}{x_i}>10^{-10},$$

$$=10^{-10}, \textit{ otherwise}.$$

9.2.3 New local linearization method

The NLL method is suggested by Shoji and Ozaki (1997: 490–491). It approximates the drift coefficient $b(X_s)$ up to the second-order terms by using the Itô formula

$$dX_s=(b(X_{t_i})+b'(X_{t_i})(X_s-X_{t_i})+\frac{1}{2}b''(X_{t_i})a^2(X_{t_i})(s-t_i))ds+a(X_{t_i})dW_s. \tag{9.4}$$

while the diffusion coefficient is still kept as a constant. The equation (9.4) can be solved analytically, and the solution at t_{i+1} is given by

$$\begin{aligned}X_{t_{i+1}}-X_{t_i}&=\frac{b(X_{t_i})}{b'(X_{t_i})}(e^{b'(X_{t_i})(t_{i+1}-t_i)}-1)\\&+\frac{b''(X_{t_i})}{(b'(X_{t_i}))^2}\frac{a(X_{t_i})^2}{2}(e^{b'(X_{t_i})(t_{i+1}-t_i)}-1-b'(X_{t_i})(t_{i+1}-t_i))\\&+a(X_{t_i})\int_{t_i}^{t_{i+1}}e^{b'(X_{t_i})(t_{i+1}-z)}dW_z.\end{aligned} \tag{9.5}$$

The distribution of the last term can is given by

$$a(X_{t_i})\int_{t_i}^{t_{i+1}}e^{b'(X_{t_i})(t_{i+1}-z)}dW_z\overset{dis.}{\sim}\mathrm{N}\left(0,a(X_{t_i})^2\int_{t_i}^{t_{i+1}}e^{2b'(X_{t_i})(t_{i+1}-z)}dz\right). \tag{9.6}$$

9.2.4 Equivalence of the Euler and NLL predictors

Here we show the Euler and NLL predictors of the SDE (1) are actually statistically equivalent due to the linearity of the drift coefficient in Equation (9.1). The Euler approximation is obtained according to Equation (9.3) and given by

$$X_{(i+1)\Delta t} - X_{i\Delta t} = (c - \beta X_{i\Delta t})\Delta t + \sigma X^{\gamma}_{i\Delta t}\Delta W_{i\Delta t} \tag{9.7}$$

while the NLL approximation is obtained according to Equation (5)

$$X_{(i+1)\Delta t} - X_{i\Delta t} = \frac{h_1(\beta)}{\beta}(c - \beta X_{i\Delta t}) + \sigma h_2(\beta) X^{\gamma}_{i\Delta t} U_{i+1} \tag{9.8}$$

Then we can observe an equivalent mapping under the reparametrization

$$\beta_{eu}\Delta t = h_1(\beta_{nll}) := 1 - e^{-\beta_{nll}\Delta t}, \quad c_{eu}\Delta t = \frac{c_{nll}}{\beta_{nll}} h_1(\beta_{nll})$$

$$\gamma_{eu} = \gamma_{nll}, \quad \sigma_{eu} = \sigma_{nll} h_2(\beta_{nll}) := \sigma_{nll}\sqrt{\frac{1 - e^{-2\beta_{nll}\Delta t}}{2\beta_{nll}\Delta t}},$$

where $U_i, i = 1, \ldots$ are i.i.d $\mathsf{N}(0, \Delta t)$-distributed.

9.3 Empirical results on modeling short-term interest rates

9.3.1 Data

We apply only the Euler and the Milstein approximations for the empirical short-rate data. The short-rate data are interest rates with a one-day maturity, which are the call money rate of Germany, the overnight interbank rate of the UK, and the federal funds rate of the USA. All data are monthly data from *"The International Statistical Yearbook"*[6] for the time period January 1983 to December 1997 (180 observations) for estimation. The further period January 1998 to June 2000 (30 observations) is reserved for prediction. We take data after the oil crisis, for January 1983–June 2000, because many researchers have found evidence of regime changes for the crisis period 1979–1982. The time series of the rates are plotted in Figures 9.1, 9.4, and 9.7.

9.3.2 Specification test

The main idea of the specification tests is to check whether there is still deterministic structure in the residuals. Two specification tests are adopted. The first test is to check whether the residuals are auto-correlated. The second one is to test whether the residuals have thick tails.

Checking autocorrelation

Let $U_1, \ldots, U_N$ be identically distributed random variables with that $E[U_i] = 0$, $Var[U_i] = 1$ and $E|U_i|^s < \infty$, for all $s \geq 2$. Let $\hat{R}_k$ be the sample autovariance function

$$\hat{R}_k = \frac{1}{N-k} \sum_{i=k+1}^{N} U_i U_{i-k}.$$

Under the null we have $E[\hat{R}_k] = 0$ and

$$Var[\hat{R}_k] = \frac{1}{N-k},$$

for $k \geq 1$. We normalized $\hat{R}_k$ into

$$\hat{r}_k = \frac{\hat{R}_k - E[\hat{R}_k]}{\sqrt{Var[\hat{R}_k]}} = \sqrt{N-k}\hat{R}_k = \frac{1}{\sqrt{N-k}} \sum_{i=k+1}^{N} U_i U_{i-k}. \tag{9.9}$$

Consider the sequence $(U_i U_{i-k})_{i=k+1,\ldots,N}$ for a fixed k. It is *near epoch dependent* on $(U_i)_{i=1,\ldots,N}$.[7] Using the central limit theorem for near epoch processes,[8] $\hat{r}_k$ converges to $\mathsf{N}(0,1)$ in distribution as $N \to \infty$. Applying the test for our discrete-time approximations, we let $U_i = W_i - W_{i-1}$.

We remark here that $\hat{r}_k \sim \mathrm{N}(0,1)$ means $\hat{R}_k \sim \mathrm{N}(0, \frac{1}{N-k})$. It is similar with the result $Var[\hat{R}_k] \sim 1/N$ in Box et al. (1994: 32) when N is large enough.

Testing normality

We employ here χ^2-test for histogram to test whether the distribution of samples is $\mathsf{N}(0,1)$ distribution.[9] The idea is to compare the relative frequency of samples on intervals I_m

$$\hat{p}_m = \frac{\textit{number of } \{i; U_i \in I_m\}}{N}$$

and p_m the probability of $\mathsf{N}(0,1)$-distribution on the intervals I_m where $\{I_m, m = 1, \ldots, M\}$ are disjoint intervals of the real line.

The weighted distance

$$d = \sum_{m=1}^{M} \frac{N}{p_m(1-p_m)} (\hat{p}_m - p_m)^2 \tag{9.10}$$

measures the distance between the sample and the normal distributions. It converges to $\chi^2(M-1)$ in distribution as $N \to \infty$.

9.3.3 Results of estimating the CKLS model

In Tables 9.1, 9.2, and 9.3, the empirical results are reported. The first two columns show the results of the Euler and Milstein approximations for the CKLS model (9.1). The notations of the parameters are adjusted for Section 9.4 later. All estimates in the drift coefficients are not significantly different from zero. The forecast errors are given in the lower part of the tables. We found that the CKLS model does not provide better data prediction than a "naïve" forecast without any model. In the row "relative," the relative forecast errors comparing a "naïve" forecast are quoted. The naïve forecast just uses today's data to forecast the next period. The relative errors both for the in-sample and out-of-sample forecast are all about 100 percent or even above.

The estimated white noises based on the two methods are very similar as plotted in Figures 9.2, 9.5, and 9.8. Figures 9.3, 9.6, and 9.9. plot the normalized autocorrelations given in Equation (9.9) for the Euler approximation. The normalized autocorrelations should be within $[-2, +2]$ band. However, in the figures they are about 3.5 for Germany and the UK and about 5 for the USA. This finding indicates strong autocorrelation in the estimated residuals. Durbin (1970) and Box and Pierce (1970) have pointed out that the sample autocorrelations will be underestimated for close time differences.[10] In this case, the underestimation indicates an even stronger autocorrelation than the values given above. As a reference, we run a Monte Carlo simulation for 1,000 repetitions using the result of the US estimations. Most of them (96 percent) have the maximal normalized autocorrelations smaller than 2.8 (we take the first ten normalized autocorrelations), and the maximal of them is only 4.2.

We also observe that the estimated residuals are more concentrated around zero than the standard normal distribution, which implies they have thick tails.[11] This fact can be inferred from Figures 9.2, 9.5, and 9.8, and the large d-statistics of the normality test in Tables 9.1–9.3 indicate that the distributions of the residuals are far from a normal distribution.

9.4 Searching for new models

Because of the results that show high correlations and thick tails for the model (9.1) shown in the last section we search for new models.

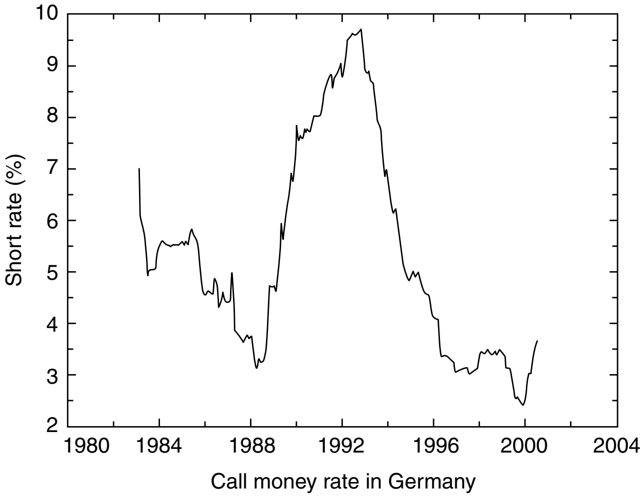


Figure 9.1 Call money rate, Germany

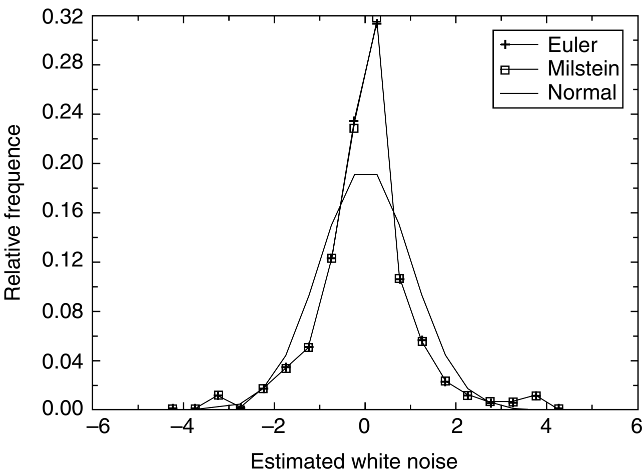


Figure 9.2 Distribution of estimated white noise (I), Germany

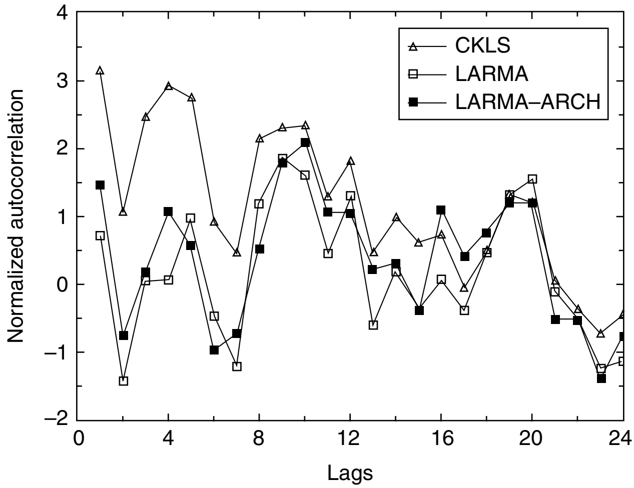


Figure 9.3 Normalized autocorrelation of the estimated noise, Germany

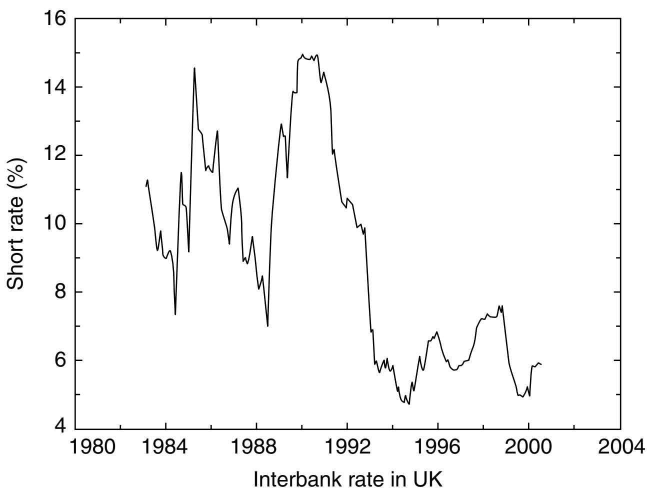


Figure 9.4 Interbank rate, UK

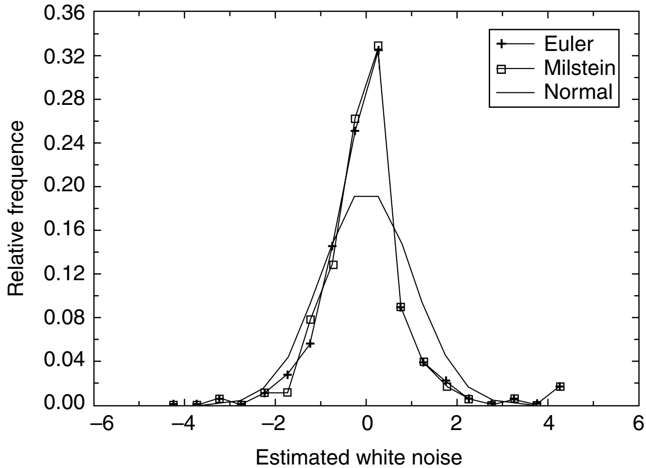


Figure 9.5 Distribution of estimated white noise (I), UK

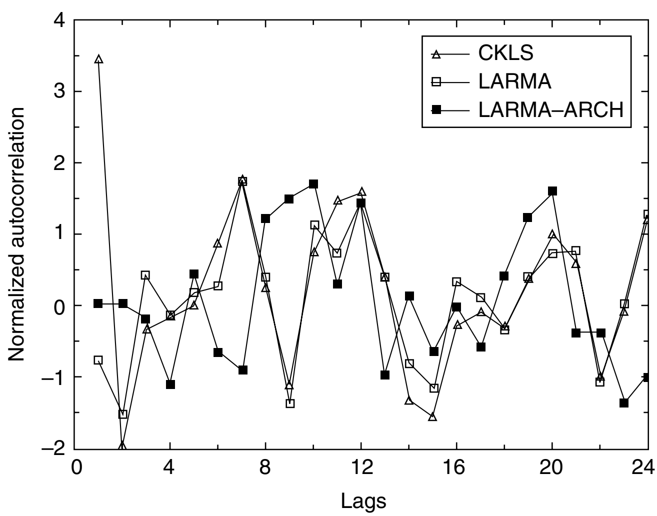


Figure 9.6 Normalized autocorrelation of the estimated noise, UK

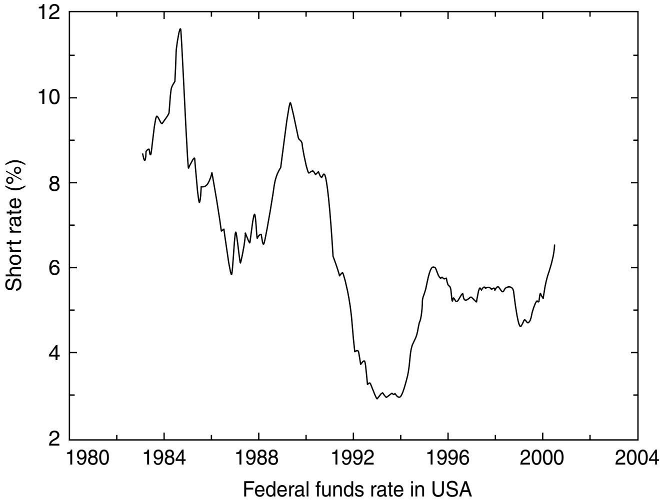


Figure 9.7 Federal funds rate of the USA

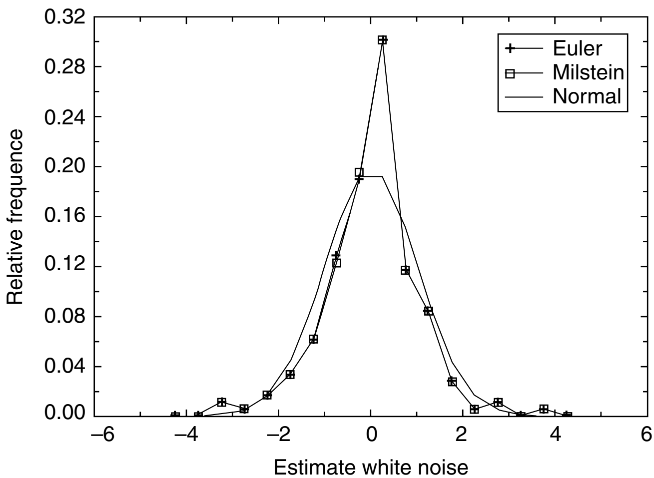


Figure 9.8 Distribution of estimated white noise (I), USA

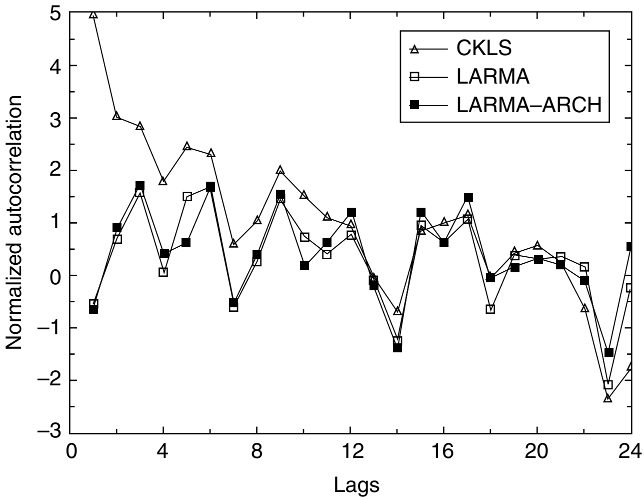


Figure 9.9 Normalized autocorrelation of the estimated noise, USA

9.4.1 Improvement in the continuous-time framework

In the literature, there are further works to improve the model (9.1) for modeling the short-term rate in the framework of continuous-time models. For example, Ait-Sahalia (1996) suggests a nonlinear drift coefficient:

$$dr_t = (\alpha_0 + \alpha_1 r_t + \alpha_2 r^2 + \frac{\alpha_3}{r_t})dt + \sqrt{\beta_0 + \beta_1 r_t + \beta_2 r_3^{\beta}}\, dW_t$$

and Andersen and Lund (1997) suggest a stochastic volatility model:

$$dr_t = \kappa_1(\mu - r_t)dt + \sigma_t r_t^{\gamma}\, dW_{1t},$$
$$d\log\sigma_t^2 = \kappa_2(\alpha - \log\sigma_t^2)dt + \xi dW_{2t}.$$

We simulate data using the models specified in Ait-Sahalia (1996) and Andersen and Lund (1997).[12] We plot them in Figures 9.10 and 9.11. The model of Ait-Sahalia cannot reproduce a similar time series of the real data. It stays always in a narrow band around the steady state. The normalized autocorrelation functions from these two models are plotted in Figure 9.12. We observe that there is no extreme autocorrelation in the estimated noise.

Table 9.1 Results of estimation and forecast for Germany

Germany	Milstein CKLS	Euler CKLS	Euler LALARMA	Euler LARMA-ARCH
Model Identification				
p	1	$\{1,6\}$	$\{1,6\}$	
q		0	0	0
k		0	0	$\{1,7\}$
$\alpha_0(c)$	0.020	0.017	0.068	0.065
(t-stat.)	(0.37)	(0.34)	(1.38)	(1.74)
$\alpha_1(-\beta)$	−0.007	−0.007	0.095	0.079
	(−0.73)	(−0.71)	(3.89)	(3.54)
α_6			−0.107	−0.091
			(−4.28)	(−4.01)
γ	0.417	0.378	0.186	0.485
	(2.21)	(2.01)	(1.00)	(2.14)
$c_0(\sigma)$	0.113	0.122	0.153	0.062
	(3.11)	(3.11)	(3.16)	(2.48)
c_1				0.297
				(2.43)
c_7				0.272
				(2.47)
Log-Likelihood	0.0054	0.0054	0.0061	0.0067
d-Statistics (χ^2)	161	160	95	31
(p-value)	($1.78e^{-25}$)	($2.91e^{-25}$)	($6.39e^{-13}$)	(0.02)
Avg. Forecast Errors				
- Level Forecast				
In Sample	0.0540%	0.0541%	0.0439%	0.0441%
(Relative)	(99%)	(99%)	(90%)	(90%)
Out of Sample	0.0192%	0.0193%	0.0153%	0.0156%
(Relative)	(100%)	(100%)	(79%)	(81%)
- Volatility Forecast				
In Sample	0.0144%	0.0144%	0.0090%	0.0082%
Out of Sample	0.0017%	0.0017%	0.0015%	0.0013%

9.4.2 Modeling autocorrelations in the estimated noise

We employ the ARMA process[13] to model the autocorrelation of the estimated noise

$$\Delta W_t = \sum_{i=1}^{p} \phi_i \Delta W_{t-i} + \sum_{j=0}^{q} \psi_i \varepsilon_{t-j}. \tag{9.11}$$

Table 9.2 Results of estimation and forecast for UK

United Kingdom	Milstein CKLS	Euler CKLS	Euler LALARMA	Euler LARMA-ARCH
Model Identification				
p	{1}	{1}	{1}	
q		0	{1}	{1}
k		0	0	{1}
$\alpha_0(c)$	0.153	0.155	0.289	0.210
(t-stat.)	(1.30)	(1.23)	(1.63)	(1.71)
$\alpha_1(-\beta)$	−0.018	−0.019	−0.034	−0.025
	(−1.25)	(−1.26)	(−1.67)	(−1.71)
β_1			0.431	0.313
			(5.38)	(2.18)
γ	0.974	0.742	0.574	0.527
	(4.97)	(3.45)	(2.91)	(2.21)
$c_0(\sigma)$	0.067	0.115	0.157	0.136
	(2.31)	(2.11)	(2.29)	(1.86)
c_1				0.498
				(2.31)
Log-Likelihood	0.00038	0.00019	0.00052	0.00091
d−Statistics (χ^2)	1639	1675	349	235
(p-value)	(0.00)	(0.00)	($9.59e^{-64}$)	($2.22e^{-40}$)
Avg. Forecast Errors				
- Level Forecast				
In Sample	0.3668%	0.3886%	0.3155%	0.3212%
(Relative)	(99%)	(99%)	(85%)	(87%)
Out of Sample	0.0701%	0.0701%	0.0777%	0.0714%
(Relative)	(105%)	(105%)	(116%)	(107%)
- Volatility Forecast				
In Sample	1.1705%	1.1503%	0.6469%	0.8050%
Out of Sample	0.0218%	0.0306%	0.0298%	0.0277%

If the noise ΔW_t has an autoregressive coefficient of order one then $\Delta W_t = \phi \Delta W_{t-1} + \varepsilon_t$.

By replacing ΔW_t using (9.7), we obtain

$$\frac{\Delta X_t - (c - \beta X_{t-1})}{\sigma X_{t-1}^{\gamma}} = \phi \frac{\Delta X_{t-1} - (c - \beta X_{t-2})}{\sigma X_{t-2}^{\gamma}} + \varepsilon_t.$$

Table 9.3 Results of estimation and forecast for USA

US	**Milstein CKLS**	**Euler CKLS**	**Euler LALARMA**	**Euler LARMA-ARCH**
Model Identification				
p	1	{1,2}	{1,2}	
q		0	0	0
k		0	0	{1,6}
$\alpha_0(c)$	0.048	0.047	0.055	0.028
(t-stat.)	(1.03)	(1.01)	(1.23)	(0.64)
$\alpha_1(-\beta)$	−0.010	−0.010	0.361	0.456
	(−1.22)	(−1.20)	(5.219)	(6.11)
α_2			−0.371	−0.461
			(−5.39)	(−6.29)
γ	0.827	0.839	0.767	0.808
	(5.70)	(5.74)	(5.25)	(3.57)
$c_0(\sigma)$	0.055	0.054	0.057	0.037
	(3.70)	(3.68)	(3.68)	(2.23)
c_1				0.225
				(1.26)
c_6				0.330
				(2.07)
Log-Likelihood	0.0050	0.0050	0.0054	0.0057
d-Statistics (χ^2)	65	63	76	36
(p-value)	($1.32e^{-7}$)	($3.43e^{-7}$)	($2.33e^{-9}$)	(0.0053)
Avg. forecast errors				
- Level forecast				
In sample	0.0732%	0.0732%	0.0614%	0.0618%
(Relative)	(99%)	(99%)	(82%)	(83%)
Out of sample	0.0252%	0.0252%	0.0190%	0.0187%
(Relative)	(102%)	(102%)	(77%)	(76%)
- Volatility forecast				
In sample	0.0256%	0.0256%	0.0183%	0.0178%
Out of sample	0.0020%	0.0020%	0.0017%	0.0014%

Rearranging it we obtain:

$$\Delta X_t = (c - \beta X_{t-1}) + \phi \left(\frac{X_{t-1}}{X_{t-2}}\right)^{\gamma} (X_{t-1} - (c - \beta X_{t-2})) + \sigma X_{t-1}^{\gamma} \varepsilon_t. \tag{9.12}$$

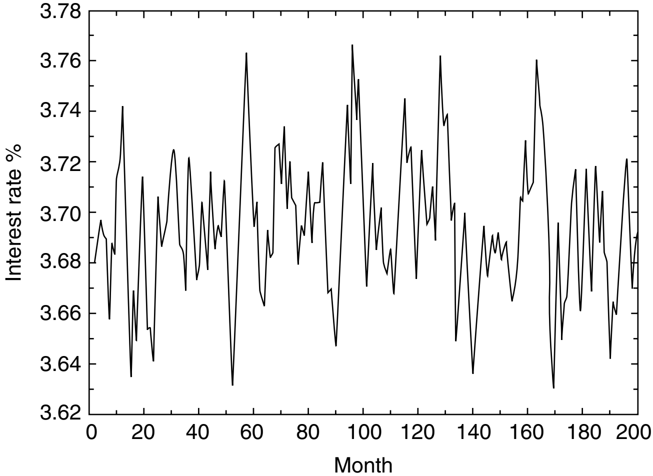


Figure 9.10 Simulated data from Ait-Sahalia's model

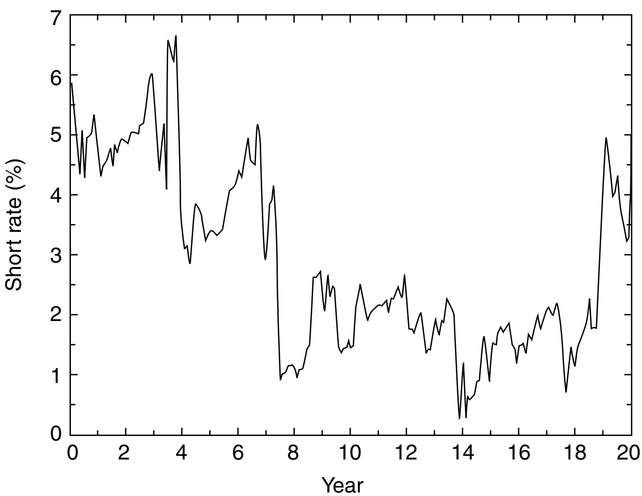


Figure 9.11 Simulated data from Andersen-Lund's model

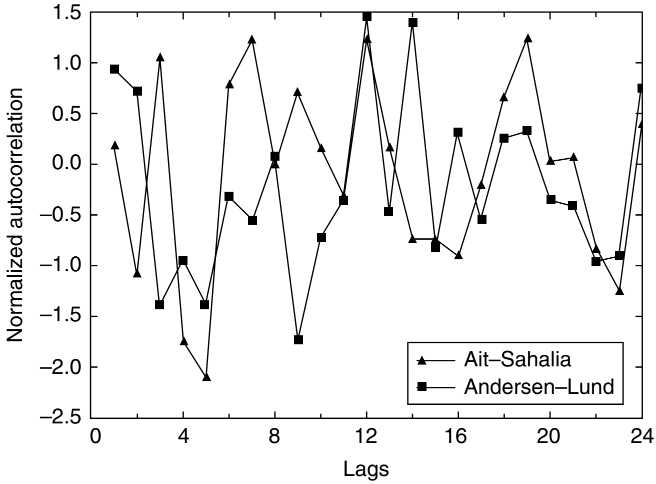


Figure 9.12 Normalized autocorrelation of the estimated noise for the continuous-time models

Rewriting (9.12) using the approximation $(\frac{X_{t-1}}{X_{t-2}})^{\gamma} \approx 1$, we obtain a model with two lags in the drift term

$$\Delta X_t = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \sigma X_{t-1}^{\gamma} \varepsilon_t.$$

So, the noise ΔW_t, with an autocorrelation of order one, gives us an model with two lags. Using this idea, we give the general structure as:

$$\Delta X_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i} + X_{t-1}^{\gamma} \left(\sum_{i=0}^{q} \beta_i \varepsilon_{t-i} \right). \tag{9.13}$$

9.4.3 Modeling thick tails in the estimated noise

For modeling thick tails of the noise we employ the idea of Brenner et al. (1996) and Koedijk et al. (1997). The common feature of their constructions is that they apply the ARCH[14] to model the thick tail. Moreover, the conditional variance (the volatility) of X_t is level-dependent. Brenner et al. (1996) argue that both level and ARCH effects are significant for short-term rates. We follow their idea and build the ARCH-structure into the model (9.13) where ε_t is $N(0, h_t)$ distributed with

$$h_t = c_0^2 + \sum_{i=1}^{k} c_i \varepsilon_{t-i}^2. \tag{9.14}$$

We employ (9.13) and (9.14) as our model class to model short rates. For the unique specification of the parameter we normalize $\beta_0 = 1$. We have a more general model than Brenner et al. (1996) by considering the ARMA structure (9.13). This can correct the autocorrelations of the residuals.[15] We note that we employ the ARCH structure instead of the GARCH structure in Brenner et al. (1996). The GARCH model is a technical improvement over the ARCH model[16] when the lags of ε_t^2 are long. The results during the model-identification stage suggest we do not need to employ the GARCH structure.

9.4.4 Results

In Tables 9.1, 9.2, and 9.3, we report the empirical results for the short rate of Germany, the UK, and the USA. The first and second columns are results of the CKLS model (1) using the Euler and the Milstein approximation as mentioned. The third and fourth columns report the results of the ARMA and the ARMA-ARCH models, respectively. The abbreviation "LARMA" denotes "level + ARMA" meaning the ARMA structure with a level effect.

The "Model Identification" in the tables means the determination of the orders p, q, and k in (9.13) and (9.14). We follow the standard procedure of Box et al. (1994). p and q are determined based on the autocorrelation function of ε_t. Next, k is chosen according to the autocorrelation function of ε_t^2. The procedure chooses the most parsimonious model where the estimated noise does not have significant autocorrelations.

The parameters of the drift coefficients α are not significant in the CKLS model (the t-statistics are quoted in parenthesis). However, they become significant after the ARMA and ARMA-ARCH structures are introduced in the third and fourth column (with the UK as an exception). All likelihood values increase when the ARMA structure is considered and are improved further when the ARCH components are introduced.

The forecast errors are reported in the tables for both the in-sample and the out-of-sample forecasts. According to (9.13), the predictor of X_{t+1} in the LARMA and the LARMA-ARCH model is given by:

$$\hat{X}_{t+1} = E_t[X_{t+1}](\hat{\theta}) = X_t + \hat{\alpha}_0 + \sum_{i=1}^{p} \hat{\alpha}_i X_{t-i+1} + X_t^{\hat{\gamma}} (\hat{\beta}_1 \varepsilon_t + \ldots + \hat{\beta}_q \varepsilon_{t-q+1}).$$

Thus, the forecast error of the level is the difference:

$$X_{t+1} - \hat{X}_{t+1} = X_t^{\hat{\gamma}} \varepsilon_{t+1}$$

and the forecast error of volatility is given by:

$$(X_{t+1} - \hat{X}_{t+1})^2 - E_t[(X_{t+1} - \hat{X}_{t+1})^2] = (X_t^{\hat{\gamma}} \varepsilon_{t+1})^2 - X_t^{2\hat{\gamma}} \hat{h}_t.$$

We observe in Tables 9.1, 9.2, and 9.3 that the in-sample and out-of-sample forecasts have improved, with the exception of the out-of-sample level forecast for the UK. We relate this improvement with the results that all estimates become significant after the introduction of the ARMA and ARMA-ARCH structure (again with the exception of the drift coefficients for the UK).

We can see that the major improvement of the level forecast is due to the introduction of the ARMA structure and the ARMA-ARCH structure contributes to the forecast improvement of volatility. The parameter γ is significantly different from zero for all three countries. This corresponds to the existence of the level effect in Brenner et al. (1996). For the data of Germany and the UK, the estimates of the parameter γ are not far from 0.5 as in the model of Cox et al. (1985).

The normalized autocorrelations with respect to the lags are plotted in Figures 9.3, 9.6, and 9.9. The normalized autocorrelations for the chosen LARMA and LARMA-ARCH models are controlled within $[-2, +2]$. The distribution of the noise can be found in Figures 9.13, 9.14, and 9.15, and the χ^2-statistics for the normality test are reported in Tables 9.1–9.3.

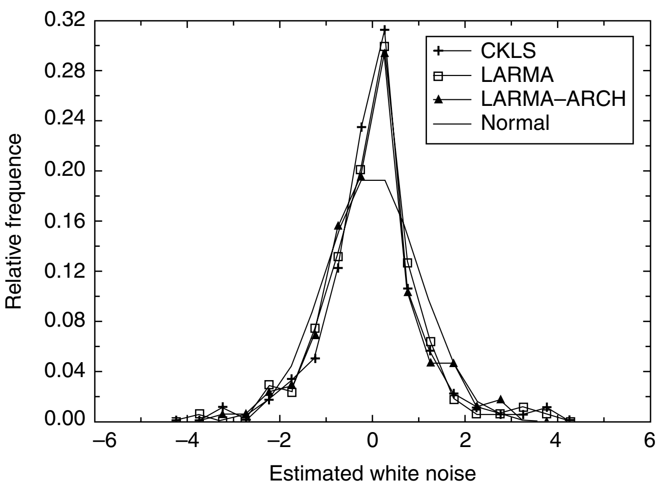


Figure 9.13 Distribution of estimated white noise (II), Germany

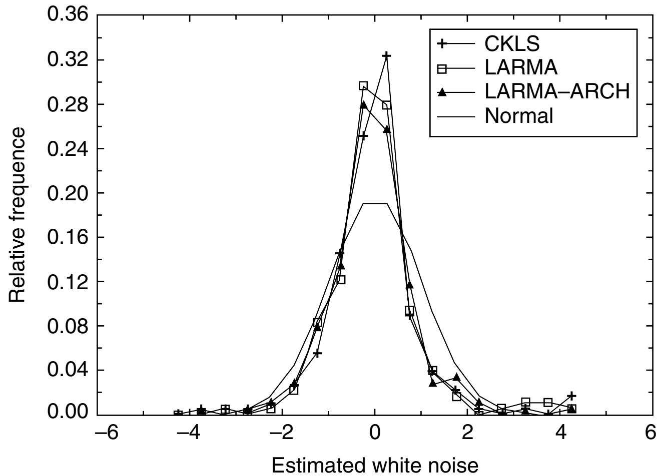


Figure 9.14 Distribution of estimated white noise (II), UK

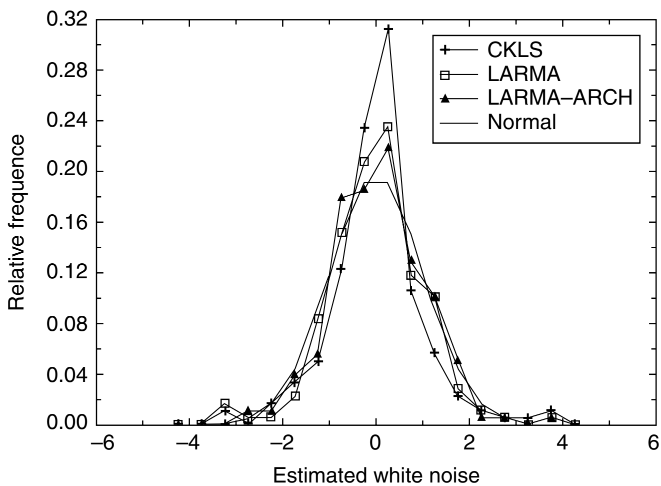


Figure 9.15 Distribution of estimated white noise (II), USA

Although we already have reduced the concentration of the distributions by introducing the ARCH structure, we still note that they are significantly different from the normal distribution at the 5 percent level. The distance is greatest for the short rate in the ÚK.

Comparing all three countries, we can observe that the modeling of the short rate for the UK is less successful. The t-statistics of the estimated parameters are not significantly different from zero, and the distance between the distribution of the estimated noise and the normal distribution is still sizeable, even after the introduction of the ARCH structure.

The parameters p,q, and k are the model identification parameters as given in Equation (9.13) and (9.14). In the parenthesis are t-statistics. d-Statistics is χ^2-distributed given by Equation (9.10). The relative in-sample forecast error in level is 99 percent of the error of the naive forecast. The naive forecast assumes the value in the next period is just that of today.

9.5 Conclusions

The objective of this chapter is to empirically model the short-term interest rates. We begin with the continuous-time CKLS model (9.1), and we apply the Euler, Milstein, and NLL approximations. For evaluating the quality of the discrete-time approximations, we compare the errors of the parameter estimations and the one-step-ahead predictions. Our results do not show an improvement of the NLL and Milstein approximations over the Euler approximation frequently found in the literature. The NLL approximation is equivalent to the Euler approximation due to the linearity of the drift coefficient. The Milstein and the Euler approximations behave similarly.

We apply the Euler and the Milstein approximations to the short-term interest rates of Germany, the UK, and the USA. It indicates evidence of model misspecification where the estimated residuals have high autocorrelation and thick tails. We show that two further continuous-time models of Ait-Sahalia (1996) and Andersen and Lund (1997) cannot sufficiently model the autocorrelation of the estimated white noise either. Therefore, we decide to model the short rates in a discrete-time framework. The model of the ARMA-ARCH structure with level-dependent volatility copes with the autocorrelation problem successfully. This model can also provide higher likelihood values and improve the level and volatility forecasts by a significant amount. However, the results regarding the distribution normality of the residuals display only a moderate success. In addition, the out-of-sample level forecasts of the UK data do not show marked improvements. This suggests one needs to broaden the framework and consider other models such as the multifactor and regime-switching models.

Notes

1. If the process deviates from $\frac{c}{\beta}$ (the mean), for example, $X_t > \frac{c}{\beta}$, then the process is drifting down and it is pulled up when $X_t < \frac{c}{\beta}$.
2. The application of the Milstein method for approximating diffusion processes is independently developed by the authors. In the appendix of this chapter we present our application and show that it is equivalent to that of Elerian (1998).
3. It means the ML estimator by using the Euler method.
4. See Kloeden and Platen (1992: 345).
5. See Kloeden and Platen (1992: Chap.10).
6. See http://www.ub.uni-bielefeld.de/english/library/databases/, then choose International Statistical YearBook 2000, for "Datenbank" choosing "OECD" and "main economic indicators", for "Period" choose "monthly data", for "Search" choose "indicator-search", then "interest rates", then "immediate rates".
7. See Gallant and White (1988) Def. 3.13, p. 27 with $Z_{ni} = U_i U_{i-k}$. One can see $v_m = 0$ when $m \geq k$.
8. See Gallant and White (1988), Theorem 5.3, p. 76. The conditions of the theorem are satisfied because under null (U_i) is independent and $v_n = n - k$.
9. See Breiman (1973: 189).
10. The is why the "Q-statistic" is developed, see Box and Pierce(1970) and Ljung and Box(1978).
11. Because the variance is normalized to 1. The concentration of the distribution around 0 leads to a smaller variance. In order to keep the variance as 1, there must be more weight in the tail.
12. We undertake simulation with an interval 0.01 and then pick up the simulated series with an interval, 1.
13. See Box, Jenkins and Reinsel (1994).
14. See Engle (1982).
15. See Brenner et al. (1996) p. 95 *"The Ljung-Box Q(ε_t/σ_t) statistics indicate that both models have significant serial correlation in the residuals."*
16. See Bollerslev (1986).

References

Ait-Sahalia, Y. (1996) "Testing Continuous-Time Models of the Spot Interest Rate," *Review of Financial Studies*, 9 (2): 385–426.

Ait-Sahalia, Y. (1997) "Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach," *Econometrica*, 70 (1): 223–262.

Ait-Sahalia, Y. (1999) "Transition Densities for Interest Rate and Other Nonlinear Diffusions," *Journal of Finance*, 54 (4): 1361–1395.

Andersen, T. G. and Lund, J. (1997) "Estimating Continuous-Time Stochastic Volatility Models of the Short-Term Interest Rate," *Journal of Econometrics*, 77 (1): 343–377.

Bollerslev, T. (1986) "Generalized Autoregressive Conditional Heteroscedasticity," *Journal of Econometrics*, 31 (3): 307–327.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994) *Time Series Analysis*, 3rd edn, Englewood Cliffs, NJ: Prentice Hall.

Box, G. E. P. and Pierce, D. A. (1970) "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models," *Journal of the American Statistical Association*, 65 (332): 1509–1526.

Breiman, L. (1973) *Statistics,* Boston, Mass.: Houghton Mifflin.

Brenner, R. J., Harjes, R. H., and Kroner, K. F. (1996) "Another Look at Models of the Short-Term Interest Rate," *Journal of Financial and Quantitative Analysis*, 31 (1): 85–107.

Chan, K. C., Karolyi, G. A., Longstaff, F. A., and Sanders, A. B. (1992) "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate," *Journal of Finance*, 47 (3): 1209–1227.

Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985) "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica*, 53 (2): 363–384.

Cox, J. C., Ingersoll, J. E. and Ross, S. A. (1985) "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53 (2): 385–407.

Dacunha-Castelle, D. and Florens-Zmirou, D. (1986) "Estimation of the Coefficients of a Diffusion from Discrete Observations," *Stochastics*, 19 (4): 263–284.

Duffie, D. and Glynn, P. (2004) "Estimation of Continuous-Time Markov Processes Sampled at Random Time Intervals," *Econometrica*, 72 (6): 1773–1808.

Durbin, J. (1970) "Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors Are Lagged Dependent Variables," *Biometrika*, 38 (3): 410–421.

Elerian, O. (1998) "A Note on the Existence of a Closed Form Conditional Transition Density for the Milstein Scheme," Nuffield College Economics Working Papers 1998-W18.

Engle, R. F. (1982) "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50 (4): 987–1007.

Frohn, J. (1995) *Grundausbildung in Oekonometrie*, 2nd edn, Berlin: Walter de Gruyter.

Fuller, W. A. (1996). *Introduction to Statistical Time Series*, 2nd edn, New York: Wiley.

Gallant, A. R. and Tauchen, G. (1996) "Which Moments to Match?" *Econometric Theory*, 12 (4): 657–681.

Gallant, A. R. and White, H. (1988) *A Unified Theory of Estimation and Inference for Nonlinear Dynamic System,* Oxford: Blackwell.

Gourieroux, C., Monfort, A., and Renault, E. (1993) "Indirect Inference," *Journal of Applied Econometrics*, 8 (S): S85–S118.

Hansen, L. A. and Scheinkman, J. A. (1995) "Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes," *Econometrica*, 63 (4): 767–804.

Karatzas, I. and Shreve, S. E. (1991) *Brownian Motion and Stochastic Calculus,* New York: Springer.

Koedijk, K. G., Nissen, F. G. J. A., Schotman, P. C. and Wolff, C. C. P. (1997) "The Dynamics of Short-Term Interest Rate Volatility Reconsidered," *European Finance Review*, 1 (1): 105–130.

Ljung, G. M. and Box, G. E. P. (1978) "On Measure of Lack of Fit in Time Series Models," *Biometrika*, 65 (2): 297–303.

Lo, A. W. (1988) "Maximum Likelihood Estimation of Generalized Ito Processes with Discretely Sampled Data," *Econometric Theory*, 4 (August): 231–247.

Pedersen, A. R. (1995). "A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations," *Scandinavian Journal of Statistics*, 22 (1): 55–71.

Shoji, I. and Ozaki, T. (1997) "Comparative Study of Estimation Methods for Continuous Time Stochastic Processes," *Journal of Time Series Analysis*, 18 (5): 485–506.

Shoji, I. and Ozaki, T. (1998) "Estimation for Nonlinear Stochastic Differential Equations by Local Linearization Method," *Stochastic Analysis and Applications*, 16 (4): 733–752.

Thompson, S. B. (2002) "Evaluating the Goodness of Fit of Conditional Distributions, with an Application to Affine Term Structure Models," Working Paper, Department of Economics, Harvard University.

Vasicek, O. (1977) "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5 (1): 177–188.

10
Testing the Expectations Hypothesis in the Emerging Markets of the Middle East: An Application to Egyptian and Lebanese Treasury Securities

Sam Hakim and Simon Neaime

10.1 Introduction

For many years, and despite many rejections,[1] the expectations hypothesis remains the widely accepted premise believed to explain the shape of the yield curve. In its simplest form, the expectations theory suggests that the current long-term interest rate is a weighted average of current and expected future short-term rates. In this setting, the spread between long- and short-term rates predicts future changes in short rates. Changes in the slope of the yield curve depend on interest expectations, with steeper yield curves foreboding greater expectations of rate changes.

This chapter focuses on testing the expectations hypothesis in two emerging capital markets. The theory assumes that securities of different maturities are substitutes and investors' arbitrage away yield spreads, which are caused by the relative excess supply or demand of a particular security over the term structure. Specifically, investors shift from one maturity sector to another in order to take advantage of yield differentials due to differences between expectations and forward rates. In this framework, the term structure is shaped by market expectations regarding the future direction of rates.

To evaluate the validity of the expectations hypothesis in the Middle East, we analyze the stochastic nature of interest rates representing the yields on securities for the entire maturity spectrum of the Egyptian and Lebanese term structures. To our knowledge, this is the first time that a study of the term structure in Middle Eastern countries has been undertaken despite the popularity of the region's fixed income instruments amongst mutual and hedge funds investing in emerging markets.[2] Our findings show that Egyptian and Lebanese interest rates are non-stationary and can be modeled as unit-root processes. Further investigation of their common relations showed that the interest rates in each country are bound by a cointegrating relation and that a unique common trend between them exists. This property suggests that Egyptian and Lebanese interest rates do not diverge consistently apart from one another and that a change in one interest rate is rapidly transmitted to the entire term structure. Overall, our results lend theoretical support to the expectations hypothesis and confirm the analysis of bond markets in more mature economies.

The remainder of this chapter is organized as follows. Section 10.2 briefly examines the empirical and theoretical literature on the term structure. Section 10.3 describes the data and provides descriptive statistics. The tests for unit roots are done in Section 10.4. We investigate the cointegrating relations between interest rates in each of the two countries and discuss the results in Section 10.5. Section 10.6 concludes this chapter.

10.2 The existing literature

The recent theoretical underpinnings of the term structure assume that the yield curve is represented by a stochastic process. Several models have evolved. In single state models, all yields, and correspondingly all discount bonds, are affected by movements in the short rate. Given the spot-rate dynamics and the structure of the market price of interest rates, default-free bonds of all maturities can be priced (Cox et al. 1985). Two factor models of the term structure were also developed by Brennan and Schwartz (1979). By and large, the interest rate is assumed to follow an Ornstein–Uhlenbeck – or mean-reverting – process, where the underlying distribution is normal. Unfortunately, this last property has the perverse implication that interest rates could also become negative, a likelihood largely mitigated by an appropriate choice of parameters of the stochastic process or by imposing certain boundary restrictions.

In general, the single-factor models are based solely on the initial short-term rate and overlook any other information on rates which can be imputed from the yield curve. Models that incorporate more information include Heath et al. (1992), Ritchken and Sankarasubramanian (1995), and Hull and White (1996).

The empirical literature on the term structure includes Engel et al. (1987) who pioneered the use of autoregressive conditional heteroscedastic models in the analysis of interest-rate series. This versatility allows for the conditional variance to affect the excess holding yield on a long-term bond. Their approach leads to a time-varying premia on securities of different maturities and a relaxation of the assumption of constant heteroscedasticity in the disturbances, based on earlier results obtained by Shiller (1979) and Campbell and Shiller (1987).

Historical support for the expectations hypothesis in the USA is documented in Bradley and Lumpkin (1992) and Hall et al. (1992), who find the rates for treasury securities cointegrated with unit roots.[3] Similar conclusions drawn from the analysis of British and Danish data is found in McDonald and Speight (1988), Mills (1991), Lee and Tse (1991), Taylor (1992), and Engsted and Tangaard (1994). More recently, Beechey et al. (2009) also found evidence of the expectations hypothesis in eight developed and six emerging economies. Nevertheless, there is also mixed evidence documented in Boothe (1991), Hall et al. (1992), Zhang (1993), and Lardic and Mignon (2004). The academic attention to interest rates in the Middle East is small but growing. Instead of interest rates, studies have focused on the region's thriving equity markets and the more traditional currency exchange rates. For example, Hammoudeh et al. (2009) examine the co-movements among the prices of four strategic commodities and their causal relationships with interest and exchange rates. The goal is to shed some light on the predictive behavior of those individual commodity prices relative to the selected financial variables and to establish a transmission link between commodity prices and the dollar exchange rate. Another set of studies examined the spread on sovereign bonds and how these measures are influenced either by the geopolitical landscape of the region (Haddad and Hakim 2007, 2009) or by US macroeconomic news (Ozatay et al. 2009).

10.3 Data and diagnostic statistics

Our data consists of the weekly yields on Egyptian treasury securities auctioned between July 21, 2006, and April 3, 2009, in five maturity sectors: one year, two years, three years, five years, and seven years. Treasury securities have only recently been auctioned in Egypt, and, therefore, there is a lot of interest to determine the efficiency of the term structure for that country. For Lebanon, our data is based on monthly observations since these securities were first introduced in 1991 as a tool of monetary policy. The data period for Lebanon covers October 1991 through to February

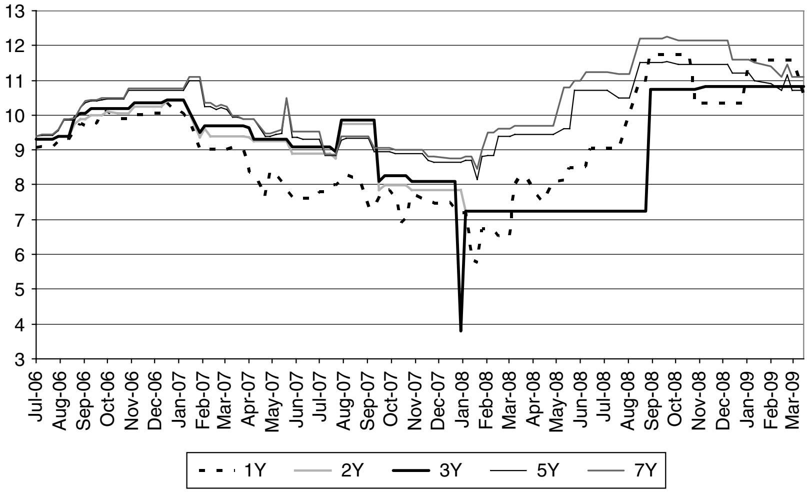


Figure 10.1 Egyptian treasury securities

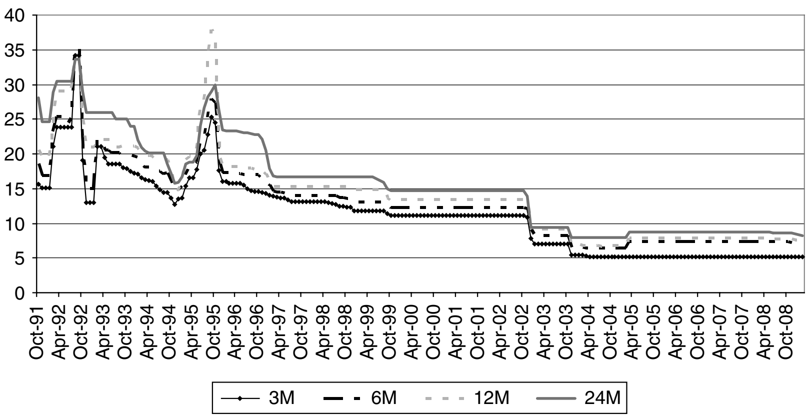


Figure 10.2 Lebanese treasury securities

2009. The Lebanese treasury securities are auctioned by Banque du Liban, the Lebanese central bank, and their yield is determined in a competitive environment. The figures were obtained from the quarterly reports published by the central bank. In both countries, the subscribers to the treasury auctions include commercial banks, financial institutions, public agencies, and private (domestic and foreign) buyers, most notably mutual funds that invest in emerging markets. Figures 10.1 and 10.2

Table 10.1 Descriptive statistics yields on Egyptian and Lebanese treasury securities

	Lebanon				Egypt				
	3M	**6M**	**1Y**	**2Y**	**1Y**	**2Y**	**3Y**	**5Y**	**7Y**
Mean	11.38	12.87	14.17	15.65	8.94	9.08	9.13	10.00	10.24
Median	11.18	12.12	13.43	14.64	9	9.35	9.4	9.88	9.97
Mode	5.22	7.24	7.75	8.68	9	7.25	7.25	10.72	10.76
Standard deviation	5.61	5.60	6.42	6.73	1.44	1.30	1.39	0.89	1.03
Kurtosis	1.53	1.50	1.48	−0.51	−0.65	−1.38	0.20	−1.17	−0.88
Skewness	0.95	1.03	1.09	0.64	0.18	−0.23	−0.71	0.11	0.39
Range	29.08	28.97	31.17	25.7	5.98	3.57	7.03	3.4	3.8
Minimum	5.1	6.31	6.68	7.89	5.75	7.25	3.79	8.15	8.45
Maximum	34.18	35.28	37.85	33.59	11.73	10.82	10.82	11.55	12.25
No. of observations	209	209	209	209	120	120	120	120	120

Egypt: February 2006–April 2009 (weekly)
Lebanon: October 1991–April 2009 (monthly)

show the behavior of interest rates from the two countries over the study periods.

The descriptive statistics of the term structure in each country are provided in Table 10.1. The mean interest rate during the study period ranged between 11.38 percent and 15.65 percent for Lebanon and 8.94 percent and 10.24 percent for Egypt. It appears that the term structure in each country was upward-sloping as the interest rates increased with longer maturity, an observation that is consistent with the normal shape of the yield curve. The distribution of each maturity class over time is not normal as is clear from the skewness statistic reported in Table 10.1 (a normal distribution has 0 skewness). The Lebanese interest rates are skewed to the right while the Egyptian interest rates are mixed. Looking at the standard deviation of the interest rates over time in relation to their mean, we find that the two-year security in Lebanon has offered its holders the lowest risk per unit of return at 0.43 (= 15.65/6.73). The best comparable rate in Egypt is the five-year treasury security at 0.09 (= 0.89/10).

10.4 Tests of unit roots in Lebanese and Egyptian interest rates

We begin by examining the stochastic nature of the yields offered on the treasury securities across the Egyptian and Lebanese term structures. We show that each set of series represents a unit-root process which we assume can be written as:

$$y_t = \beta + \alpha y_{t-1} + u_t \tag{10.1}$$

where β is a drift parameter. The process can be rewritten in first difference form as:

$$\Delta y_t = \beta + (\alpha - 1)y_{t-1} + u_t \tag{10.2}$$

The test for unit root is essentially one which tests for $\alpha = 1$. If the errors are assumed to follow an AR (1) process

$$u_t = \rho u_{t-1} + \varepsilon_t \tag{10.3}$$

the model can be rewritten as:

$$\Delta y_t = Z_t \beta_+ (\alpha - 1)(\rho - 1)y_{t-1} + \alpha\rho\Delta y_{t-1} + \varepsilon_t \tag{10.4}$$

where Z_t is the drift term. The new test is an Augmented Dickey Fuller ($\hat{\tau}$) test of $(\alpha - 1)(\rho - 1) = 0$. Elliott et al. (1996) perform asymptotic power calculations and show that the modified DF-GLS test can achieve substantial power gain over the testing procedure outlined in (10.4). In the DF-GLS test, we consider the same null hypothesis on the *t*-ratio tests but critical values change as reported in Elliott et al. (1996). To see whether our results are sensitive to serial correlation in the disturbances, we ran the tests with AR (2), AR (3), and AR (4) errors[4] without a change in our results. In the end, we base the model selection on Schwartz's criterion.[5]

We also employ the KPSS test as developed in Kwiatkowski (1992) with bandwidth set at 4 and Bartlett kernel. For the KPSS test, we perform estimation with the modified Akaike information criterion (AIC).

The test statistics for all variables reported in Table 10.2 clearly reject the null hypothesis of unit root in the differences but not in the levels. This leads us to believe that each variable is an I(1) process, which becomes stationary after differencing it once.

Overall, our results confirm the analysis of the bonds market in other countries, notably Lee and Tse (1991), Mills (1991), Taylor (1992), Engsted and Tangaard (1994), Chong et al. (2006), Kleimeier and Sander (2006), De Graeve et al. (2007), and Liu et al. (2008).

The preceding findings suggest that the Lebanese and Egyptian interest rates can be modeled as unit-root processes. As such, the variance for each interest-rate series goes to infinity as the sample size increases. Furthermore, from equation (10.1), any innovation u_t has a permanent effect on the value y_t which can be written as a sum of all previous innovations.

Table 10.2 Unit-root tests for interest-rate spreads of Egyptian and Lebanese treasury securities

		ADF-GLS (2 lags)		KPSS (4 lags)	
Egypt					
Spreads with 1Y Treasury					
	2Y–1Y	−2.48		0.28	**
	3Y–1Y	−2.75	*	0.28	**
	5Y–1Y	−2.62	*	0.4	**
	7Y–1Y	−2.52		0.34	**
Lebanon					
Spreads with 3m Treasury					
	6m–3m	−1.33		0.72	**
	12m–3m	−3.5	**	1.27	**
	24m–3m	−0.72		2.08	**

Significant at 10% (*) or 1 percent (**)

10.5 Testing the expectations hypothesis and cointegration analysis

In this section, we turn our attention to the expectations hypothesis in the term structure of each country. To that end, we begin with a simple model of the expectations hypothesis of the term structure which can be written as:

$$y_{m,t} = \frac{1}{m} \cdot \sum_{k=0}^{m-1} E_t(y_{1,t+k}) + \Lambda_m \tag{10.5}$$

where $y_{m,t}$ represents the yield on a pure discount treasury bond with maturity *m*, E(.) is the expectations operator, and Λ_m is the term premium. Equation (10.5) suggests that the yield at time *t* of a pure discount bond with *m* maturity can be written as the average expected yield on *m* future bonds. In our case, we consider four specific yields with time to maturity equal to three, six, twelve, and twenty-four months in Lebanon and five specific yields for Egypt with maturities of one year, two years, three years, five years, and seven years. These interest rates enable us to construct three possible pairs for Lebanon and four pairs for Egypt that obey equation (10.5) and define stationary interest-rate spreads. These are:

$$\text{Lebanon}: \{y_{6m}, y_{3m}\}, \{y_{12m}, y_{3m}\}, \{y_{24m}, y_{3m}\}$$

$$\text{Egypt}: \{y_1, y_2\}, \{y_1, y_3\}, \{y_1, y_5\}, \{y_1, y_7\}$$

Based on the results of the preceding section suggesting the unit-root nature of the series, we now examine their cointegrating relationship. Let Y_t represents a $k \times 1$ vector of non-stationary I(1) time series. A $k \times r$ matrix of cointegrating vectors γ is said to exist if the linear combination $\gamma' \; Y_t$ is stationary or I(0). Each cointegrating vector suggests the existence of a long-term relationship between the series, namely an equilibrium.

We define $Y_{Lebanon} \equiv [y_{3m}, y_{6m}, y_{12m}, y_{24m}]'$ as the vector of four interest rates composing the Lebanese term structure. The expectations hypothesis asserts that if each component of Y_t is I(1), then the spreads defined as $S_i(Lebanon) \equiv y_i - y_{3m}$ for $i = 6m,\ 12m,\ 24m$ may be I(0), and, hence, any two yields must be cointegrated with cointegrating vectors determined by the spread vectors. A similar vector Y_{Egypt} is defined for the Egyptian term structure with the corresponding spreads $S_j(Egypt) \equiv y_j - y_{1y}$ for $j = 2y,\ 3y,\ 5y,$ *and* $7y$. Table 10.3 reports the ADF-GLS and KPSS tests for unit roots applied to the interest-rate spreads in each country. The results of ADF-GLS are somewhat mixed but hold more for the intermediate term spreads. For example, the Egyptian 7y–1y interest-rate spread is non-stationary. The same applies to the 2y–1y spread for Egypt and the 24m–3m and the 12m–3m for Lebanon. The

Table 10.3 Unit-root tests for yields on Egyptian and Lebanese treasury securities*

	ADF–GLS (2 lags)			KPSS (4 lags)		
Series	**Levels**	**1st differences**		**Levels**		**1st differences**
Egypt						
1M	−1.12	−5.74	**	0.47	**	0.08
1Y	−1.47	−5.47	**	0.51	**	0.08
2Y	−1.41	−6.18	**	0.37	**	0.09
3Y	−1.45	−7.48	**	0.36	**	0.08
5Y	−1.30	−6.53	**	0.47	**	0.14
7Y	−1.29	−6.60	**	0.47	**	0.15
Lebanon						
3M	−1.30	−8.93	**	3.72	**	0.02
6M	−1.08	−5.06	**	3.70	**	0.02
1Y	−1.29	−7.30	**	3.56	**	0.02
2Y	−0.22	−3.35	**	3.76	**	0.05

Significant at 10 percent (*) or 1 percent (**)
Egypt: 1-month, 1-year, 2-year, 3-year, 5-year, and 7-year maturity
Lebanon: 3-month, 6-month, 1-year, and 2-year maturity

KPSS tests are more consistent and show that all interest-rate spreads are stationary in both countries.

When the expectations hypothesis holds, the spreads between the long-term rate and the short-term rate provide information about the future level of the short-term rate. However, if the fluctuations in the short-term rate are unpredictable, the spread between long-term and short-term rates cannot provide useful information about the market's expectation for the short-term rate unless the two rates share a cointegrating relation. Specifically, having established that the interest rates are unit roots, they can be stationary in levels in the direction of the cointegrating vector.

We now turn our attention to the cointegration analysis. This is important because the lack of cointegration represents strong evidence against the expectations hypothesis as there is no stable long-run relationship between interest rates of different maturities.

To test for cointegration, we consider a multivariate model of the form:

$$Y_t = \mu + A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_q Y_{t-q} + \varepsilon_t \quad t = 1, 2, \cdots, T \qquad (10.6)$$

where each Y_t is an N-dimensional vector of yields defined above and ε_t's are independent k-dimensional Gaussian errors with covariance matrix Σ. N represents the number of yields in each country (four for Lebanon, and five for Egypt). The model can be written in difference form as:

$$\Delta Y_t = \mu + \Pi_1 \Delta Y_{t-1} + \Pi_2 \Delta Y_{t-2} + \cdots + \Pi_q \Delta Y_{t-q} + \Phi Y_{t-q} + \varepsilon_t \qquad (10.7)$$

where μ is a linear deterministic trend,

$$\Pi_i \equiv -I + A_1 + A_2 + \cdots + A_i \qquad (10.8)$$

$$\Phi \equiv -I + A_1 + A_2 + \cdots + A_q \qquad (10.9)$$

and I is the identity matrix. Note that equation (10.7) is equivalent to a q dimensional VAR except for the term ΦY_{t-q}. Note also that the inclusion of the term ΦY_{t-q} makes equation (10.7) an error-correction model, where the matrix Φ contains the information about the cointegrating relations. If there are r cointegrating vectors then Φ can be expressed as:

$$\Phi = \varphi \gamma' \qquad (10.10)$$

where φ and γ are *kxr* matrices of rank *r*, $0 < r < k$, and γ is a matrix of cointegrating vectors. Therefore, the rank[6] of Φ is also *r*. Johansen (1988) and Johansen and Juselius (1990) maximize the likelihood function for Y_t conditional on the restrictions $\Phi = \varphi\gamma'$. The likelihood ratio (LR) is derived by applying least squares on the following equations:

$$Y_{t-q} = \mu + V_1 \Delta Y_{t-1} + V_2 \Delta Y_{t-2} + \cdots + V_{q-1} \Delta Y_{t-q+1} + u_{0t} \quad (10.11)$$

$$\Delta Y_t = \mu + W_1 \Delta Y_{t-1} + W_2 \Delta Y_{t-2} + \cdots + W_{q-1} \Delta Y_{t-q+1} + u_{1t} \quad (10.12)$$

and computing the residual sample second-moment matrices:

$$\begin{aligned} \hat{\Omega}_{00} &= T^{-1} \sum_{t=1}^{T} \hat{u}_{0t} \hat{u}'_{0t} \\ \hat{\Omega}_{11} &= T^{-1} \sum_{t=1}^{T} \hat{u}_{1t} \hat{u}'_{1t} \\ \hat{\Omega}_{10} &= T^{-1} \sum_{t=1}^{T} \hat{u}_{1t} \hat{u}_{0t} \end{aligned} \quad (10.13)$$

where u_0 and u_1 are i.i.d. Under the null hypothesis $H(r)$ that there are *r* cointegrating vectors against the alternative of no cointegrating vectors, the LR statistic, can be written as:[7]

$$-2 \ln Q(H(r)|H_0) = -T \sum_{i=r+1}^{n} \ln(1 - \hat{\lambda}_i) \quad (10.14)$$

This is referred to as the λ-trace test, where $\hat{\lambda}_{r+1}, \ldots, \hat{\lambda}_n$ are the $n - r$ smallest eigenvalues that solve the detrimental equation:

$$|\lambda \hat{\Omega}_{11} - \hat{\Omega}_{10} \hat{\Omega}_{00}^{-1} \hat{\Omega}_{01}| = 0 \quad (10.15)$$

The λ-trace test statistics are provided in Table 10.4. The results reveal the existence of a single cointegrating vector for Lebanon and three cointegrating vectors for Egypt. Among the Lebanese treasury securities, there is a single common stochastic trend, and there are two among the Egyptian securities. The number of trends is simply the difference between the number of interest variables and the number of cointegrating vectors. Generally, with *n* yields to maturity, it is possible to form $n - 1$ linearly independent yield spreads. The yields are I(1) cointegrated processes and the spreads between them represent cointegrating relations. If the expectations hypothesis is valid, this leaves three common

Table 10.4 Tests of cointegration of interest rates

Lebanon Rank	Eigenvalue	Trace test	p-value	Lmax test	p-value
0	0.413	166.28	0.0000	110.29	0.0000
1	0.192	55.993	0.0000	44.341	0.0000
2	0.0491	11.653	0.1765	10.44	0.1878
3	0.0058	1.2131	0.2707	1.2131	0.2707
Cointegration vector: (12m) + 0.40(3m) − 1.31(6m) − 0.19(24m) = −1.04					
Egypt Rank	**Eigenvalue**	**Trace test**	**p-value**	**Lmax test**	**p-value**
0	0.366	129.80	0.000	53.732	0.000
1	0.245	76.068	0.000	33.125	0.0065
2	0.191	42.944	0.0007	25.021	0.0114
3	0.100	17.923	0.0195	12.447	0.0945
Cointegration vector: (1y) − 0.42(2y) + 0.04(3y) − 0.66(5y) − 0.39(7y) = −5.10					

Egypt: 1-month, 1-year, 2-year, 3-year, 5-year, and 7-year maturity
Lebanon: 3-month, 6-month, 1-year, and 2-year maturity
Egypt: February 2006–April 2009 (weekly)
Lebanon: October 1991–April 2009 (monthly)

non-stationary I(1) factors for Lebanon and two for Egypt that represent exogenous elements to the system of yields, as, for example, the growth in monetary aggregates which derives from the central bank's monetary policy. The evidence in support of a unique common factor is found in several studies (Engle and Granger 1987; Stock and Watson 1988).

The cointegrating vectors for Egypt and Lebanon are:

Lebanese yields cointegrating vector:

$$(12m) + 0.40\ (3m) - 1.31\ (6m) - 0.19\ (24m) = -1.04$$

Egyptian yields cointegrating vector:

$$(1y) - 0.42(2y) + 0.04(3y) - 0.66(5y) - 0.39(7y) = -5.10$$

The preceding results suggest that Lebanese and Egyptian bond yields do not drift apart from one another indefinitely. Even though each interest rate is stochastic and non-predictable, the rates are linked together by a stable and long-term relation. They move together over time, a basic feature of the expectations theory.

10.6 Conclusion

This chapter tested the expectations hypothesis and found it to hold in two emerging bond markets. We investigated in detail the time-series properties of interest rates offered on distinct securities forming the entire Egyptian and Lebanese term structures. The results showed that long- and short-term interest rates are characterized as unit-root processes with stationary spreads. This suggests that interest rates in each country are well behaved in the sense that they do not diverge consistently and permanently from one another. Further analysis showed that it is appropriate to model the Egyptian and Lebanese term structures as a cointegrated system. The results pointed to the existence of three cointegrating vectors for Egypt and a single vector for Lebanon, a property that suggests that the liquidity premia are stationary across the maturity spectrum. Overall, our results offer hope for the development of capital markets in other infant emerging economies, particularly in the Middle East.

Notes

1. See Froot (1989) for a historical recap of the expectations hypothesis.
2. See Russian, Kazakh, Middle East bonds fill bulging pipeline." Euroweek (September 15, 2006): 25–25.
3. Campbell (1995) reviews the latest literature on the term structure, and Crockett (1998) examines the assumptions underlying the expectations hypothesis.
4. MacKinnon (1990) provides a more complete set of critical values for $\hat{\tau}$. A non-parametric analog to $\hat{\tau}$ which allows for a wide range of serial correlation and heterogeneity patterns was suggested by Phillips (1987a and 1987b) and Phillips and Perron (1988). The critical values for the statistic are presented in Phillips and Ouliaris (1990). However the finite sample properties of the statistic are not fully investigated.
5. Schwartz's (1978) criterion is defined as: $BIC = \log|\hat{\Sigma}| + (p^2 q)T^{-1}\log T$ where p is the dimension of the VAR; q the order of the VAR model being tested for q = m against q = m − 1; m is the maximum order considered; and $\hat{\Sigma}$ is the estimated covariance of the errors. In effect, BIC includes a penalty adjustment for the number of estimated parameters. Note that Schwartz's (BIC) is similar to Akaike's Information Criterion (AIC) with the distinction that the former can be shown to be strongly consistent.
6. Without arbitrary constraints it is impossible to uniquely identify the elements of $\phi\gamma$ 'because for any rxr matrix Γ, we have $(\varphi\Gamma^{-1})(\Gamma\gamma') = \varphi\gamma'$.
7. The test statistic has an asymptotic χ^2 distribution with r(v-w) degrees of freedom where r is the number of cointegrating vectors, v the number of variables in the VAR system, and w is the number of variables left in the VAR system after testing the restrictions of the cointegrating relations. Note that if r = w,

then the null hypothesis places no restrictions on the cointegrating relations, whereas if $w = r$, all the VAR system variables need to be I(0) under the null. In general we have $r \leq w \leq v$.

References

Boothe, P. (1991) "Interest Parity, Cointegration, and the Term Structure in Canada and the United States," *Canadian Journal of Economics*, 24 (3): 499–516.

Beechey, M., Hjalmarsson, E., and Osterholm, P. (2009) "Testing the Expectations Hypothesis When Interest Rates Are Near Integrated," *Journal of Banking and Finance*, 33 (5): 934–943.

Bradley, M. G. and Lumpkin, S. A. (1992) "The Treasury Yield Curve as a Cointegrated System," *Journal of Financial and Quantitative Analysis*, 27 (3): 449–463.

Brennan, M. and Schwartz, E. (1979) "A Continuous Time Approach to the Pricing of Bonds," *Journal of Banking and Finance*, 3 (2): 133–155.

Campbell, J. Y. (1995) "Some Lessons from the Yield Curve," *Journal of Economic Perspectives*, 9 (3): 1986–1998.

Campbell, J. Y. and Shiller, R. J. (1987) "Cointegration and Tests of Present Value Models," *Journal of Political Economy*, 95: 1062–1088.

Chong, B. S., Liu, M.-H., and Shrestha, K. (2006) "Monetary Transmission Via the Administered Interest Rate Channels," *Journal of Banking and Finance*, 30 (5): 1467–1484.

Cox J., Ingersoll, J., and Ross, S. (1985) "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53 (2): 385–407.

Crockett, J. A. (1998) "Rational Expectations, Inflation and the Nominal Interest Rate," *Journal of Econometrics*, 83 (1–2): 349–363.

De Graeve, F., De Jonghe, O., and Vander Vennet, R. (2007) "Competition, Transmission and Bank Pricing Policies: Evidence from Belgian Loan and Deposit Markets," *Journal of Banking and Finance*, 31: 259–278.

Engel, R. F. and Granger, W. J. (1987) "Co-Integration and Error Correction: Representation Estimation and Testing," *Econometrica*, 55: 251–276.

Engel, R. F., Lilien, D. M., and Robins, R. P. (1987) "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model," *Econometrica*, 55 (2): 391–407.

Engsted, T. and Tanggaard, C. (1994) "Cointegration and the US Term Structure," *Journal of Banking and Finance*, 18 (3): 167–181.

Froot, K. A. (1989) "New Hope for the Expectations Hypothesis of the Term Structure of Interest Rates," *Journal of Finance*, 44 (2): 283–305.

Haddad, M. and Hakim, S. (2008) "The Impact of War and Terrorism on Sovereign Risk in the Middle East," *Journal of Derivatives and Hedge Funds*, 14: 237–250.

Haddad, M. and Hakim, S. (2007) "The Cost of Sovereign Lending in the Middle East after September 11," *Journal for Global Business Advancement*, 1 (1): 127–139.

Hall, A. D., Anderson, H. M., and Granger, C. W. (1992) "A Cointegration Analysis of Treasury Bill Yields," *The Review of Economics and Statistics*, 74 (2): 117–126.

Hammoudeh, S., Sari, R., and Ewing, B. (2009) "Relationships among Strategic Commodities and with Financial Variables: A New Look," *Contemporary Economic Policy*, 27 (2): 251–264.

Heath, D., Jarrow, R., and Morton A. (1992) "Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation," *Econometrica*, 60 (2): 77–105.

Hull, J. and White, A. (1996) "Using Hull–White Interest Rate Trees," *Journal of Derivatives*, 3 (3): 26–36.

Jarque, C. and Bera, A. (1980) "Efficient Tests for Normality Homoskesdasticity and Serial Independence of Regression Residuals," *Econometric Letters,* 6 (6): 255–259.

Johansen, S. (1988) "Statistical Analysis of Cointegrating Vectors," *Journal of Economics Dynamics and Control*, 12 (3): 231–254.

Johansen, S. (1991) "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Auto Regressive Models," *Econometrica*, 59 (6): 1551–1580.

Johansen, S. and Juselius, K. (1990) "Maximum Likelihood Estimation and Inference on Cointegration with Application to the Demand for Money," *Oxford Bulletin of Economics and Statistics*, 52 (4): 169–210.

Johansen, S. and Juselius, K. (1992) "Testing Structural Hypotheses in a Multivariate Cointegration Analysis of the PPP and the UIP for UK," *Journal of Econometrics*, 53 (3): 211–244.

Kleimeier, S. and Sander, H. (2006) "Expected Versus Unexpected Monetary Policy Impulses and Interest Rate Pass-through in Euro-Zone Retail Banking," *Journal of Banking and Finance*, 30 (2): 1839–1870.

Lardic, S. and Mignon, V. (2004) "Fractional Cointegration and the Term Structure," *Empirical Economics*, 29 (1): 723–736.

Lee, T. K. and Tse, Y. K. (1991) "Term Structure of Interest Rates in the Singapore Asian Dollar Market," *Journal of Applied Econometrics,* 6 (3): 143–152.

Liu, M.-H., Margaritis, D., and Tourani-Rad, A. (2008) "Monetary Policy Transparency and Pass-Through of Retail Interest Rates," *Journal of Banking and Finance*, 32 (4): 501–511.

Macdonald, R. and Speight, A. (1988) "The Term Structure of Interest Rates in the UK," *Bulletin of Economic Research*, 40 (1): 287–299.

Mackinnon, J. G. (1991) "Critical Values for Cointegration Tests," UC San Diego Discussion Paper 90-4, San Diego.

Mills, T. C. (1991) "The Term Structure of UK Interest Rates: Tests of the Expectations Hypothesis," *Applied Economics*, 23 (3): 599–606.

Ozatay, F., Ozmen, E., and Sahinbeyoglu, G. (2009) "Emerging Market Sovereign Spreads, Global Financial Conditions and US Macroeconomic News," *Economic Modelling*, 26 (2): 526–531.

Pesando, J. (1978) "On the Efficiency of the Bond Market: Some Canadian Evidence," *Journal of Political Economy*, 86 (4): 1057–1076.

Phillips, P. C. B. (1987) "Time Series Regression with a Unit Root," *Econometrica,* 55 (6): 227–301.

Phillips, P. C. B. (1987) "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika,* 74 (6): 535–547.

Phillips, P. C. B. and Ouliaris, S. (1988) "Testing for Cointegration Using Principal Components Methods," *Journal of Economic Dynamics and Control*, 12 (3): 205–230.

Phillips, P. C. B. and Perron, P. (1988) "Testing for Unit Roots in Time Series Regression," *Biometrika*, 75 (1): 335–346.

Ritchken, P. and Sankarasubramanian, L. (1995) "Volatility Structures of Forward Rates and the Dynamics of the Term Structure," *Mathematical Finance*, 3 (1): 55–72.

Shiller, R. (1979) "The Volatility of Long-Term Interest Rates and the Expectations Model of the Term Structure," *Journal of Political Economy*, 87 (3): 1190–1219.

Stock, J. and Watson, M. (1988) "Testing for Common Trends," *Journal of the American Statistical Association*, 83 (5): 1097–1107.

Taylor, M. P. (1992) "Modeling the Yield Curve," *The Economic Journal*, 102 (4): 524–537.

Zhang, H. (1993) "Treasury Yield Curves and Cointegration," *Applied Economics*, 25 (7): 361–367.

Index