Roberto Dieci · Xue-Zhong He
Cars Hommes *Editors*

Nonlinear Economic Dynamics and Financial Modelling

Essays in Honour of Carl Chiarella

Springer

Nonlinear Economic Dynamics and Financial Modelling

Roberto Dieci · Xue-Zhong He
Cars Hommes
Editors

Nonlinear Economic Dynamics and Financial Modelling

Essays in Honour of Carl Chiarella

Springer

Editors
Roberto Dieci
Department of Mathematics
University of Bologna
Bologna
Italy

Cars Hommes
Amsterdam School of Economics
University of Amsterdam
Amsterdam
The Netherlands

Xue-Zhong He
UTS Business School
University of Technology, Sydney
Sydney, NSW
Australia

ISBN 978-3-319-07469-6 ISBN 978-3-319-07470-2 (eBook)
DOI 10.1007/978-3-319-07470-2
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014944539

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The MDEF Workshop has been held at the University of Urbino since 2000. The 2014 Workshop is particularly dedicated to Carl Chiarella for his 70th birthday. As the second home (along with the University of Bielefeld, another second home), Carl visited Urbino in 1998 for the first time and the visit has become an almost annual event since then. In order to commemorate the occasion, a number of Carl's colleagues from around the world gladly agreed to contribute chapters to a special book dedicated to this event. The book is the outcome of this process. It contains the latest developments in nonlinear economic dynamics, financial market modeling, and quantitative finance, the three most active research areas Carl has been involved in.

This book is a collection of essays written by colleagues of Carl Chiarella in honour of his 70th birthday. Most of the authors have been collaborating with Carl in the past. We would first of all like to thank Laura Gardini and Gian Italo Bischi for stimulating discussion on the initiation of this special book and suggestion to dedicate it to Carl's 70th birthday in the 2014 MDEF Workshop. We would also like to express our gratitude to all contributors and in particular those who have collaborated with Carl, as well as to the referees involved in the review process. Finally, we would like to acknowledge the assistance of Kai Li who has worked on the book under much pressure.

Born in March 1944 in Sydney, Carl realized in his final high school years that he wanted to do something in life that would involve the use of mathematics, although that "something" would involve economics and finance was totally absent from his mind then. After completing his B.Sc. (Hons.) from the University of Sydney in 1965, Carl completed an M.Sc. at the University of Sydney in 1967, a Ph.D. at the University of New South Wales in 1969 in applied mathematics, and wrote a thesis on nuclear reactor physics. After spending two years at the University of Nancy as a postdoc Carl returned to Australia in 1971, and joined the School of Mathematical Science at University of Technology, Sydney (UTS). He has built his entire subsequent career at UTS since then (apart from a three year spell at the University of New South Wales from 1986 to 1989).

From his teenage years, Carl had an interest in the origins of the economic cycle. Despite heavy teaching load, Carl managed to pursue his long held interest in economics and completed a Master of Commerce in 1977 and a second Ph.D. in Economics at the University of New South Wales in 1988. His Ph.D. thesis was on

the nonlinear viewpoint in economics. The thesis led to his first book, *The Elements of a Nonlinear Theory of Economic Dynamics*, published in the Springer-Verlag Lecture Note Series in 1990. In 1989, Carl was appointed as a Professor in the School of Finance and Economics at UTS, a position that he still occupies. Apart from his early work on nuclear reactor theory, Carl has made numerous scientific achievements and important contributions to the economics and finance area, in particular to nonlinear dynamic economic, financial market modeling, and option pricing.

As a mathematical economist, Carl has a strong research interest in modeling key economic adjustment processes as nonlinear dynamical systems. Carl's earlier work on chaotic economic dynamics in 1980s, in particular *The Cobweb Model: Its Instability and the Onset of Chaos* published in Economic Modeling in 1988 and *The Dynamics of Speculative Behaviour*, published in Annals of Operations Research in 1992, have been pioneering contributions in this area, which had profound influence on many researchers in this field. Carl has made a significant contribution to at least two areas of dynamic economic modeling. The first is on out-of-equilibrium models of macroeconomic dynamics. It develops a systematic approach to the disequilibrium tradition of macroeconomic dynamic analysis, leading to nine jointly authored books (with Peter Flaschel and others) on integrated Keynesian dynamic macroeconomic models, including three with Cambridge University Press and three with Springer-Verlag. The other is on financial market models with heterogeneous boundedly rational economic agents, showing that price movements of financial assets are the result of nonlinear dynamic feedback processes driven by the interaction of investors with heterogeneous beliefs and bounded rationality.

Through his many conferences and visits, the University of Urbino and the University of Bielefeld have become second home for Carl. Carl's visits to Urbino started in 1998 and have become regular since then. The attraction of Urbino for Carl is not only the glorious history, beautiful palaces, and churches, but also a group of brilliant researchers around Laura and Gian Italo in the theory of nonlinear dynamical systems. Through the Vienna Workshops on Economic Dynamics initiated by Gustav Feichtinger, Carl established his intensive research collaboration with the research groups around Peter Flaschel, Willi Semmler, and Volker Böhm in Bielefeld. Carl's collaborations with these groups belong to the highlights of his career.

As one of the main organizers of the annual Quantitative Methods in Finance conference at UTS since 1997, Carl has made a significant contribution to American option pricing, where he has mainly contributed to the development and numerical implementation of various solution methods. He has also been active in pricing interest rate derivative securities along two directions. The first is to implement on market data the various interest rate term structure and interest rate derivative pricing models that have been developed over the last two decades using nonlinear filtering and Bayesian updating methods. The second consists in finding improved computational procedures within the stochastic calculus framework of the term structure and option prices by allowing the volatility function of

the Heath–Jarrow–Morton model to depend on the forward rate, and allowing for jump-processes in the underlying forward rate dynamics of this framework.

Carl has published more than 15 books and 200 papers, supervised more than 10 Ph.D. students, been involved in more than 30 research projects including the Australian Research Council (ARC) Discovery Grants. He was the Co-Editor of the *Journal of Economic Dynamics and Control* from 2004 to 2012 and has been Associate Editor of many leading finance and economics journals, including *Journal of Economic Behavior and Organization, Macroeconomic Dynamics, Computational Economics, Studies in Nonlinear Dynamics and Econometrics, European Journal of Finance, Quantitative Finance,* and *Asia-Pacific Financial Markets*. Of course, this is not a full list of Carl's numerous scientific achievements. The papers in this book deal with some of the many research topics Carl has addressed in many of his papers and books. They reflect the breadth of topics Carl has worked on during his career. We are grateful for the inspiration his work has given to all of us over so many years. Indeed his work inspires a new generation to further develop this exciting and challenging research agenda.

Bologna, April 2014 Roberto Dieci
Sydney Xue-Zhong He
Amsterdam Cars Hommes

Contents

Contributors

Anna Agliari DISES, Catholic University Piacenza, Piacenza, Italy

Toichiro Asada Faculty of Economics, Chuo University, Hachioji, Tokyo, Japan

Ramaprasad Bhar School of Risk and Actuarial Studies, The University of New South Wales, Sydney, NSW, Australia

Gian Italo Bischi Università di Urbino Carlo Bo, Urbino, Italy

John Breslin Lacima, Sydney, NSW, Australia

Matthieu Charpe International Labor Organization, Geneva, Switzerland

Gerald H. L. Cheang School of Information Technology and Mathematical Sciences, Centre for Industrial and Applied Mathematics, University of South Australia, Adelaide, SA, Australia

Wai-Mun Chia Division of Economics, Nanyang Technological University, Singapore, Singapore

Carl Chiarella UTS Business School, University of Technology, Sydney, NSW, Australia

Les Clewlow Lacima Group, Sydney, NSW, Australia

Mark Craddock School of Mathematical Sciences, University of Technology, Sydney, Broadway, NSW, Australia

Corrado Di Guilmi UTS Business School, University of Technology, Sydney, Sydney, NSW, Australia

Roberto Dieci Dipartimento di Matematica, Università di Bologna, Bologna, Italy

Peter Flaschel Bielefeld University, Bielefeld, Germany

Reiner Franke University of Kiel, Kiel, Germany

Mauro Gallegati DiSES, Università Politecnica delle Marche, Ancona, Italy

Laura Gardini DESP, Università di Urbino Carlo Bo, Urbino, Italy

Xue-Zhong He UTS Business School, University of Technology, Sydney, Sydney, NSW, Australia

Cars Hommes CeNDEF, Amsterdam School of Economics, University of Amsterdam, Amsterdam, The Netherlands

Giulia Iori Department of Economics, City University, London, UK

Boda Kang Department of Mathematics, University of York, Heslington, York, UK

Masaaki Kijima Tokyo Metropolitan University, Hachiohji, Tokyo, Japan

Polina Kovaleva Department of Economics, City University, London, UK

Daniel Ladley Department of Economics, University of Leicester, Leicester, UK

Simone Landini I.R.E.S. Piemonte, Turin, Italy

Mengling Li Division of Mathematical Sciences, Nanyang Technological University, Singapore, Singapore

Xihao Li DiSES, Università Politecnica delle Marche, Ancona, Italy

A. G. Malliaris Department of Economics and Finance, Quinlan School of Business, Loyola University, Chicago, IL, USA

Akio Matsumoto Department of Economics, Chuo University, Hachioji, Tokyo, Japan

Gunter H. Meyer School of Mathematics, Georgia Institute of Technology, Atlanta, GA, USA

Laura Morino Dipartimento di Matematica Pura ed Applicata, Università di Padova, Padua, Italy; Deloitte Consulting Srl., Milan, Italy

Yukio Muromachi Tokyo Metropolitan University, Hachiohji, Tokyo, Japan

Christina Sklibosios Nikitopoulos UTS Business School, University of Technology, Sydney, Sydney, NSW, Australia

Paolo Pellizzari Department of Economics, Ca'Foscari University, Venice, Italy

Christian R. Proaño The New School for Social Research, New York, NY, USA

Andreea Röthig Institute for Automatic Control and Mechatronics, Technische Universität Darmstadt, Darmstadt, Germany

Andreas Röthig Deutsche Bundesbank, Frankfurt am Main, Germany

Wolfgang J. Runggaldier Dipartimento di Matematica Pura ed Applicata, Università di Padova, Padua, Italy

Willi Semmler The New School for Social Research, New York, NY, USA

Joseph E. Stiglitz Columbia Business School, Columbia University, New York, NY, USA

Chris Strickland Lacima, Sydney, NSW, Australia

Iryna Sushko Institute of Mathematics NASU, Kiev, Ukraine

Ferenc Szidarovszky Department of Applied Mathematics, University of Pécs, Pécs, Hungary

Gim-Aik Teh BNP Paribas, Hong Kong SAR, China

Huanhuan Zheng Institute of Global Economics and Finance, The Chinese University of Hong Kong, Shatin N. T, Hong Kong; Department of Economics and Related Studies, The University of York, Heslington, York, UK

Introduction

Roberto Dieci, Xue-Zhong He and Cars Hommes

The book opens with two brief articles summarising Carl's view on a broad range of research-related issues, mostly concerning the role of mathematical modelling in Economics and Finance. Both articles were originally published in Italian in the Springer journal "Lettera Matematica Pristem". The first article is the result of an interview given by Carl to Gian Italo Bischi—one of Carl's collaborators from the 'Urbino group'—during the annual meeting of AMASES (Italian Association for Mathematics Applied to Economics and Social Science) in Lecce, in September 2007. Besides some biographical details, the interview focuses on Carl's prominent research themes, on some issues related to the use of mathematical modelling in social sciences, on alternative approaches to economic modelling such as Econophysics and Agent-Based models, on Carl's direct experience (as a world traveller) of research organisation and funding in Italy, Europe, USA, Australia and so-called Asian emerging countries. The second chapter authored by Carl himself, is a brilliant discussion of the debates that have gone on amongst economists in the past century about what is the correct approach to modelling economic behaviour, of the future of Mathematical Economics and of the possible impact of the recent economic crisis on economic theorising.

The second part of the book containing six chapters deals with Nonlinear Economic Dynamics, the area of Economic Theory that mostly attracted Carl's interest

R. Dieci
Dipartimento di Matematica, Università di Bologna, Piazza di Porta San Donato, 5,
40126 Bologna, Italy
e-mail: roberto.dieci@unibo.it

X.-Z. He (✉)
UTS Business School, University of Technology, Sydney,
Sydney, NSW, Australia
e-mail: tony.he1@uts.edu.au

C. Hommes
CeNDEF, Amsterdam School of Economics, University of Amsterdam, Amsterdam,
The Netherlands
e-mail: C.H.Hommes@uva.nl

R. Dieci et al. (eds.), *Nonlinear Economic Dynamics and Financial Modelling*,
DOI: 10.1007/978-3-319-07470-2_1,

since the beginning of his scientific career. The first two chapters stand in the tradition of the joint work of Carl with many co-authors in the field of disequilibrium macroeconomic dynamics, which typically deals with high-dimensional nonlinear dynamic models in continuous time. *Matthieu Charpe, Peter Flaschel, Christian R. Proaño and Willi Semmler* incorporate the basic elements of a firms' debt-finance model into a larger scale disequilibrium macroeconomic framework along the lines of Chiarella et al. (2005). In a fully interdependent macro-model they investigate both analytically and through numerical simulation the feedback impact of endogenously generated debt of firms on aggregate economic activity, on investors' confidence and on the stability of the economic system. The chapter by *Toichiro Asada* deals with the impact of macroeconomic stabilization policies under a Minsky-type "financial instability" hypothesis, again using the analytical framework of high-dimensional nonlinear Keynesian macrodynamic models. The chapter starts from a two-dimensional fixed price model without active stabilisation policy and considers, as extensions of this core version, a four-dimensional model of monetary stabilisation policy with flexible prices and a six-dimensional case with a monetary and fiscal policy mix. Besides providing a number of theoretical results concerning the stability of the steady state of the economy (depending on the fiscal and monetary parameters, and central bank's credibility), the chapter offers an economic interpretation of the main feedback mechanisms operating in the dynamic models. The next two chapters deal more closely with the bifurcations and the cyclical and complex dynamics that may emerge from traditional economic models when standard rationality and full-information assumptions are abandoned in favour of agents' bounded rationality and the use of simple rules of thumb in making decisions, which often results in nonlinearities. This is a research field in which Carl has made important contributions. The chapter by *Anna Agliari, Laura Gardini and Iryna Sushko* is inspired by Carl's early work (Chiarella 1990) on the issue of so-called dynamic instability of saddle-point type under perfect foresight, discussed within a continuous-time model of monetary dynamics (Sargent and Wallace 1973). A key feature of Carl's version of the monetary model was the assumption of a nonlinear S-shaped demand function, justified by realistic portfolio rules adopted by economic agents. Building on earlier work in collaboration with Carl (Agliari et al. 2004), the authors consider a discrete-time version of the perfect-foresight model with a similar (log) money demand function, linear within a 'normal' range of the expected inflation rate, and constant outside this range. As a result, the price dynamics of the physical good is described by a piecewise-linear one-dimensional map having two "kink points". Bounded cyclical orbits of any period and even chaotic dynamics may appear if the demand function is sufficiently sloped and price reacts slowly to excess money demand. The study is based on advanced and up-to-date analytical and numerical methods for piecewise linear models. *Akio Matsumoto and Ferenc Szidarovszky* build a dynamic monopoly model in which a bounded rational monopolist has partial and delayed information about the inverse demand function. In order to deal with this kind of uncertainty and to react smoothly to sudden market changes, the monopolist adopts a 'gradient' output decision rule based on average past data. After the benchmark case of fixed delay, the authors investigate the case of continuously distributed delays, under different weighting

functions of past data. Unlike the case of static rational monopoly, cyclic dynamics can emerge from a quite simple economic structure in this case. In particular, the parameter representing the length of the delay has a threshold value above which stability is lost via a Hopf bifurcation. The bifurcation boundary in the parameter space is investigated analytically and numerically, under different time averaging patterns. The last two chapters in this part are concerned with the joint impact of nonlinearity and stochastic factors in macroeconomic dynamics, which also represents a major topic among Carl's research interests in recent years. The chapter by *Simone Landini, Mauro Gallegati, Joseph E. Stiglitz, Xihao Li and Corrado Di Guilmi* develops an Agent-Based Model (ABM) of an economy with heterogeneous and interacting firms subject to financial constraints. It focuses on the macro effects of firms' learning and decision process, according to the notion of "social atom" (Buchanan 2007). In a nonlinear stochastic environment, the aggregate observables generated by the ABM are analysed by means of master equations and combinatorial master equations. The chapter is concerned with the dominance and survival of firms' behavioural rules, and the role played by "financial fragility" in a complex environment. It is found that financially fragile firms—the most active ones in learning—contribute more to growth and determine periods of expansion sustained by credit supply but, at the same time, their behaviour may compromise system stability. Besides providing insights into an alternative micro-foundation of macromodels, the chapter offers a new interpretation of system phase transitions. *Reiner Franke* starts from recent empirical evidence against the hypothesis of normal distribution of aggregate output, and reconsiders this issue for quarterly US output data using a number of statistical tests, among which is the "shape parameter" of the exponential power distribution, the two polar values of which constitute the normal and the Laplace distribution with its fatter tails. It turns out that evidence against normality of output growth rates is weaker than one might expect, once a structural break between the periods of Great Inflation (GI) and Great Moderation (GM) is properly taken into account. However, if the Laplace can be rejected in favour of normality in one subsample (GI), in the other subsample (GM) normality is rejected and the Laplace cannot be ruled out. The chapter provides new empirical results and methodological insights on the important issues of nonlinearity and non-normality of economic time series.

The third part of the book focusses on Financial Market Modelling, one of the main areas of Carl's work. The first two chapters present agent-based models of financial markets with limit order books, extending some of Carl's earlier work on this topic in Chiarella et al. (2009). *Giulia Iori and Polina Kovaleva* consider an ABM of an order-driven market in which agents hold heterogeneous beliefs and study the interrelations between pre-trade quote transparency and stylised properties of order-driven markets. Their ABM is able to replicate stylised facts such as negative skewness of stock returns and clustered volatility when book depth is visible to traders. Full quote transparency contributes to convergence in traders' actions, while partial transparency restrictions may lead to long-range dependencies. *Daniel Ladley and Paolo Pellizzari* study optimal trading strategies in order book-based continuous double auction markets. Their framework is still analytically tractable and optimal

trading strategies can be identified using numerical techniques. They find that the optimal strategies are well approximated by linear strategies using only the best quotes. This study illustrates that, in complex markets, optimal behaviour may be well approximated by simple (linear) heuristics, a major theme in Carl's work. The following chapter by *Wai-Mun Chia, Mengling Li and Huanhuan Zheng*, studies regime switching models in foreign exchange markets and fits in a rapidly growing literature on estimating heterogeneous agent models (HAMs), an area where Carl has made major contributions, e.g. recently in Chiarella et al. (2012). Three different empirical models are compared, endogenous switching with fractions determined by relative performance, endogenous switching based on macroeconomic fundamentals and models with heterogeneous beliefs based on a Markov-switching process. In-sample and out-of-sample forecasting are compared across these different HAMs using monthly AUD/USD exchange rate data. The last two chapters in this part of the book are concerned with time series modelling and empirical analysis of financial markets. *Andreas Röthig and Andreea Röthig* study the time-varying cross-market trading activities of speculators in US currency futures markets, extending earlier work in Röthig and Chiarella (2011). They investigate linkages between speculative activities in different currency futures markets. The results show positive responses of total/long/short speculative activities in the GBP, CAD and JPY futures markets to an increase in total/long/short speculation in the CHF futures markets, indicating the presence of cross-market herding activities. These cross-market linkages between speculative activities are relatively stable over time from 1994–2013 and therefore do not suggest that changes in regulation or new market participants or trading strategies had a significant and lasting impact on cross-market speculative activities. The final chapter in this part of the book, by *Ramaprasad Bhar and A.G. Malliaris*, deals with the Stochastic Discount Factor (SDF) methodology as a general empirical framework for asset pricing. In particular, the authors suggest a multifactor model for the SDF taking both macroeconomic fundamental factors, such as the yield curve, the VIX index and a measure for trading liquidity, as well as behavioural factors into account to identify significant determinants of the daily equity premium. The chapter proposes to include momentum return as a behavioural factor in the SDF and offers a way to address this issue empirically. The chapter also shows how copula methods can be used in this context to overcome analytical complexities for software implementation in defining the dependence between asset returns and the SDF. These five chapters illustrate the broad contributions of Carl to Financial Market Modelling.

The fourth part of the book consisting of seven chapters focusses on Quantitative Finance, another main area of Carl's work. The first two chapters develop a new framework for risk management of interest-rate products, extending some of Carl's earlier work on this topic in Chiarella and Kwon (2003), Chiarella et al. (2007) and Chiarella et al. (2010). To develop a new methodology for risk management of interest-rate sensitive products, *Masaaki Kijima and Yukio Muromachi* present a risk evaluation model for interest-rate sensitive products within the no-arbitrage framework. They first consider a yield-curve model under the observed probability measure, based on the results of the principal component analysis (PCA), to generate future scenarios of interest rates, and then iden-

tify market prices of risk for the pricing of interest-rate derivatives under the risk-neutral measure at any future time. Thus risk measures such as Value-at-Risk (VaR) of portfolios with interest-rate sensitive products can be evaluated through simple Monte Carlo simulation. They also show, however, that some market models often used in practice are not consistent with the no-arbitrage paradigm. Motivated by a significant increase in the spread between LIBORs of different tenors as well as the spread between LIBOR and the discount curve during the financial crisis, *Laura Morino and Wolfgang J. Runggaldier* extend Carl's work (Chiarella et al. 2007, 2010) beyond a pure credit risk setting to a more general post-crisis multicurve set-up. While Carl's work follows an HJM-based approach, here the authors use a short rate modelling with a short rate spread and consider a two-curve model with one curve for discounting (OIS swap curve) and one for generating future cash flows (LIBOR for a give tenor). The clean-valuation approach of pricing FRAs and CAPs without counterparty risk exhibits an "adjustment factor" when passing from the one-curve to the two-curve setting. The bottom-up short rate modelling where the short rate and a short rate spread are driven by affine factors allows for correlation between short rate and short rate spread as well as to exploit the convenient affine structure methodology. The next two chapters contribute to price American call option and futures price volatility with the framework of stochastic volatility, two areas Carl has made a significant contribution, see for example, Chiarella and Kwon (2001), Chiarella et al. (2009, 2010, 2013) and Adolfsson et al. (2013). To price an American call option when the underlying dynamics follow the Heston's stochastic volatility and the Cox-Ingersoll-Ross (CIR) stochastic interest rate, *Boda Kang and Gunter H. Meyer* formulate the call as a free boundary PDE problem on a finite computational domain with appropriate boundary conditions. Comparing with finite difference approximation, they find that the time discrete method of lines is accurate and efficient in producing option prices, early exercise boundaries and option hedge ratios such as delta and gamma. Using a continuous time forward price model with stochastic volatility, *Les Clewlow, Boda Kang and Christina Sklibosios Nikitopoulos* introduce three distinct volatility structures to capture the impact of long-term, medium-term and short-term futures price volatility in commodity futures markets. They then use an extensive 21-year database of commodity futures prices to estimate the model for six key commodities: gold, crude oil, natural gas, soybean, sugar and corn. They identify the shape and the persistence of each volatility factor, their contribution to the total variance, the extent to which commodity futures volatility can be spanned and the nature of the return-volatility relation. In the next chapter, based on the structural relationships in the electricity market, *John Breslin, Les Clewlow and Chris Strickland* develop a general framework for the modelling of Australian electricity market risk. The framework is consistent with temperature and load mean forecasts, market forward price quotes, the dependence of load on temperature and the dependence of price on load. The model uses basic building blocks of an HJM form for which Carl has contributed important results. The model can be used not only for accurate evaluation of the market risk of an electricity generation and retail company, but also for the valuation of electricity market derivatives and assets. They demonstrate the application of the framework to the Australian National Electricity

Market. In the following chapter, *Mark Craddock* presents a tractable solution to the Yakubovich parabolic PDE $u_t = x^2 u_{xx} + x u_x - x^2 u$, which arises in a number of financial mathematical problems such as Asian option pricing, the Hartman-Watson law and pricing zero coupon bonds in the Dothan model. After deriving the heat kernel for the PDE, Mark uses the Fourier sine transform to reduce the kernel to a simple form, which may be explicitly evaluated as a series of error functions. Some financial applications are then discussed. In the final chapter, by applying the approach of changing numeraire, *Gerald Cheang and Gim-Aik Teh* extend the European call option pricing formula in the literature to the case when both stock prices and interest rates are driven by jump-diffusion processes. The pricing model is an extended Merton jump-diffusion stock price model with a stochastic interest rate term structure that is an HJM-type model with jumps. The approach does not require Fourier transforms as used in the existing literature. It allows us to price the option when the bond price dynamics is also discontinuous. When the jump-sizes are fixed instead, then they get the special case of the HJM model with fixed jumps as in Chiarella and Nikitopoulos Sklibosios (2003).

References

Adolfsson, T., Chiarella, C., Ziogas, A., & Ziveyi, J. (2013). Representation and numerical approximation of American option prices under Heston stochastic volatility dynamics, Quantitative Finance Research Centre, University of Technology Sydney. Working paper No. 327.

Agliari, A., Chiarella, C., & Gardini, L. (2004). A stability analysis of the perfect foresight map in nonlinear models of monetary dynamics. *Chaos, Solitons and Fractals*, *21*, 371–386.

Buchanan, M. (2007). *The social atom: why the rich get richer, cheaters get caught, and your neighbor usually looks like you*. New York: Bloomsbury.

Chiarella, C. (1990). *The elements of a nonlinear theory of economic dynamics*. New York: Springer.

Chiarella, C., Flaschel, P., & Franke, R. (2005). *Foundations for a disequilibrium theory of the business cycle*. Cambridge, UK: Cambridge University Press.

Chiarella, C., He, X.-Z., Huang, W., & Zheng, H. (2012). Estimating behavioural heterogeneity under regime switching. *Journal of Economic Behavior and Organization*, *83*(3), 446–460.

Chiarella, C., Iori, G., & Perello, J. (2009). The impact of heterogeneous trading rules on the limit order book and order flows. *Journal of Economic Dynamics and Control*, *33*(3), 525–537.

Chiarella, C., Kang, B., Meyer, G. H., & Ziogas, A. (2009). The evaluation of American option prices under stochastic volatility and jump-diffusion dynamics using the method of lines. *International Journal of Theoretical and Applied Finance*, *12*(3), 393–425.

Chiarella, C., Kang, B., Nikitopoulos, C. S., & Tô, T. (2013). Humps in the volatility structure of the crude oil futures market: New evidence. *Energy Economics*, *40*, 989–1000.

Chiarella, C., & Kwon, O. (2001). Forward rate dependent Markovian transformations of the Heath-Jarrow-Morton term structure model. *Finance and Stochastics*, *5*(2), 237–257.

Chiarella, C., & Kwon, O. (2003). Finite dimensional affine realisations of HJM models in terms of forward rates and yields. *Review of Derivatives Research*, *6*, 129–155.

Chiarella, C., Maina, S. C., & Nikitopoulos, C. S. (2010). Markovian defaultable HJM term structure models with unspanned stochastic volatility. Quantitative Finance Research Centre Research Paper no. 283 (2010), University of Technology, Sydney.

Chiarella, C., & Nikitopoulos, C. S. (2003). A class of jump-diffusion bond pricing models within the HJM framework. *Asia-Pacific Financial Markets*, *10*, 87–127.

Chiarella, C., Nikitopoulos, C. S., & Schloegl, E. (2007). A Markovian defaultable term structure model with state dependent volatilities. *International Journal of Theoretical and Applied Finance*, *10*, 155–202.

Chiarella, C., Ziogas, A., & Ziveyi, J. (2010). Representation of American option prices under Heston stochastic volatility dynamics using integral transforms. In *Contemporary quantitative finance: essays in honour of Eckhard Platen* (pp. 281–315). New York: Springer.

Röthig, A., & Chiarella, C. (2011). Small traders in currency futures markets. *Journal of Futures Markets*, *31*, 898–913.

Sargent, T. J., & Wallace, W. (1973). The stability of models of money and growth with perfect foresight. *Econometrica*, *41*, 1043–8.

Part I
Carl Chiarella: An Interview and Some Perspectives

An Interview to Carl Chiarella, an Italo-Australian Globe Trotter Who Studies Dynamic Models for Economics and Finance

Gian Italo Bischi

Gian Italo. *Carl, what caused the shift of your researches from mathematical modelling in nuclear physics to economics and finance?*

Carl. From about my teenage years I had an interest in trying to understand the origins of the economic cycle, I often wondered about their causes. This interest was probably driven by the fact that both my grandfathers had emigrated to Australia from Italy in the mid-1920s, in good economic times that soon turned into the great depression. They experienced quite a deal of hardship during this period, experiences that were shared by both my parents who arrived in Australia as teenagers in the 1930s. So it was probably quite normal that these personal experiences, as well as the impact of the depression on the broader society, formed part of family discussions as I was growing up. Of course such stories could be repeated by many young people growing up during the 1950s. I was also aware of the earlier depression in Australia in the 1890s and of earlier recessions during the nineteenth century in Britain. I was fascinated by the fact that the economic cycle was a constantly recurring event, and often pondered as to its causes, though at that time without doing any formal modelling. As my high-school years unfolded I found that I had an aptitude for mathematics and physics and, after considering engineering studies, decided to major in applied mathematics at university. It would be nice for the purposes of this story to recount that I studied applied mathematics with a view to eventually working in finance and economics, but that was not at all the case. Young people of my generation gifted in mathematics were naturally led to careers in the sciences. For this reason I decided to write my mathematics Ph.D. in the area of nuclear reactor physics, the

Originally published (in Italian) in the journal "Lettera Matematica PRISTEM", n. 74–75, Springer (2010), pp. 108–111.

G. I. Bischi (✉)
Università di Urbino Carlo Bo, via Saffi n.42, 61029 Urbino, Italy
e-mail: gian.bischi@uniurb.it

R. Dieci et al. (eds.), *Nonlinear Economic Dynamics and Financial Modelling*,
DOI: 10.1007/978-3-319-07470-2_2,

problems and mathematics used interested me and the skills I developed seemed to promise to some sort of "useful" career. After returning to Australia from France in 1971 after a two-year post-doctoral scholarship at l'Université de Nancy, I took a job as a Lecturer in Mathematics at the University of Technology, Sydney. After two years, I felt sufficiently on top of my teaching duties to finally undertake studies in economics, in which I had always retained an interest. After doing the basic first-year economics course, and doing quite well, I was admitted into a course work master's degree in Economics. It was during the undertaking of this course that I came into contact with a couple of professors who inspired my interest in several of areas of economics and finance and who "took me under their wings" so to speak. The path to a second PhD in Economics followed quite naturally after that. I have written in more detail elsewhere how my particular interests in finance and economics developed.[1]

Gian Italo. *There are no doubts about the fact that mathematical models are useful in physics and engineering; however, many economists are quite sceptic about their effectiveness in economics and social sciences, as they consider them more similar to academic exercises than useful tools for solving real problems. What is your opinion?*

Carl. The question of the utility of mathematical models in the social sciences has probably been debated for some time, but I would say that as far as their use in economics and finance is concerned the debate is over, and the use of mathematical models is here to stay. This is probably due to a number of factors. First, a maturing of expectations of what one can achieve with mathematical models in economics and finance. The initial early hope was that economics and finance could become predictive sciences in the same way as celestial mechanics can predict perfectly the motion of the planets (actually not perfectly as we know that relativity corrections are necessary to predict perfectly for some planets). We now understand that the science

[1] C. Chiarella "My chaotic career—from billiard balls to economic dynamics and financial markets", Chaos, Solitons and Fractals 29 (2006) 517–519—special issue on 'Dynamic Modelling in Economics and Finance' in honour of Professor Carl Chiarella, edited by Bischi G.I. and I. Sushko..

of economics and finance is more like medical science. My doctor can inform me about all the risk factors I should be wary of if I want to avoid, say heart disease. But she cannot predict when, if at all, heart problems will start to occur if I choose to ignore her good advice. Indeed, she cannot even guarantee that I will not have heart problems even if I do follow her advice. There is nevertheless no call that the study of medical science should be abandoned because it cannot give us perfect predictions of the outcomes of following certain medical advice. So in economics and finance we have come to appreciate that perfect prediction is impossible, in part because the economy is too complex with many feedback loops, and is buffeted by many stochastic factors that it will never be possible to model. Economic science can tell us the general tendency of the economy if for example interest rates are increased or decreased, what are the essential feedback chains in the economy, how they operate and which parameters we need to tie down empirically to know the direction of certain effects, in financial markets we appreciate (at least by and large) that we cannot predict the market movements, but we can quantify and minimise the risk of our exposure to such movements. Indeed, if ever we could predict the movements of financial markets their very nature would any way change as they are institutions that have been established in order to allow society to deal with the inherent uncertainty and risks of the movements of the prices of risky assets.

Gian Italo. *Can you give some examples of dynamic models you studied for which you are particularly proud because they gave important suggestions to policy makers, economists or financial operators?*

Carl. I would cite two areas of dynamic economic modelling about which I am particularly pleased if not to say proud. The first is my work on macrodynamic modelling which has its roots in my economics Ph.D. thesis (though the ideas were germinating in my head from my initial studies of economics), but took on a new dimension when I started collaborating with Peter Flaschel in the early 1990s. Our recent book "A Disequilibrium Theory of the Business Cycle", joint with Reiner Franke, brings a lot of this work together and to the point where it could be taken by policy economists and developed into useful policy models. This work will be on-going for some time yet as we refine the basic model, do better estimations and calibrations, carry out policy experiments and so forth, but this book marks an important milestone in this long research agenda. The other area is my work on financial market models of heterogeneous boundedly rational economic agents. This developed out of my dissatisfaction with the standard paradigm of homogeneous rational representative agents, which I believe cannot even serve as a benchmark model for what goes on in real markets. From my original paper "The Dynamics of Speculative Behaviour" published in 1992 (but actually written in 1989) work on this topic has been intense with my UTS colleague (and former Ph.D. student) Tony Xue-Zhong He, Roberto Dieci (Bologna) and Laura Gardini (Urbino) being my principal co-authors. We have recently been joined by Min Zheng whose skills with the theory of random dynamical systems is helping us to elucidate the interaction of the nonlinear and stochastic elements, both of which I think are important in understanding how the interactions among heterogeneous agents bring about the type of price and return behaviour we observe in financial markets, such as fat tails, volatility clustering and so forth. The

work on deterministic effects has reached a mature stage with the publication of "Heterogeneous Expectations and Speculative Behavior in a Dynamic Multi-Asset Framework", (joint with Tony He and Roberto Dieci), and "Asset Price and Wealth Dynamics in a Financial Market with Heterogeneous Agents" (joint with Roberto Dieci and Laura Gardini). The work on the nonlinear and stochastic elements using the theory of random dynamical systems has just started with the working paper "The Stochastic Price Dynamics of Speculative Behaviour" and this will be a major focus of research effort in the coming few years.

Gian Italo. *What do you think about the approach known as "Econophysics"? Economists seem to be a bit reluctant to accept this kind of approach.*

Carl. First of all let me say that I dislike the term econophysics as it does not describe the new ideas and concepts that physicists are bringing to economic analysis. The term already creates a separation that I feel makes communication difficult. Recently, in my capacity as one of the editors of the Journal of Economic Dynamics and Control I was asked to oversee a special issue on econophysics (that I think has just recently appeared) with Doyne Farmer and Thomas Lu as guest editors. I suggested that the title of the special issue be "Applications of Statistical Physics to Economics and Finance" as I felt that the use of techniques and concepts from that discipline were the main ideas that were being used in the economics and finance context. I think there are a number of reasons why the econophysics community has had a minimum impact in economics. First of all there is the communication barrier, the physicists are really writing for other physicists who have developed an interest in economics and finance so they do not try to translate their language into that used by economists. As a result it is very difficult for economists to gain any message from this literature. Second, and this is also related to the communication barrier problem, economists have gained the impression (rightly or wrongly) that physicists have come down from the mount to teach them how to use advanced mathematical tools, whereas economists have been borrowing mathematical tools and concepts as needed from a range of disciplines over the last six decades. To see this one only needs to consider the use in economic analysis of the theory of nonlinear dynamical systems, stochastic differential equations, stochastic optimal control theory, Monte Carlo simulation methods, the numerical solution of partial differential equations; many of these tools had their origins in physics, and economists feel they have not been laggards in adapting them to the needs of economic analysis as required. Third, it seems to me that economists have a higher standard of empirical analysis, so economists remain unconvinced by the type of empirical analysis one encounters in the econophysics literature. It is my fourth point that I believe is at the root of the fact that economists ignore the econophysics literature, and this is to do with the concepts of equilibrium and disequilibrium. Mainstream economists have become firmly wedded to the notion that the economic system is in an equilibrium state, by which they usually mean a stable fixed point of some dynamical system. Physicists see disequilibrium as the more normal state of affairs since this is what they are used to in physics, and so they frequently seek to model the economic process in disequilbrium terms, for instance using concepts from statistical physics. It may further be added that econophysicists seem to not accept the notions of rational expectations and the

representative agent which are regarded almost as an article of religious faith by some mainstream economists. Those who know my work will know that I also share the view of the econophysicists concerning the representative agent and the rational expectations paradigm, and I think the use of concepts from statistical physics to model the interaction of heterogeneous agents similar to an ensemble of particles in physics should (and is indeed already) provide a fruitful avenue for research.

Gian Italo. *Carl, you spend a lot of time to travel all around the world to meet research collaborators, and when you are in Sydney you often invite researchers to visit your department. Do you think that these travels are still useful for making research even if Iinternet connections and email are so diffused, easy and cheap*?

Carl. It is certainly the case that I spend many short but intense periods with my various research collaborators either at my home institution UTS or visiting the universities of my colleagues. We also collaborate very much and very effectively by email, skype and telephone. However, I find that the really significant progress on a joint research project comes about when the co-authors are sitting around the same desk staring at and thinking about difficult aspects of some question on a piece of paper or a white board. Also, often the conversations over a coffee or a meal together on such occasions lead to some breakthroughs. Finally, I often think that just the fact that we have gone to so much trouble to be working together for such a short period forces the pace so to speak and leads to a lot of progress in a short period of time.

Gian Italo. *Can you make some comparisons among the different ways to organise research in the universities and research institutions in Europe, USA and Asian emerging countries?*

Carl. Interestingly I would say that the actual way of doing research on the part of the scholar is universal, and is the same in all the countries I have visited and where I have collaborators. What I have observed differs a lot is the infrastructure support within which the scholar finds himself/herself working, such as the teaching load, support to travel to conferences, general support (computers, libraries etc.), and availability of research grant funding, and more importantly, whether the bureaucratic structures in the universities are aiding or hindering the research enterprise. The views that I have formed are just personal impressions that I hasten to add are not backed up by hard statistics. I would say overall it is best in the USA on pretty well all of the fronts I have mentioned, it is also quite good in Japan though I personally find their bureaucratic structures somewhat Byzantine. In Europe, it may be hard to make a general statement, in some countries (e.g. the Scandinavian countries, the UK and Holland) support seems quite good, but in other countries it could be much better (e.g. France, Italy and Germany), certainly in these latter countries the bureaucratic structures are tending more towards the hinderance side of things. Speaking more broadly for a moment, in my view the difficulties of research support in Europe transcend the bigger question that I think is starting to be debated in several European countries as to whether the university system should remain very open or become more selective. In my own country I would say we are somewhere between the European and North Amrerican situation, support is reasonable but could be better. Probably the more market-oriented approach pursued since the late 1980s is pushing us more towards the North American situation. The countries where I have seen the

biggest changes with respect to research support are the emerging countries of Asia, such as Singapore, mainland China, Taiwan and Hong Kong (I mention these last three separately only because their developments have followed quite different paths over the last 60 years). I recall visiting one of the universities in Singapore in the early 1970s, and was very conscious of the fact that they lagged behind universities in the USA, Europe and Australia. I have spent three periods at a university in Singapore since 2000, and I have become very conscious of how far they have progressed. There has been a tremendous investment in teaching and research infrastructure, and there is a real desire on the part of key decision-makers for the Singapore universities to be ranked among the top in the world. Indeed, that is already happening if one takes seriously the rankings one reads about in major international newspapers and, whatever one may think about these rankings (and I have my doubts on several counts), they are nevertheless indicating certain trends that cannot be ignored. There is also a similar trend with universities in China, Hong Kong and Taiwan. I recently read in a major international publication that ranks universities globally a prediction that within the next decade the top universities in Asia will surpass the top European universities. I would have to say that based on my personal observations this is a forecast with which I would tend to agree. This fact will come as a surprise to many Europeans who may still think of these countries as emerging, whereas on many counts they have already emerged.

Gian Italo. *How did you start your contacts with Italian researchers? What is your opinion about the level of research here in Italy?*

Carl. I have already described in the Chaos, Solitons and Fractals (see footnote 2) article how I became aware of some very interesting and good quality research being done in Italy through the AMASES association, and I published a paper in the AMASES journal[2] in 1985. A contingent of Italians came to the first QMF conference organised annually at UTS, and it was through them that I established contact with Laura Gardini that led to my first visit to Urbino (of which there have been many since) in 1998. The initial contact has led to the blossoming of many research flowers, I will here just highlight the collaborations with Laura herself, Roberto Dieci and Anna Agliari. I also have contacts with the group around Mauro Gallegati in Ancona, the Cagliari-Genoa group around Michele Marchesi and Silvano Cincotti, the group around Matteo Marsili at ICTP in Trieste, and more recently the group in Siena around Alessandro Vercelli and Serena Sordi whom I will visit after the AMASES meeting in Lecce. I think the level of research in Italy is quite high in certain areas, especially in nonlinear dynamics and in particular its applications to economics and finance. I never cease to be amazed at the quality of research in Italy given the many impediments against which the Italian researchers have to struggle. I have already alluded to the general problems in responding to the previous question, but in addition they often have to cope with holding a position in towns far from where their families live (because it takes so long to get a permanent position), quite heavy teaching loads

[2] The AMASES journal is now known as “Decisions in Economics and Finance”, published by Springer.

and rather patchy infrastructure support. I think the fact that high quality research can be maintained may be due to the long tradition of good research in Italy and the fact that Italian researchers seem to network quite well. Nevertheless, I think it is not a good sign for the future of Italian research that I find many good young Italian researchers working not only in the US, but also in the other European countries that give much better support to research.

What's Beyond? Some Perspectives on the Future of Mathematical Economics

Carl Chiarella

There is no more difficult, and some might say futile, task than that of trying to predict the future as one is almost certainly bound to be wrong. This aphorism is doubly true in economics as economists themselves remain divided over what is the correct approach to modelling economic behaviour. So perhaps it is best that I start by discussing this point first of all.

The great debates that have gone on among economists for much of the past century have been reflected in many of the essays in this special issue, but let me state what I see as the main issue. This point is important since the mathematical economic tools that one needs to develop are different depending on which economic paradigm one adopts. Early in the development of economic theory economists felt that the discipline required some guiding principle, just as in the classical physics of the nineteenth century there was the principle of least action. The principle in economics that is supposed to be the force driving individuals is that of expected utility maximisation, with utility functions being rooted in the five axioms of choice. However, in order to calculate expected utility the economic agent needs to have some concept of the probability distribution of future outcomes in the economy. To handle this issue in a way that eliminates any arbitrary specification of expectations, some economists in the late 1960s and early 1970s proposed the notion of rational expectations. According to this postulate agents know the true distribution and indeed the laws of motion driving the state variables of interest. Thus was born the rational representative economic agent who is able to perform a very precise decision calculus because it is able to calculate (at least in theory) all required quantities.

Originally published (in Italian) in the journal "Lettera Matematica PRISTEM", n. 68, Springer (2008) pp. 11–17.

C. Chiarella (✉)
UTS Business School, University of Technology, Sydney,
PO Box 123, Broadway, NSW 2007, Australia
e-mail: carl.chiarella@uts.edu.au

R. Dieci et al. (eds.), *Nonlinear Economic Dynamics and Financial Modelling*,
DOI: 10.1007/978-3-319-07470-2_3,

The rational economic agent, efficient markets view (known as the neoclassical model) has come to dominate economic theorising, at least in North American universities and some universities in Europe and some parts of Asia that seek to mimic American style institutions. The fundamental reason for this dominance was the perceived failure of the previously dominant Keynesian viewpoint of the late 1940s, the 1950s and the 1960s, which was deemed to be ineffective in face of the period of economic stagnation that gripped the major industrial economies from around the mid-1970s. Keynesian ideas in turn had displaced the previously dominant classical viewpoint in the 1930s after it was perceived to be the cause of policies that led to the great depression, an economic event which had a profound impact on economic thought at the time and for many decades subsequently. It is still too early to tell whether the current economic crisis will have the same profound impact upon economic theorising. This at least seems likely given the fact that in order to manage the crisis the world's major economic institutions have been forced to adopt policies that run counter to the currently dominant paradigm, and indeed those policies have very much a Keynesian flavour about them. Throughout the period of the development of the neo-classical school of thought during the 1970s and 1980s there nevertheless continued a tradition that did not consider economic agents as totally rational, omniscient decision-makers nor markets as perfectly efficient. Rather it posited economic agents as boundedly rational and having limited computational power; it also allowed for the view that markets are not self-regulating. This view became properly articulated from the early 1990s with seminal works like those by Kirman (1992), (1993) or Day and Huang (1990), and has developed into a number of related directions; see for example the survey articles of Hommes (2006), Lebaron (2006), Chiarella et al. (2009b) and the books of Aoki and Yoshikawa (2007), Chiarella et al. (2009a) and Delli Gatti et al. (2008).

One reason why the two different viewpoints on economic life lead to different styles of mathematical economics is that each viewpoint has a different concept of the long-run outcome of economic activity. The neoclassical view sees the economy as arriving at some equilibrium at which all necessary economic relationships (such as budget constraints) are satisfied. This approach is very much rooted in methods of optimisation and general equilibrium theory and indeed the term dynamic stochastic general equilibrium theory has been coined to describe this general approach. The reader may consult Stokey and Lucas (1989) and Ljungqvist and Sargent (2000) for a good overview of what this approach involves. The mathematical workhorses in this approach are Markov chains, stochastic linear difference equations and dynamic programming. The main object of analysis becomes the so-called stochastic Euler equations which are essentially stochastic difference equations expressing the first order condition of optimality for the dynamic programming problem of interest. These equations link quantities of interest at two successive time periods via an expectation operator (and hence the need for some theory about expectations, which as stated earlier for the neoclassical approach is that of rational expectations). Implementation of this approach involves one in practical issues of dynamic programming algorithms, such as the curse of dimensionality, the best approach to discretisation of the state space and ways to approximate the value function.

Many details and references can be found in the book by Ljungqvist and Sargent. At a conceptual level one is dealing with general equilibrium theory, rational expectations, overlapping generations and game theory. Despite its use of the term "dynamic" this overall approach is essentially a static one, since it posits the economy as arriving at some steady-state situation which is perturbed by external random factors, to which the economic agents are reacting optimally, and hence, the strong focus on tools of optimisation theory and practice. One reason that the neoclassical view focuses on stable equilibrium situations is almost by assumption as the rational expectations assumption puts the economic system on a stable path by construction—any paths that lead to a consideration of instability are simply ruled out as impossible outcomes. This is the case for example with the employment of the so-called jump variable technique that in the face of a saddle point solution (that occurs naturally in many models of rational expectations) assumes that somehow the economy will manage to put itself onto the stable branch of the saddle; though the mechanism by which this can come about is never explained. Broadly speaking one may regard stochastic dynamic general equilibrium theory as the latest manifestation of Walrasian economics. The alternative view based on boundedly rational economic agents is in contrast a disequilibrium theory as it allows for a much wider range of possible equilibrium outcomes than simply a fixed point. The wider range of equilibrium outcomes includes limit cycles, strange attractors, multiple basins of attraction. If one also allows for the influence of external noise factors then one has noisy versions of all of these possibilities. At a mathematical level one is dealing with random nonlinear dynamical systems. While I have depicted the boundedly rational agent view as if it were one coherent overall theory, it in fact involves many different subthemes and I should stress that the broad classification is my own and several people

working in some of the sub-fields that I mention may not agree entirely with this classification. I have already pointed out a characteristic feature of this alternative approach is disequilibrium. This comes about because the underlying models allow for nonlinear elements, and so the loss of local stability does not necessarily entail some unacceptable divergent outcome. Economically, apart from allowing the wider range of equilibrium outcomes referred to earlier, this class of models makes room for financial markets that are not always efficient. The issue of market efficiency had almost become an article of faith in neoclassical economics, with the belief that markets are self-correcting and the role of government was at most one of a light touch regulator. Mathematically, one needs to deal with the full array of methods developed over recent decades for the analysis of nonlinear dynamical systems, a good account of which can be found in Puu (2000) for deterministic models. For an example of the developing use of random nonlinear models in economics the reader may consult Chiarella et al. (2009).

Let me spend a paragraph describing some of the major themes that have developed under the boundedly rational economic agent paradigm. First, there is the expression in the modern theory of nonlinear dynamics of the classical Keynesian system, updated in light of developments in economic theory since the mid-1970s (when most American schools started to abandon this approach); a good representation of this approach and the mathematical tools involved (mainly the theory of nonlinear dynamical systems) is contained in Chiarella et al. (2008). A very ambitious research agenda that has developed in recent years consists in treating economic agents as particles in statistical physics, and overcomes the aggregation problem[1] by using the master equation to statistically aggregate the agents and describe their aggregate behaviour via statistical distributions. The best expression of this approach is given by Aoki and Yoshikawa (2007), and how it can be applied to macroeconomic questions of relevance to current debates is demonstrated by Delli Gatti et al. (2008). For a nice overview of the mathematical tools involved, which are essentially ideas adapted from statistical physics, we refer the reader to Landini and Uberti (2008). Certainly, this approach is in stark contrast to the approach based on equilibrium theory as it is an attempt to apply to economics the ways developed in statistical mechanics to analyse systems in a permanent state of disequilibrium. In the view of this writer this attempt could mark an important paradigm shift for economic theory. It is too early to tell whether it will have the impact on economic theory that it should. Certainly, the current financial crisis has given great impetus to the search for alternative economic paradigms, as the current one has been found to be sorely inadequate to explain the crisis or to propose appropriate policy advice. But the pull of the general equilibrium viewpoint remains very strong in economics, and only

[1] Economists have long struggled with the aggregation problem without really finding a satisfactory solution. The basic problem is to determine the aggregate behaviour of the economy by aggregating the optimal behaviour of each economic agent. The difficulty of carrying out such a procedure is one reason for the dominance of the representative economic agent paradigm. Although we should point out that in recent years there have been attempts to incorporate heterogeneous agents into the neoclassical paradigm (den HaanW 2010).

the writer of a similar survey article two decades hence will really be able to make a pronouncement on the outcome of this struggle between paradigms. Certainly we live in interesting times.

References

Aoki M., & Yoshikawa H. (2007). *Reconstructing macroeconomics: A perspective from statistical physics and combinatorial stochastic processes*, Cambridge: Cambridge University Press.

Chiarella, C., He, X., Wang, D., & Zheng, M. (2008). The stochastic bifurcation behavior of speculative financial markets. *Physica A*, *387*, 3837–3846.

Chiarella, C., Dieci, R. & He, X. (2009a). Heterogeneity, market mechanisms, and asset price dynamics. In T. Hens & K. R. Schenk-Hoppe (Eds.), *Handbook of financial markets: Dynamics and evolution*, (pp. 277–344). New York: Elsevier.

Chiarella, C., Flaschel, P., Franke, R., & Semmler, W. (2009b). *Financial markets and the macroeconomy*. Routledge: A Keynesian Perspective.

Day, R., & Huang, W. (1990). Bulls, bears and market sheep. *Journal of Economic Behavior and Organization*, *14*, 299–329.

Delli, Gatti D., Gaffeo, E., Gallegati, M., Giuliani, G., & Palestrini, A. (2008). *Emergent macroeconomics: An agent based approach to business fuctuations*. Heidelberg: Springer.

den Haan, W.. (2010). Assessing the accuracy of aggregate laws of motion in models with heterogeneous agents. *Journal of Economic Dynamics ands Control*, *34*, 79–99.

Gandolfo, G. (2009). *Economic dynamics*. 4th edn. Heidelberg: Springer.

Hommes, C. (2006). *Heterogeneous agent models in economics and finance*. In L. Tesfatsion & K. Judd (Ed.), (pp. 1109–1186). North Holland: Agent Based Computational Economics.

Kirman, A. (1992). Whom or what does the representative agent represent? *Journal of Economic Perspectives*, *6*, 117–136.

Kirman, A. (1993). Ants, rationality and recruitment. *Quarterly Journal of Economics*, *108*, 137–156.

Landini, S., & Uberti, M. C. (2008). A statistical mechanics view of macrodynamics. *Economics, Computational Economics*, *32*, 121–146.

LeBaron, B. (2006). *Agent based computational finance* In L. Tesfatsion & K. Judd (Eds.), North Holland: Agent Based Computational Economics.

Ljungqvist, L. & Sargent, T. J. (2000), *Recursive macroeconomic theory*, MIT Press.

Puu, T. (2000). *Attractors, bifurcations and chaos: Nonlinear phenomena in economics*. Heidelberg: Springer.

Stokey, N. L. & Lucas, R. E. (1989). *Recursive methods in economic dynamics*, Harvard University Press.

Part II
Nonlinear Economic Dynamics

Expectations, Firms' Indebtedness and Business Fluctuations in a Structural Keynesian Monetary Growth Framework

Matthieu Charpe, Peter Flaschel, Christian R. Proaño and Willi Semmler

1 Introduction

In recent times, and especially since the recent global financial crisis, there has been a renewed interest in understanding the interplay between the financial markets and the macroeconomy, as well as the role played therein by the expectations of the different agents in the economy. As the role of both equity and debt financing has not only increased especially in the United-States since the late 1980s, and since both equity and debt issuance have become more volatile as well as correlated to each other, understanding the interaction between the firms' indebtedness levels and the dynamics of the economy remains a central question in economic research.[1]

In the literature numerous and quite heterogeneous approaches to this topic have been developed over time, and many such approaches are directly linked to the assessment of the efficiency of financial markets. For instance, while according to the Modigliani and Miller (1958) theorem the value of a firm would be unaffected by its financing structure if capital markets are frictionless, in theoretical studies along

[1] On more empirical grounds, recent work by Covas and Den Haan (2012) has stressed the fact that the relative importance of these two types of financing depends on the firm's size, with small firms and start-ups being more reliant on equity finance, and large and more established firms being more debt-finance oriented.

M. Charpe
International Labor Organization, Geneva, Switzerland
e-mail: charpe@ilo.org

P. Flaschel (✉)
Bielefeld University, Bielefeld, Germany
e-mail: pflaschel@wiwi.uni-bielefeld.de

C. R. Proaño · W. Semmler
The New School for Social Research, New York, NY, USA
e-mail: christian.proano@gmail.com

W. Semmler
e-mail: semmlerw@newschool.edu

R. Dieci et al. (eds.), *Nonlinear Economic Dynamics and Financial Modelling*,
DOI: 10.1007/978-3-319-07470-2_4,

the lines of Minsky (1975, 1982), high debt-asset ratios may reflect a dangerous liability structure of firms, increasing the risk of bankruptcy and developing fragile macroeconomic dynamics.

Already in the 1980s, early attempts were made to fully endogenise debt-financing in an interdependent economic system. For example, while in Foley (1986, 1987) borrowing, lending and capital outlays of firms are essentially determined by endogenously determined profit rate and liquidity, in Taylor (1985) and Taylor and O'Connell (1985), the savings generated by the rentier households, as already suggested by Kalecki, are channeled through the banking system to investing firms. However, despite the fact that this class of models develops rich macroeconomic dynamics—including financial instability as studied by Minsky in his financial crisis theory—the role of firms' indebtedness and debt payment commitments is not fully worked out given the partial nature of those frameworks.

In this contribution we incorporate the main elements of the small-scale firms' debt-finance model by Franke and Semmler (1989) into a medium-scale disequilibrium macroeconomic framework along the lines of Chiarella et al. (2005). In a fully interdependent model incorporating investing firms, savings of rentier households, commercial banks and the government, the endogenously generated debt of firms (created through borrowing) feeds back dynamically to the investment behaviour of firms, their borrowing of funds, the asset market, the interest rate and the expected rate of return (representing the confidence of investors with regard to future development). The impact of debt-financing of firms on aggregate economic activity will be studied within this context.

The remainder of this chapter is organised as follows: In Sect. 2, after describing briefly the main features of the Franke and Semmler (1989) framework, we discuss a medium-scale macroeconomic model which incorporates these features into a much richer Keynesian disequilibrium framework. We then investigate the dynamics of the resulting model by means of numerical simulations in Sect. 3. Finally, we draw some concluding remarks and mention the possible lines of future research in Sect. 4.

2 A Medium-Scale Keynes-Metzler Macroeconomic Framework of Firms' Indebtedness and Business Fluctuations

The main objective of this contribution is to integrate the approach of Franke and Semmler (1989) towards financial markets, debt-financing of firms and expectations about the future state of the economy into a medium-scale macroeconomic framework of the Keynes–Metzler variety along the lines of Chiarella et al. (2005).

In a nutshell, the theoretical framework originally proposed by Franke and Semmler (1989) studies the complex macroeconomic dynamics resulting from a debt-financing strategy by the entrepreneurial sector in a closed economy and the state of confidence in the economy, and therefore, on the general expectations of the economic agents with respect to the current and future performance of the economy. The modelling strategy is as follows: In the first step, Franke and Semmler

characterise the short-run equilibrium in goods and financial markets using a slight modification of a framework already employed several times in the literature (see e.g. Taylor and O'Connell (1985), Foley (1986, 1987)) and investigate the short-run stability properties of the resulting equilibrium. In this context, the ratio of indebtedness of firms and the economy's "state of confidence" are considered to be exogenously given. Further, in a second step the debt-to-capital ratio and the state of confidence are then endogenised through the formulation of behavioural equations concerning their evolution over time based on the assumption of continuous market clearing in the goods and financial markets.

Concerning the latter dynamic variable, Franke and Semmler (1989) assume that it evolves over time according to

$$\dot{\psi} = v(\rho - r, \lambda), \quad v_1 > 0, \ v_2 < 0, \tag{1}$$

where ρ represents the actual real rate of profit and λ the firms' debt-to-capital ratio, i.e. $\lambda = \Lambda/(pK)$, where Λ represents the firms' aggregate debt level, p the aggregate price level and K the capital stock level. Accordingly, the state of confidence in the economy depends negatively on the firms' debt-to-capital ratio, and positively on the differential between the actual profit rate and the real rate of interest, as originally proposed by Kalecki (1937) [84–95] in the following manner: "…the rate of investment decision is an increasing function of the gap between the prospective rate of profit and the rate of interest" and the difference between the "prospective rate of profit, and the rate of interest, is equal to the risk incurred".

The growth rate of the firms' indebtness is assumed to be a positive function of the difference between the expected profit rate and the current rate of interest, and a negative function of the debt-to-capital ratio,

$$\dot{\Lambda}/\Lambda = b(\rho^e - r, \lambda), \quad b_1 > 0, \ b_2 < 0, \tag{2}$$

where the expected rate of profit is defined through

$$\rho^e = \rho + \psi. \tag{3}$$

In contrast, the growth rate of the capital stock $\dot{K}/K = I/K$ (where I represents aggregate investment) is assumed to be driven solely by the relative expected profitability, i.e.

$$\dot{K}/K = g(\rho^e - r), \quad g_1 > 0, \ g_2 < 0. \tag{4}$$

Since $\hat{\lambda} = \hat{\Lambda} - \hat{p} - \hat{K}$, the resulting 2D nonlinear system of differential equations of the Franke and Semmler (1989) framework is given by:

$$\dot{\lambda} = (b(\rho + \psi - r, \lambda) - g(\rho + \psi - r)) \cdot \lambda \tag{5}$$

$$\dot{\psi} = c(\rho - r, \lambda) \tag{6}$$

under the assumption that $\hat{p} = 0$, and where we rewrite Eq. (1) again here for expositional clarity.

While different assumptions on the diverse reaction functions $b(\cdot)$, $g(\cdot)$ and $c(\cdot)$ make it possible to generate quite differentiated and complex dynamics, in Franke and Semmler (1989) essentially three types of scenarios are investigated. Two of them exhibit local (if not global) stability with respect to a unique long-run equilibrium of steady growth. The other one, the most interesting case, generates persistent fluctuations in the debt-to-capital ratio and the state of confidence in the economy tending to a closed orbit (so that a growth cycle results).

The Franke and Semmler (1989) framework is capable of delivering valuable insights on the interaction between the firms' indebtedness levels, the economy's state of confidence and macroeconomic activity. However, it contains still various assumptions, such as the continuous goods market equilibrium and the exogenously given labour share, which are not only unlikely on empirical grounds, but also at odds with a truly Keynesian framework from a theoretical point of view. Accordingly, in the following we incorporate their approach into a medium-scale macroeconomic framework along the lines of Chiarella et al. (2005).

The household sector is assumed to comprise two different types of agents: workers and asset holders. Workers are assumed to consume all their labour income, while asset holders are assumed to consume only a fraction of $1 - s_c$. Under the assumption that all interest payments to banks are transferred to the household sector (and more specifically to the asset holders), the aggregate household consumption is determined by

$$C = \omega L^d + (1 - s_c)(\rho K + r\Lambda/p), \tag{7}$$

where ω denotes the real wage w/p, L^d the labour demand (which equals the actual level of employment given the underutilisation macroeconomic regime assumed here), K the capital stock, r the rate of interest, Λ the total level of the households' interest bearing deposits and p the aggregate price level. Further, the rate of profit ρ is defined as

$$\rho = (Y^d - \delta K - \omega L^d - r\lambda K)/K, \tag{8}$$

where Y^d represents the aggregate demand level, as usual defined as

$$Y^d = C + I + G + \delta K, \tag{9}$$

where I represents aggregate net investment, G aggregate government expenditures and δ the depreciation rate of capital.

The households' (meaning, asset holders') real savings are allocated in new money holdings $\dot{M}$, new loans to the entrepreneurial sector $\dot{\Lambda}$ and new equity holding $\dot{E}$,

$$S_p = s_c(\rho K + r\Lambda/p) = (\dot{M} + \dot{\Lambda} + p_e\dot{E})/p. \tag{10}$$

Accordingly, the households' financial wealth (consisting of course only of asset holders' wealth) consists of money holdings, debt claims and equities (all expressed here in nominal terms), i.e.

$$W = M + \Lambda + p_e E. \tag{11}$$

Concerning the firms' behaviour, labour demand L^d is assumed to be fully determined by the aggregate output level Y (see Eq. (17)), where $x = const.$ is the average labour productivity, i.e.

$$L^d = Y^d / x. \tag{12}$$

Further, aggregate net investment I is assumed to depend not only on the difference between the expected profit rate and the rate of interest (as in Eq. (4)), but also on the deviation of the capacity utilisation u from its normal level $\bar{u}$, and the trend growth of investment $\gamma = const.$ (which for simplicity is assumed to equal the growth rate of the population), namely

$$I = i_1(\rho^e - (r - \pi^e))K + i_2(u - \overline{u})K + \gamma K, \tag{13}$$

where $u = Y/Y^p$, with $Y^p = y^p K$ and $y^p = Y^p/K = const.$ representing the potential output-capital ratio. Further, Eq. (14) represents the budget restriction of firms, which states that net investment is either financed by new borrowing or by issuing new equities, i.e.

$$I = (\dot{\Lambda} + p_e \dot{E})/p. \tag{14}$$

with $\hat{K} = \dot{K}/K = I/K$ being the growth rate of capital by definition.

Total aggregate investment includes net investment, the replacement of depreciated capital and the actual change in the firms' inventories N, i.e.

$$I^a = I + \delta K + \dot{N}, \tag{15}$$

where $\dot{N}$ results from the discrepancies between actual output Y and actual aggregate demand Y^d, i.e.

$$\dot{N} = Y - Y^d = S_f. \tag{16}$$

As extensively discussed in Chiarella and Flaschel (2000) and Chiarella et al. (2005), in an uncertain environment, firms have to decide on production before actual sales are known. Accordingly, actual aggregate output equals the expected output level Y^e plus the additional change in actual inventories, i.e.

$$Y = Y^e + \mathscr{I}. \tag{17}$$

with the change in the desired level of inventories N^d being assumed to be determined through

$$\mathscr{I} = \gamma N^d + \beta_\gamma (N^d - N), \tag{18}$$

where

$$N^d = \beta_{\gamma^d} Y^e, \tag{19}$$

as in Chiarella et al. (2005). Further, the evolution of Y^e is assumed to be given by:

$$\dot{Y}^e = \gamma Y^e + \beta_{y^e}(Y^d - Y^e). \tag{20}$$

Accordingly, the expected output level is determined in an adaptive manner by aggregate demand Y^d, as well as by the trend population growth γ.

The firms' pursued debt policy is described by Eq.(21), according to which the growth rate of debt financing of firms depends positively on the expected rate of profit (in nominal terms), negatively on interest payments (in capital stock terms) $r\lambda$.[2]

$$\dot{\Lambda} = d(\rho^e + \pi^e, r\lambda) \cdot \Lambda, \quad \text{with } d_1 > 0, d_2 < 0 \tag{21}$$

For the sake of simplicity, the government's behaviour is described in a very parsimonious manner. Real government expenditures are proportional to net investment and are totally money financed, so that government dissavings equal the current government expenditures as formulated in Eqs. (22)–(24), i.e.[3]

$$G = \phi I, \quad \phi = const. \tag{22}$$

$$\dot{M} = pG \tag{23}$$

$$S_g = -G, \tag{24}$$

Turning next to the asset markets we first of all consider the balance sheets of the economy and their aggregation towards the determination of the nominal wealth of asset owning households, as shown in Table 1.

The scheme illustrated in Table 1 is based on the assumption that there are no currency holdings of the public that holds (and use) liquidity in the form of interest-free demand deposits (with no fixed multiplier relationship connecting M and Λ^o, or Λ^o and Λ^r).[4] Interest bearing deposits of the public, by contrast, are not "liquid" and they exhibit a fixed price (as a deposit) and as the result of market interaction a variable rate of interest r. Further, it is assumed that this rate only holds ex ante (as a

[2] Equity financing is therefore considered here as a residual in the present formulation of the firms' behaviour.

[3] We assume in addition that initial conditions for the application of these rules are such that we get from (22)–(24) the stock relationship $M = \phi p K$.

[4] Money receipts—injections of new money—are assumed to be immediately channeled to the banks and held there as demand or time deposits.

Table 1 The asset markets structure of the economy

	Assets	Liabilities
Central banks	High powered money M	Deposits of commercial banks Λ^c
Commercial banks	Bank reserves Λ^c	Interest free deposits of the public Λ^o
	Bank loans to firms Λ^r	Interest bearing deposits of the public Λ^r
Firms	Value of capital stock $p_k K$	Loans from banks Λ
		Value of equity stock $p_e E$
Asset owners	Interest free deposits Λ^o	Nominal wealth W
	Interest bearing deposits Λ^r	
	Equities $p_e E$	

market clearing signal), while the ex post rate includes profits or losses of commercial banks $r\Lambda - r\Lambda^r$ which are distributed as windfall profits to asset owning households ($r^a = r\Lambda/\Lambda^r$). Commercial banks are thus pure intermediaries (which hold the high-powered money—issued by way of the government expenditure rule—as reserve for transaction processes) with no costs of production, no investment and no income, since the interest they receive from firms $r\Lambda$ is completely distributed to asset owning households.

The above scheme implies for the nominal wealth of households the identity:

$$W = M + \Lambda + p_e E = \Lambda^o + \Lambda^r + p_e E = M + p_k K,$$

and for its intended and realised change (at current market prices):

$$pS_p = \dot{\Lambda}^o + \dot{\Lambda}^r + p_e \dot{E} = \dot{M} + \dot{\Lambda} + p_e \dot{E},$$

which explains the use of the wealth concept employed above and the allocation rule for the savings plans made by households.

On the basis of these identities we can now describe the reallocation of wealth that takes place in each moment of time (before the trading of any flows) by means of

$$p_e E = e(\rho^e + \pi^e, r)W, \quad e_1 > 0, e_2 < 0, \tag{25}$$

$$\Lambda^o = d^o(\rho^e + \pi^e, r)W, \quad d_1^o < 0, d_2^o > 0, \tag{26}$$

$$\Lambda^r = d^r(\rho^e + \pi^e, r)W, \quad d_1 < 0, d_2 > 0, \tag{27}$$

with $e(\cdot) + d^o(\cdot) + d^r(\cdot) = 1$, and $e(\cdot)$, $d^o(\cdot)$ and $d^r(\cdot) \geq 0$. The portions of nominal wealth that are desired to be held as equities, interest-free and interest-bearing deposits depend on two rates of return: the expected nominal rate of profit $\rho^e + \pi^e$ and the current market rate of interest r.[5]

[5] As we assume that the two types of deposit holdings, and their particular composition does not feed back into the rest of the economy's structure, both of them will be neglected in the following.

As in Franke and Semmler, equity prices are determined by the canonical equation

$$p_e E = \frac{(\rho^e + \pi^e) pK}{r}. \tag{28}$$

Accordingly, as in Franke and Semmler (1989), we assume here the rate of interest and the equity prices are jointly determined by Eqs. (25) and (28), and thus that they adjust automatically to guarantee asset market equilibrium in every moment in time.[6] In contrast, the composition of $\Lambda^o + \Lambda^r = M + \Lambda$ is chosen so as to be in line with the above allocation rules for these two asset demands.

Expressed in a different way the equilibrium situation on the markets for financial assets can also be derived and described as follows. The condition that the equity market clears is equivalent to assuming the following condition for interest-free and interest bearing deposits

$$M + \Lambda = (d^o(\rho^e + \pi^e, r) + d^r(\rho^e + \pi^e, r))pW$$

and to assume that the interest rate on the latter deposits adjusts in order to get this equality between the high-powered money plus loans to firms and total deposit demand. The interest rate determination is thus described here through market forces, but is of course the same as the one we have determined through equity market equilibrium described by Eq. (25). Again the division of deposits in interest-free and interest-bearing deposits is made on the basis of the rates r and $\rho^e + \pi^e$ and is of no consequence for the working of the economy. The public is therefore completely free to choose the degree of liquidity of its financial assets.

Concerning the determination of wage and price inflation, and thus of the real wage, following the work by Flaschel and Krolzig (2006), the wage and price inflation rates are assumed to be determined by the disequilibrium situations in the labour and goods markets, cross-over expectational terms and the medium-term inflationary climate, i.e.

$$\hat{w} = \beta_w(e - \bar{e}) + \kappa_w \hat{p} + (1 - \kappa_w)\pi^e \tag{29}$$

$$\hat{p} = \beta_p(u - \bar{u}) + \kappa_p \hat{w} + (1 - \kappa_p)\pi^e \tag{30}$$

$$\dot{\pi}^e = \beta_{\pi_1}(\hat{p} - \pi^e) + \beta_{\pi_2}(\hat{M} - \gamma - \pi^e) \tag{31}$$

where $e = L^d/L = Y/(xL)$ represents the economy's employment rate, and π^e the medium-run inflationary climate in the economy.

[6] Since we make use of Metzlerian delayed output adjustment in place of an IS-equilibrium condition, we cannot assume as Franke and Semmler (1989) that the actual rate of profit automatically adjusts to bring about equilibrium in the goods markets.

3 The Model in Intensive Form

Since the current model describes the dynamics of a growing economy, it is convenient to express the main dynamic equations of the model in intensive form. Accordingly, aggregate demand per unit of capital (see Eqs. (7), (13), (22) and (9)) is given by

$$y^d = vy + (1 - s_c)[\rho + r\lambda] + (1 + \phi)(i_1(\rho + \psi - r + \pi^e)) + i_2(u - \bar{u}) + \gamma + \delta, \tag{32}$$

where $v = \omega/x$ is the wage share and $\rho = (1 - v)y - \delta - r\lambda$ is the profit rate in intensive form terms. It should be pointed out that this expression would give rise to the same type of IS-equation as employed in Franke and Semmler (1989) [p. 45] if goods' market equilibrium was assumed in place of the inventory adjustment process assumed here. The IS-equation that is implicitly contained in our model is therefore very close to the form used by Franke and Semmler (1989), who however use a different measure for relative profitability in the investment demand function, and who set i_2 and γ equal to zero.

The condition for asset market equilibrium given by Eq. (25) can be rewritten in intensive form as:

$$\rho^e + \pi^e = e(\rho^e + \pi^e, r)(\rho^e + \pi^e + r), \tag{33}$$

by making use of $p_e E/(pK) = (\rho^e + \pi^e)/r$ as given by Eq. (28), see also Franke and Semmler (1989) [p. 43].

The complete model in intensive form is given by the following system of nonlinear dynamical equations:

$$\hat{\omega} = \kappa[(1 - \kappa_p)\beta_w(e - \bar{e}) + (\kappa_w - 1)\beta_p(u - \bar{u})], \tag{34}$$

$$\hat{l} = -i_1(\rho + \psi - r + \pi^e) - i_2(u - \bar{u}), \tag{35}$$

$$\hat{m} = \phi g/m - \pi^e - \gamma - \kappa[\beta_p(u - \bar{u}) + \kappa_p\beta_w(e - \bar{e})] + \hat{l}, \tag{36}$$

$$\dot{\pi}^e = \beta_{\pi_1}\kappa[\beta_p(u - \bar{u}) + \kappa_p\beta_w(e - \bar{e})] + \beta_{\pi_2}(\phi g/m - \gamma - \pi^e), \tag{37}$$

$$\dot{y}^e = \beta_{y^e}(y^d - y^e) + \gamma y^e, \tag{38}$$

$$\dot{v} = y - y^d - (i_1(\rho + \psi - r + \pi^e) + i_2(u - \bar{u}) - \gamma)v, \tag{39}$$

$$\hat{\lambda} = d(\rho + \psi, (r - \hat{p})\lambda) - \pi^e - \kappa[\beta_p(u - \bar{u}) + \kappa_p\beta_w(e - \bar{e})] - \gamma + \hat{l}, \tag{40}$$

$$\dot{\psi} = c(\rho + \psi - r + \pi^e), \lambda), \tag{41}$$

where $u = Y/Y^p = y/y^p$, $l = L/K$, $m = M/(pK)$, $y^e = Y^e/K$, $v = N/K$, and $y = y^d + v$.

As Franke and Semmler (1989) provide a detailed analysis of the temporary equilibrium positions and their evolution in time, and the properties of macroeconomic models similar to the one discussed here have also been extensively studied in Chiarella and Flaschel (2000), Chiarella et al. (2005) and other related work, in

this paper we will focus primarily on numerical simulation of particular submodules of this extended framework, namely the interaction between debt-to-capital ratio λ, the state of confidence ψ (the two dynamical variables considered by Franke and Semmler (1989)) and the real wage ω in the following section, and leave the analysis of the complete macroeconomic framework for future research.

4 Numerical Analysis

As mentioned above, in this section we consider by means of numerical simulations only a subsystem of the complete macroeconomic framework consisting of the core dynamical variables λ, ψ and ω. The investigation of the full model must here be left for future research. Note however that the Metzlerian quantity dynamics as well as the price inflation dynamics have been studied extensively in the work of Chiarella, Flaschel, Franke and Semmler on various levels of their integration into larger models of disequilibrium macro-dynamics.

As previously mentioned, in Franke and Semmler (1989) goods market equilibrium is assumed, as well as a constant wage share. Further, they neglect the dynamics of factor proportions l since there is no feedback of this magnitude into the rest of the dynamics due to the irrelevance of the rate of employment and the rate of capacity utilisation (wages and prices do not change).

By setting $\beta_w = \beta_p = 0$, and $\pi^e = 0$, we can modify our more general framework to reflect these assumptions.

Under the assumption that firms can automatically adjust their production to the exact level of aggregate demand, so that $y = y^d$ in each moment in time goods market equilibrium (in intensive form) is then given as

$$y = \frac{1}{(1-v)(s_c - (1+\phi)i_1) - i_2/y^p} \\ ((1+\phi)i_1(-\delta - r\lambda + \psi - r) - i_2\bar{u} + \gamma + s_c\delta) \tag{42}$$

Concerning the determination of the rate of interest through asset market equilibrium, for the sake of simplicity we assume that $e(\cdot,\ )$ is such that

$$r = \varepsilon_\psi(\psi - \psi_o) - \varepsilon_\lambda(\lambda - \lambda_o). \tag{43}$$

On the basis of such temporary equilibrium position the evolution of the dynamically endogenous variables λ and ψ which gives rise to an autonomous differential equation system in these variables can then be considered under appropriate assumptions for the solution of the goods and asset markets.

The dynamical equations of the model are in the considered special situation:

$$\dot{\lambda} = [d(\rho + \psi, \lambda) - (i_1(\rho + \psi - r) + i_2(y/y^p - \bar{u}) + \gamma) - \hat{p}]\lambda, \tag{44}$$

$$\dot{\psi} = c(\rho - r, \lambda), \tag{45}$$

Table 2 Parameter values

Steady state values					
$s_c = 0.3$	$y^p = 0.4$	$u_o = 0.91$	$\delta = 0.01$	$\gamma = 0.02$	$i_o = 0.089$
$\phi = 0.0$	$\lambda_o = 0.3$	$\psi = 0.1389$	$x = 5$	$\omega_o = 3.33$	
Behavioural parameters					
$i_1 = 0.0581$	$i_2 = 0.2$	$\alpha_{d1} = 0.25$	$\alpha_{d2} = 0.01$	$\beta_\psi = 0.2$	$\beta_\lambda = 0.1$
		$\beta_u = 0.2$	$\beta_v = 0.1$		

$$\hat{\omega} = \kappa[(1 - \kappa_p)\beta_w(e - \bar{e}) + (\kappa_w - 1)\beta_p(u - \bar{u})]. \tag{46}$$

This dynamical model is basically the model that is investigated in Franke and Semmler (1989) from the analytical as well as from the numerical point of view. Differences to our formulation concern our additional investment term $i_2(\cdot)$, and thus the somewhat different specifications that are used for the functions $d(\cdot)$, $c(\cdot)$ here, and the endogenisation of the rate of profit through aggregate demand-driven business dynamics and dividends, as well as of the real wage in a dynamic set-up. The general model thus exhibits in its analytical core a financial dynamic—based on an IS-LM theory of the rate of profit and the rate of interest—which is capable of generating persistent fluctuations around a well-defined steady state.

Table 2 summarises the parameter values used in the following simulations.

Figure 1 illustrates the dynamic adjustments of the 3D dynamical system given by Eqs. (44)–(46) to an exogenous one-time positive shock to the steady-state debt-to-capital ratio. As it can be clearly observed, the fact that $\lambda_o > \lambda$ for one period leads to the emergence of long-lasting fluctuations in all model variables. In the immediate periods after the shock, aggregate output—and capacity utilisation—increases through the expansion of aggregate investment, which leads to a higher state of confidence in the economy and to a subsequent increase in the firms' indebtedness. As real wages are procyclical (a result of the special choice of the parameter in the real wage equation), the increase in output leads to an increase in the real wage, which in turn would lead ceteris paribus to a decrease in the profit rate.

This and more information can also be retrieved from the 2D cycles depicted in Fig. 1. On the one hand, as the capacity utilisation-real wage cycle shows, there is a clockwise movement (where capacity utilisation is depicted in the x-axis and the real wage in the y-axis) which clearly demonstrates the procyclical behaviour of the real wage, increasing when u increases and vice versa. However, it is also interesting to note that given the error-correction terms present in both wage and price Phillips curves, the real wage tends to fall when $v \geq v_o$ and vice versa. This interaction generates the well-known cyclical behaviour in the real wage, and would, if labour productivity was endogenised as procyclical, generate anticyclical movements of the labour share, as discussed for example in Flaschel (2014).

The dynamics with the most pronounced procyclicality are the interaction between the capacity utilisation and the state of confidence in the economy. This is particularly interesting as this is an emergent feature of the model which results truly from the

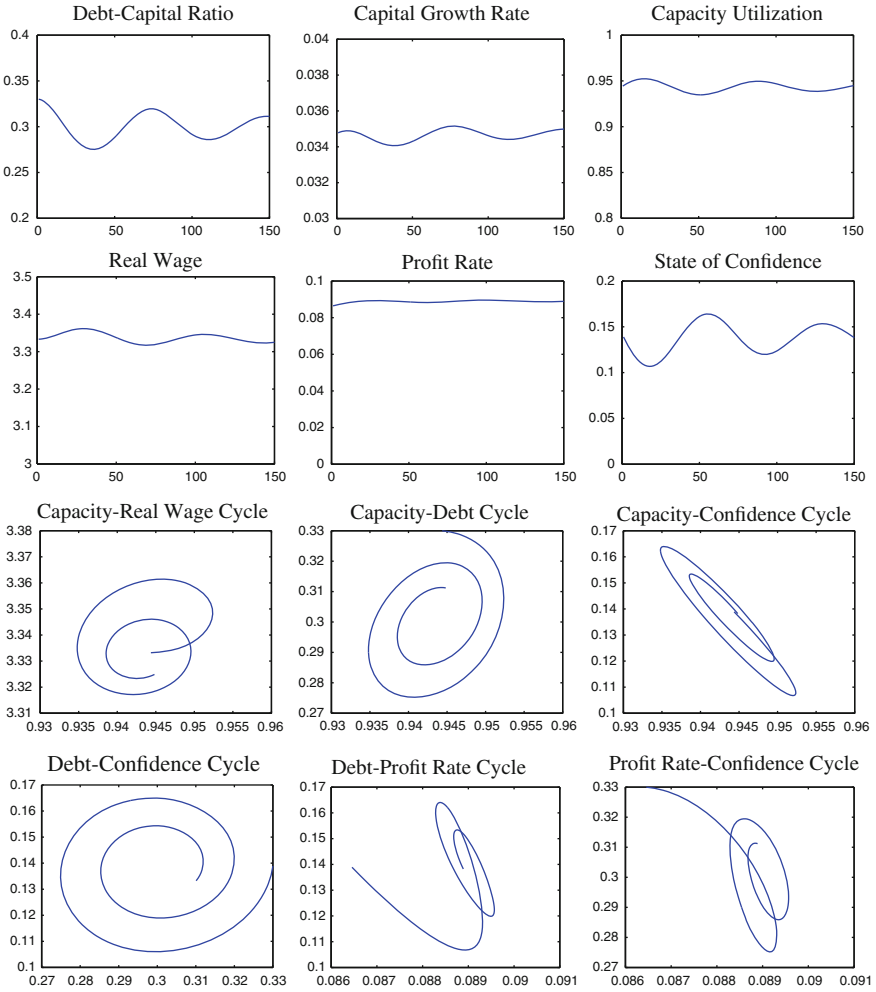

Fig. 1 Dynamic adjustments of the 3D model after an exogenous 10 % increase in the debt-to-capital ratio

interaction of all model variables and not simply by a particular specification of one of the variables.

5 Concluding Remarks

In this contribution we have incorporated the main elements of the small-scale firms' debt-finance model by Franke and Semmler (1989) into a medium-scale disequilib-

rium macroeconomic framework along the lines of Chiarella et al (2005). We illustrated by means of a simple numerical simulation of the resulting framework—which featured investing firms, savings of rentier households, commercial banks and the government—the dynamic feedback mechanisms of endogenously generated debt of firms (created through borrowing) into the investment behaviour of firms, their borrowing of funds, the asset market, the interest rate and the expected rate of return (representing the confidence of investors with regard to future development).

The theoretical framework we developed in this contribution is rich enough to allow for far more complex analytical and numerical investigations. For instance, one could investigate in detail how the model's local stability is affected by increases in certain parameters such as the rentiers' savings rate or the reactive of aggregate investment with respect to expected profitability. Also, a proper empirical analysis of the current framework which would also take into account the possibility time- and/or regime-dependent nature of various key parameters of the model seem like a promising next step of research.

References

Chiarella, C., & Flaschel, P. (2000). *The dynamics of keynesian monetary growth. Macrofoundations*. Cambridge, UK: Cambridge University Press.

Chiarella, C., & Flaschel, P. (2005). *Foundations of a disequilibrium theory of the business cycle*. Cambridge, UK: Cambridge University Press.

Covas, F., & Den Haan, W. (2012). The role of debt and equity finance over the business cycle. *The Economic Journal*, *122*(565), 1262–1286.

Flaschel, P. & Krolzig, H.-M. (2006). Wage-price Phillips curves and macroeconomic stability: Basic structural form, estimation and analysis. In: C. Chiarella, P. Flaschel, R. Franke & W. Semmler (Eds.), *Quantitative and Empirical Analysis of Nonlinear Dynamic Macromodels. Contributions to Economic Analysis 277* (pp. 7–47). Amsterdam: Elsevier.

Flaschel, P. (2014). Goodwin's MKS system. A baseline macro-model. Cambridge Journal of Economics, forthcoming.

Foley, D. (1986). Stabilization policy in a nonlinear business-cycle model. In W. Semmler (Ed.), *Competition, Instability, and Nonlinear Cycles*., Lecture Notes in Economics and Mathematical Systems New York: Springer.

Foley, D. (1987). Liquidity-profit rate cycles in a capitalist economy. *Journal of Economic Behavior and Organization*, *8*(3), 363–377.

Franke, R., & Semmler, W. (1989). Debt financing of firms, stability and cycles in a dynamical macroeconomic growth model. In W. Semmler (Ed.), *Financial Dynamics and Business Cycles: New Perspectives* (pp. 38–64). Armonk, New York: M.E. Sharpe.

Kalecki, M. (1937). A theory of the business cycle. *The Review of Economic Studies*, *4*(2), 77–97.

Minsky, H. P. (1975). *John Maynard Keynes*. New York: Columbia University Press.

Minsky, H. P. (1982). *Can "It" Happen Again? Essays on Instability and Finance*. New York: M. E. Sharpe Inc.

Modigliani, F., & Miller, M. H. (1958). The cost of capital, corporation finance and the theory of investment. *American Economic Review*, *48*(3), 261–297.

Taylor, L. (1985). A stagnationist model of economic growth. *Cambridge Journal of Economics*, *9*, 383–403.

Taylor, L., & O'Connell, S. A. (1985). A Minsky crisis. *Quarterly Journal of Economics*, *100*, 871–886.

Mathematical Modelling of Financial Instability and Macroeconomic Stabilisation Policies

Toichiro Asada

1 Introduction

Minsky's (1975, 1982, 1986) "financial instability hypothesis" implies that the financially dominated capitalist economy is inherently unstable. His hypothesis has been neglected for a long time by the mainstream economists, although it had a considerable influence among the heterodox economists such as Post Keynesians.[1] But, the situation dramatically changed soon after the worldwide financial crisis that was initiated by the so-called "subprime mortgage crisis" in 2007 in the United States. Since then, Minsky's hypothesis was "rediscovered" by some mainstream economists such as Krugman.[2]

Minsky distinguishes three forms of investment financing, that is, "hedge finance", "speculative finance", and "Ponzi finance". He defines these three financing forms as follows.

"If realized and expected income cash flows are sufficient to meet the payment commitments on the outstading liabilities of a unit, then the unit will be hedge financing. However, the balance-sheet cash flows from a unit can be larger than the expected income receipt so that the only way they can be met is by rolling over or even increasing debt; units that roll over debt are engaged in speculative finance and those that increase debt are engaged in Ponzi finance." (Minsky 1986, p. 203)

Minsky provides a description of the business cycle of the financially dominated capitalist economy that is based on the endogenous changes of these three financing

[1] For the Post Keynesian-oriented theoretical literature on Minsky's financial instability hypothesis, see, for example, Asada's (2001, 2004, 2012), Asada et al. (2010), Keen (2000), Nasica (2000), Pally (1996) and Semmler (ed.) (1989).

[2] See, for example, Eggertsoon and Krugman (2012) and Krugman (2012).

T. Asada (✉)
Faculty of Economics, Chuo University, 742-1 Higashinakano, Hachioji, Tokyo 192-0393, Japan
e-mail: asada@tamacc.chuo-u.ac.jp

R. Dieci et al. (eds.), *Nonlinear Economic Dynamics and Financial Modelling*,
DOI: 10.1007/978-3-319-07470-2_5,

forms, that is, Hedge finance→Speculative finance→Ponzi finance→Hedge finance and so on, which is called the "Minsky cycle".

By the way, it is important to note that Minsky did not think that such an inherent instability of the financially dominated capitalist economy is uncontrollable by the government and the central bank. In fact, he stressed that it is important to "stabilize an unstable economy" by means of the proper macroeconomic stabilisation policies by the government and the central bank.[3]

In this paper, we consider how to stabilise an unstable economy theoretically by using the analytical framework of "high dimensional nonlinear Keynesian macrodynamic model" that was developed by Asada et al. (2003, 2010) and Chiarella et al. (2005).[4] In Sect. 2, we formulate the basic Minskian two-dimensional fixed price model of financial instability without active macroeconomic stabilisation policy. In Sect. 3, we consider an extended flexible price four-dimensional model with central bank's monetary stabilisation policy. In Sect. 4, we study a further extended flexible price six-dimensional model of the macroeconomic stabilisation policy by means of monetary and fiscal policy mix. Finally, in Sect. 5, we provide an intuitive economic explanation of the analytical results. Some complicated mathematical proofs are relegated to the appendices.

2 Basic Model: Two-dimensional Model with Fixed Prices

The basic model that is the starting point of our analysis consists of the following system of equations.[5]

$$\dot{d} = \phi(g) - s_f(r - id) - (g + \pi)d;\ 0 < s_f < 1 \tag{1}$$

$$\dot{y} = \alpha(c + \phi(g) + v - y);\ \alpha > 0 \tag{2}$$

$$\dot{g} = g(r, \rho - \pi^e, d);\quad g_r = \partial g/\partial r > 0,\ g_{\rho-\pi^e} = \partial g/\partial(r - \pi^e) < 0,\ g_d = \partial g/\partial d < 0 \tag{3}$$

$$\begin{aligned} c &= (1 - s_1)\{y - r + (1 - s_f)r - \tau(y)\} + (1 - s_2)id + (1 - s_3)\rho b; \\ 0 &< \tau_y = \tau'(y) < 1,\ 0 < s_1 < 1,\ 0 < s_2 \le 1,\ 0 < s_3 \le 1 \end{aligned} \tag{4}$$

$$r = P/K = \beta Y/K = \beta y;\ 0 < \beta < 1 \tag{5}$$

$$i = \rho + \zeta(d) = i(\rho, d);\ \zeta(d) \ge 0,\ i_d = \zeta'(d) > 0 \text{ for } d > 0 \tag{6}$$

[3] See Minsky's (1986) and Asada et al. (2010).

[4] The "high dimensional" dynamic model means the dynamic model with many (at least three) endogenous variables.

[5] This model is essentially based on Asada's (2001) formulation.

$$m = l(y, \rho);\ l_y = \partial l/\partial y > 0,\ l_\rho = \partial l/\partial \rho < 0 \tag{7}$$

$$\pi = \pi^e = 0 \tag{8}$$

$$\rho = \text{constant} > 0 \tag{9}$$

$$v = \text{constant} > 0 \tag{10}$$

The meanings of the symbols are as follows. D = stock of firms' nominal private debt. p = price level. K = real capital stock. $d = D/(pK)$ = private debt-capital ratio. $\pi = \dot{p}/p$ = rate of price inflation. π^e = expected rate of price inflation. $g = \dot{K}/K$ = rate of capital accumulation. $\phi(g)$ = adjustment cost function of investment that has the properties $\phi'(g) \geq 1$, $\phi''(g) > 0$, which was introduced by Uzawa's (1969). P = real profit. $r = P/K$ = rate of profit. i = nominal rate of interest that is applied to firms' private debt. ρ = nominal rate of interest of the government bond. $\rho - \pi^e$ = expected real rate of interest of the government bond. s_f = firms' internal retention rate that is assumed to be constant. Y = real output (real national income). $y = Y/K$ = output-capital ratio, which is a surrogate variable of the "rate of capital utilization" and the "rate of labor employment". G = real government expenditure. $v = G/K$ = government expenditure-capital ratio. B =stock of nominal government bond. $b = B/(pK)$ = government bond-capital ratio. α = quantity adjustment speed of the disequilibrium in the goods market. C = real private consumption expenditure. $c = C/K$ = private consumption expenditure-capital ratio. T = real tax. $\tau = T/K$ = tax-capital ratio. s_1 = average saving rate out of wage and profit income after tax that is assumed to be constant. s_2 = average saving rate out of interest on private debt that is assumed to be constant. s_3 = average propensity to save out of interest on public debt that is assumed to be constant. $\beta = P/Y$ = share of profit in national income that is assumed to be constant ($0< \beta <1$). M = nominal money stock. $m = M/(pK)$ = money stock-capital ratio. L = real money demand. $l = L/K$ = money demand-capital ratio.

We can derive Eq. (1) as follows. The dynamic law of motion of the firms' private debt can be expressed by

$$\dot{D} = \phi(g)pK - s_f(rpK - iD). \tag{11}$$

On the other hand, by differentiating the definitional equation $d = D/(pK)$ by time, we have

$$\dot{d}/d = \dot{D}/D - \dot{p}/p - \dot{K}/K = \dot{D}/D - \pi - g. \tag{12}$$

Substituting Eq. (12) into Eq. (11), we obtain Eq. (1). Equation (2) describes the Keynesian quantity adjustment process of the disequilibrium in the goods market,

which is called the dynamic multiplier process.[6] Equation (3) is the Keynesian type investment function that includes the Fisher (1933) debt effect.[7] Equation (4) is the standard Keynesian type consumption function. In fact, it is assumed that

$$C = (1 - s_1)\{W + (1 - s_f)P - T\} + (1 - s_2)iD + (1 - s_3)\rho B, \tag{13}$$

$$Y = W + P \tag{14}$$

where W is the pre tax real wage income and P is the pre tax real profit. From these equations we have Eq. (4).[8] Equation (5) simply says that the share of profit in national income $\beta = P/Y$ is fixed, which is supposed to be determined by the "degree of monopoly" in the sense of Kalecki (1971). Equation (6) captures the fact that i the interest rate of the "risky assets", will be higher than ρ, the interest rate of the "safer asset", and the difference between them will reflect the degree of risk. Equation (7) is the equilibrium condition for the money market. The function $l(y, \rho)$ is the standard Keynesian real money demand function due to Keynes (1936). Equations (8)–(10) imply that the price level is fixed and both of monetary and fiscal policies are inactive. These assumptions will be relaxed step by step in the subsequent sections.

We can rewrite the system of Eqs. (1)–(10) as follows:

$$\dot{d} = \phi(g(\beta y, \rho, d)) - s_f\{\beta y - i(\rho, d)d\} - g(\beta y, \rho, d)d = f_1(d, y) \tag{15}$$

$$\begin{aligned}\dot{y} = \alpha[&(1 - s_1)\{(1 - s_f\beta)y - \tau(y)\} + (1 - s_2)i(\rho, d)d + (1 - s_3)\rho b \\ &+ \phi(g(\beta y, \rho, d)) + v - y] = \alpha f_2(d, y, b)\end{aligned} \tag{16}$$

$$m = l(y, \rho) \tag{17}$$

In this section, we assume that

$$s_3 = 1 \tag{18}$$

for simplicity of the analysis. In this case, Eqs. (15) and (16) consist of the two-dimensional subsystem of dynamic equations with respect to d and y that is independent of Eq. (17). In such a case, Eq. (17) has the only role to determine the

[6] $E = \phi(g)K$ is the real investment expenditure including the adjustment cost, so that $E/K = \phi(g)$ is the real investment expenditure including the adjustment cost per capital stock. In this formulation, international trade is neglected for simplicity.

[7] Asada's (2001) derived this type of investment function from the firms' profit maximisation behaviours by using both Kalecki's (1937) hypothesis of increasing risk of investment and Uzawa's (1969) hypothesis of increasing adjustment cost of investment, which is called "Penrose effect".

[8] In this formulation, it is assumed that the household is the creditor to both of firms and the government. Furthermore, it is assumed for simplicity that $\tau = T/K$ is independent of id and ρb but it solely depends on y. Incidentally, a possible formulation of the consumption function is $C = (1 - s_1)\{W + (1 - s_f)(P - iD) + iD - T\} + (1 - s_3)\rho B$. In this particular case, we have $s_2 = 1 - (1 - s_1)s_f$ in Eq. (13).

endogenous movement of the variable m, which does not feedback to other subsystem. In other words, this is a decomposable system.

We assume that this system has an equilibrium solution $(d^*, y^*) > (0, 0)$ such that $\dot{d} = \dot{y} = 0$. The Jacobian matrix of this system *at the equilibrium point* becomes as follows:

$$J_1 = \begin{bmatrix} f_{11} & f_{12} \\ \alpha f_{21} & \alpha f_{22} \end{bmatrix} \tag{19}$$

where

$$f_{11} = \partial f_1/\partial d = \{\underbrace{\phi'(g)}_{(+)} - d\} \underbrace{g_d}_{(-)} - g + s_f(\underbrace{i_d}_{(+)} d + i), \tag{20}$$

$$f_{12} = \partial f_1/\partial y = \beta\{(\underbrace{\phi'(g)}_{(+)} - d) \underbrace{g_r}_{(+)} - s_f\}, \tag{21}$$

$$f_{21} = \partial f_2/\partial d = (1 - s_2)(\underbrace{i_d}_{(+)} d + i) + \underbrace{\phi'(g)}_{(+)} \underbrace{g_d}_{(-)}, \tag{22}$$

$$f_{22} = \partial f_2/\partial y = (1 - s_1)(1 - s_f\beta - \tau_y) + \beta \underbrace{\phi'(g)}_{(+)} \underbrace{g_r}_{(+)} - 1, \tag{23}$$

$$\begin{aligned} f_{11}f_{22} - f_{12}f_{21} = &- \underbrace{\phi'(g)}_{(+)} \underbrace{g_d}_{(-)} \{(1 - s_1) \underbrace{\tau_y}_{(+)} + s_1(1 - s_f\beta)\} \\ &+ (1 - s_2)(\underbrace{i_d}_{(+)} d + i)(\beta s_f - d \underbrace{g_r}_{(+)}) + \beta \underbrace{\phi'(g)}_{(+)} \underbrace{g_r}_{(+)} \\ &\{-g + (s_f + s_2 - 1)(\underbrace{i_d}_{(+)} d + i)\} \\ &+ \{-g + s_f(\underbrace{i_d}_{(+)} d + i)\}\{(1 - s_1)(1 - s_f\beta - \underbrace{\tau_y}_{(+)}) - 1\}. \end{aligned} \tag{24}$$

Now, let us assume as follows.

Assumption 1 Assume

$$f_{11} < 0,\ f_{12} > 0,\ f_{21} < 0,\ f_{22} > 0,\ f_{11}f_{22} - f_{12}f_{21} > 0.$$

These inequalities will in fact be satisfied if $\phi'(g)$ and $|g_d|$ are sufficiently large *at the equilibrium point*. Under **Assumption 1,** we have the following proposition.[9]

Proposition 1 *There exists a parameter value* $\alpha_0 > 0$ *that satisfies the following properties*

(1) The equilibrium point of the dynamic system (15)–(18) *is* locally stable *for all* $\alpha \in (0, \alpha_0)$.
(2) The equilibrium point of the dynamic system (15)–(18) *is* locally totally unstable *for all* $\alpha \in (\alpha_0, +\infty)$.
(3) There exist the non-constant closed orbits *around the equilibrium point for some range of the parameter value* α *that is sufficiently close to* α_0.

Proof See Asada (2001).

Proposition 1 (3) means that the endogenous fluctuations occur for some intermediate range of the parameter value α. We can consider that this is a mathematical expression of the "Minsky cycle" that was proposed by Minsky's (1975, 1982, 1986).[10]

3 An Extension: Four-Dimensional Model of Monetary Stabilisation Policy with Flexible Prices

In the model of the previous section, it is assumed that the price level is fixed and the central bank's monetary policy is totally inactive. In this section, we relax these assumptions. We replace the Eqs. (8) and (9) in the previous section with the following equations:

$$\pi = \varepsilon(y - \bar{y}) + \pi^e;\ \varepsilon > 0,\ \bar{y} > 0 \tag{25}$$

$$\dot{\rho} = \begin{cases} \beta_1(\pi - \bar{\pi}) + \beta_2(y - \bar{y}) & \text{if } \rho > 0 \\ \max[0, \beta_1(\pi - \bar{\pi}) + \beta_2(y - \bar{y})] & \text{if } \rho = 0 \end{cases};\ \beta_1 > 0,\ \beta_2 > 0 \tag{26}$$

$$\dot{\pi}^e = \gamma[\xi(\bar{\pi} - \pi^e) + (1 - \xi)(\pi - \pi^e)];\ \gamma > 0,\ 0 \leq \xi \leq 1 \tag{27}$$

[9] In the models in this paper, the "jump variables" are not allowed for unlike the mainstream "New Keynesian" dynamic models that are represented by Woodford (2003), Galí (2008) and others, but it is assumed that all initial conditions of the endogenous variables are historically given. This means that we adopt the traditional notion of the local stability/instability that is popular in the "Old Keynesian" dynamic models represented by Tobin (1994) as well as the "Post Keynesian" models. That is to say, (1) the equilibrium point is considered to be locally stable if all characteristic roots have negative real parts, and (2) it is considered to be locally unstable if at least one characteristic root has positive real part, and (3) it is considered to be locally totally unstable if all characteristic roots have positive real parts. As for the critical assessment of "New Keynesian" dynamic models, see, for example, Asada (2013); Asada et al. (2006, 2010); Chiarella et al. (2013); Flaschel et al. (2008) and Mankiw (2001).

[10] In fact, the point $\alpha = \alpha_0$ is the Hopf Bifurcation point (Gandolfo 2009 p. 481).

Eq. (25) is the quite standard "expectation-argumented price Phillips curve". Equation (26) formalises an interest rate monetary policy rule by the central bank, which is a variant of the "Taylor rule" type monetary policy that considers both of the rate of inflation and the level of real output, which is a surrogate variable of labour employment.[11] In this formulation, the zero bound of the nominal interest rate is explicitly considered. We can consider that this is a type of the flexible inflation targeting monetary policy rule, and $\bar{\pi}$ is the target rate of inflation that is set by the central bank. Equation (27) is a mixed type inflation expectation hypothesis. This is a mixture of the "forward looking" and the "backward looking" (adaptive) inflation expectations. In case of $\xi = 0$, it is reduced to $\dot{\pi}^e = \gamma(\pi - \pi^e)$, which is a purely adaptive inflation expectation hypothesis. On the other hand, in case of $\xi = 1$, it is reduced to $\dot{\pi}^e = \gamma(\bar{\pi} - \pi^e)$, which means that the publics' expected rate of inflation gravitates towards the target rate of inflation that is set and announced by the central bank. We can consider that the parameter value ξ is a measure of the "degree of the credibility" of the central bank's inflation targeting, so that we call it the "credibility parameter".

The model in this section can be reduced to the following system of equations[12]

$$\begin{aligned}\dot{d} &= \phi(g(\beta y, \rho - \pi^e, d)) - s_f\{\beta y - i(\rho, d)d\} \\ &\quad - \{g(\beta y, \rho - \pi^e, d)d + \varepsilon(y - \bar{y}) + \pi^e\}d = F_1(d, y, \pi^e, \rho)\end{aligned} \tag{28}$$

$$\begin{aligned}\dot{y} &= \alpha[(1 - s_1)\{(1 - s_f\beta)y - \tau(y)\} + (1 - s_2)i(\rho, d)d + (1 - s_3)\rho b \\ &\quad + \phi(g(\beta y, \rho - \pi^e, d)) + v - y] = \alpha F_2(d, y, \pi^e, \rho, b)\end{aligned} \tag{29}$$

$$\dot{\pi}^e = \gamma[\xi(\bar{\pi} - \pi^e) + (1 - \xi)\varepsilon(y - \bar{y})] = F_3(y, \pi^e) \tag{30}$$

$$\dot{\rho} = F_4(y, \pi^e) = \begin{cases} \beta_1(\pi^e - \bar{\pi}) + (\beta_1\varepsilon + \beta_2)(y - \bar{y}) & \text{if } \rho > 0 \\ \max[0, \beta_1(\pi^e - \bar{\pi}) + (\beta_1\varepsilon + \beta_2)(y - \bar{y})] & \text{if } \rho = 0 \end{cases} \tag{31}$$

$$m = l(y, \rho) \tag{32}$$

Also in this section, we assume that $s_3 = 1$ for simplicity. In this case, the subsystem (28)–(31) becomes an independent four-dimensional system of dynamic equations with respect to d, y, π^e and ρ. In such a case, the role of Eq. (32) is only to determine the value of m endogenously.

---

[11] For the original exposition of the "Taylor rule", see Taylor (1993).

[12] Equations (29) and (30) imply that $\dot{y}$ is a *decreasing* function of $\rho - \pi^e$, and $\dot{\pi}^e$ is an *increasing* function of y. In other words, this model is immune from the notorious "sign reversals", which are the peculiar characteristics of the "New Keynesian" dynamic model. See, for example, Asada (2013), Asada et al. (2006, 2010), Mankiw (2001).

We can express the equilibrium solution $(d^*, y^*, \pi^{e*}, \rho^*, m^*)$ that satisfies the condition $\dot{d} = \dot{y} = \dot{\pi}^e = \dot{\rho} = 0$ as follows if we neglect the non-negative constraint of ρ.

$$F_1(d^*, \bar{y}, \bar{\pi}, z^* + \bar{\pi}) = 0 \tag{33}$$

$$F_2(d^*\bar{y}, \bar{\pi}, z^* + \bar{\pi}) = 0 \tag{34}$$

$$\pi^* = \pi^{e*} = \bar{\pi} \tag{35}$$

$$y^* = \bar{y} \tag{36}$$

$$\rho^* = z^* + \bar{\pi} \tag{37}$$

$$m^* = l(\bar{y}, z^* + \bar{\pi}) \tag{38}$$

where z^* is the equilibrium real interest rate of the government bond. We can determine the equilibrium values (d^*, z^*) from a system of the simultaneous Eqs. (33) and (34). Incidentally, ρ^* becomes positive *if and only if* the inequality

$$\bar{\pi} > -z^* \tag{39}$$

is satisfied. We assume that this inequality is satisfied. In fact, we assume that $\bar{\pi} > 0$ and $z^* > 0$.

Next, let us study the local stability/instability of the equilibrium point. The Jacobian matrix of the system (28) - (31) *at the equilibrium point* becomes as follows:

$$J_2 = \begin{bmatrix} F_{11} & F_{12} & F_{13} & F_{14} \\ \alpha F_{21} & \alpha F_{22} & \alpha F_{23} & \alpha F_{24} \\ 0 & \gamma\varepsilon(1-\xi) & -\gamma\xi & 0 \\ 0 & \beta_1\varepsilon + \beta_2 & \beta_1 & 0 \end{bmatrix}, \tag{40}$$

where $F_{11} = \partial F_1/\partial d = f_{11} - \bar{\pi}$, $F_{12} = \partial F_1/\partial y = f_{12} + \varepsilon d$, $F_{21} = \partial F_2/\partial d = f_{21}$, $F_{22} = \partial F_2/\partial y = f_{22}$, where f_{11}, f_{12}, f_{21},and f_{22} are defined by Eqs. (20)–(23) in the previous section.

Let us suppose that **Assumption 1** in the previous section is satisfied. Then, we obtain the following set of relationships:

$$F_{11} = \underbrace{f_{11}}_{(-)} - \underbrace{\bar{\pi}}_{(+)} < 0,\ F_{12} = \underbrace{f_{12}}_{(+)} + \varepsilon d > 0,\ F_{21} = \underbrace{f_{21}}_{(-)} < 0,\ F_{22} = \underbrace{f_{22}}_{(+)} > 0, \tag{41}$$

$$F_{11}F_{22} - F_{12}F_{21} = \underbrace{(f_{11}f_{22} - f_{12}f_{21})}_{(+)} - \underbrace{\bar{\pi}}_{(+)}\underbrace{f_{22}}_{(+)} - \varepsilon d\ \underbrace{f_{21}}_{(-)}. \tag{42}$$

Other partial derivatives become as follows:

$$F_{13} = \partial F_1/\partial \pi^e = -\{\underbrace{\phi'(g)}_{(+)} - d\}\underbrace{g_{\rho-\pi^e}}_{(-)} - d, \tag{43}$$

$$F_{14} = \partial F_1/\partial \rho = \{\underbrace{\phi'(g)}_{(+)} - d\}\underbrace{g_{\rho-\pi^e}}_{(-)} + s_f d, \tag{44}$$

$$F_{23} = \partial F_2/\partial \pi^e = -\underbrace{\phi'(g)}_{(+)}\underbrace{g_{\rho-\pi^e}}_{(-)} > 0, \tag{45}$$

$$F_{24} = \partial F_2/\partial \rho = \underbrace{\phi'(g)}_{(+)}\underbrace{g_{\rho-\pi^e}}_{(-)} + (1-s_2)d + (1-s_3)b, \tag{46}$$

$$\begin{aligned} F_{11}F_{23} - F_{13}F_{21} = {} & \underbrace{\phi'(g)}_{(+)}[\underbrace{g_{\rho-\pi^e}}_{(-)}\{(g+\bar{\pi}) - (1-s_f-s_2)(\underbrace{i_d}_{(+)}d + i)\} + d\underbrace{g_d}_{(-)}] \\ & + (1-s_2)(\underbrace{i_d}_{(+)}d + i)(1 - \underbrace{g_{\rho-\pi^e}}_{(-)}), \end{aligned} \tag{47}$$

$$\begin{aligned} F_{11}F_{24} - F_{14}F_{21} = {} & -\underbrace{\phi'(g)}_{(+)}\underbrace{g_{\rho-\pi^e}}_{(-)}\{(g+\bar{\pi}) - (1-s_f-s_2)(\underbrace{i_d}_{(+)}d + i)\} \\ & + [\{\underbrace{\phi'(g)}_{(+)} - d\}\underbrace{g_d}_{(-)} - (g+\bar{\pi}) \\ & + s_f(\underbrace{i_d}_{(+)}d + i)]\{(1-s_2)d + (1-s_3)b\} \\ & + (d\underbrace{g_{\rho-\pi^e}}_{(-)} - s_f d)(1-s_2)(\underbrace{i_d}_{(+)}d + i). \end{aligned} \tag{48}$$

We assume that the following assumption as well as Assumption 1 in the previous section are satisfied.

Assumption 2 Assume $F_{13} > 0$, $F_{14} < 0$, $F_{24} < 0$, $F_{11}F_{22} - F_{12}F_{21} > 0$, $F_{11}F_{23} - F_{13}F_{21} > 0$, $F_{11}F_{24} - F_{14}F_{21} > 0$.

These inequalities will be satisfied if $\phi'(g)$, $|g_{\rho-\pi^e}|$, and i_d are sufficiently large *at the equilibrium point*, ε is sufficiently large and $1 - s_f - s_2 > 0$.

The characteristic equation of the dynamic system (28)–(31) *at the equilibrium point* becomes

$$\Delta_2(\lambda) \equiv |\lambda I - J_2| = \lambda^4 + b_1\lambda^3 + b_2\lambda^2 + b_3\lambda + b_4 = 0, \tag{49}$$

where

$$b_1 = -\text{trace}J_2 = -\underbrace{F_{11}}_{(-)} - \alpha \underbrace{F_{22}}_{(+)} + \gamma\xi, \tag{50}$$

$$\begin{aligned} b_2 &= \text{sum of all principal second order minors of } J_2 \\ &= \alpha \underbrace{(F_{11}F_{22} - F_{12}F_{21})}_{(+)} - \gamma\xi \underbrace{F_{11}}_{(-)} + \alpha\gamma\{\xi \underbrace{F_{22}}_{(+)} - \varepsilon(1-\xi) \underbrace{F_{23}}_{(+)}\} \\ &\quad - \alpha(\beta_1\varepsilon + \beta_2) \underbrace{F_{24}}_{(-)}, \end{aligned} \tag{51}$$

$$\begin{aligned} b_3 &= -\text{sum of all principal third order minors of } J_2 \\ &= \alpha[-\gamma \underbrace{F_{24}}_{(-)}\{\varepsilon(1-\xi)\beta_1 + \xi(\beta_1\varepsilon + \beta_2)\} + (\beta_1\varepsilon + \beta_2) \underbrace{(F_{11}F_{24} - F_{14}F_{21})}_{(+)} \\ &\quad + \gamma\varepsilon(1-\xi) \underbrace{(F_{11}F_{23} - F_{13}F_{21})}_{(+)} + \gamma\xi \underbrace{(F_{11}F_{22} - F_{12}F_{21})}_{(+)}] > 0, \end{aligned} \tag{52}$$

$$b_4 = \det J_2 = \alpha\gamma\{(\beta_1\varepsilon + \beta_2)\xi + \beta_1\varepsilon(1-\xi)\} \underbrace{(F_{11}F_{24} - F_{14}F_{21})}_{(+)} > 0. \tag{53}$$

It is well known that a set of *necessary* (but not sufficient) conditions for the local stability of the equilibrium point of the dynamic system (28)–(31) is given by the following set of inequalities (cf. Asada et al. 2010, Mathematical appendix p. 416).

$$b_j > 0 \text{ for all } j \in \{1, 2, 3, 4\} \tag{54}$$

The following "instability proposition" is a direct corollary of this fact.

Proposition 2 (Instability Proposition) *Suppose that the parameter values* α, β_1 *and* β_2 *are fixed at any positive levels. Furthermore, suppose that (1) the "credibility" parameter of the central bank's inflation targeting* (ξ) *is close to zero (including the case of* $\xi = 0$*), and (2) the adjustment speed of the inflation expectation* (γ) *is sufficiently large. Then, the equilibrium point of the dynamic system* (28)–(32) *becomes* locally unstable.

Proof Suppose that $\xi = 0$. In this case, Eq. (51) becomes

$$b_2 = \alpha\{\underbrace{(F_{11}F_{22} - F_{12}F_{21})}_{(+)} - \gamma\varepsilon \underbrace{F_{23}}_{(+)} - (\beta_1\varepsilon + \beta_2) \underbrace{F_{24}}_{(-)}\}. \tag{55}$$

Then, we have $b_2 < 0$ for all sufficiently large values of $\gamma > 0$, which violates one of the necessary conditions for local stability (54). It must be noted that we have $b_2 < 0$ for all sufficiently large values of $\gamma > 0$ even if $0 < \xi < 1$, as long as ξ is sufficiently close to zero, by continuity. □

On the other hand, we have the following "stability proposition" in contrast to the above "instability proposition".

Proposition 3 (Stability Proposition) *Suppose that (1) the adjustment speed of the goods market disequilibrium* (α) *is sufficiently small, and (2) the "credibility" parameter of the central bank's inflation targeting* (ξ) *is close to 1 (including the case of* $\xi = 1$*). Then, the equilibrium point of the dynamic Eqs.* (28)–(32) *is* locally stable.

Proof Suppose that $\xi = 1$. Then, the Jacobian matrix (40) becomes

$$J_2 = \begin{bmatrix} F_{11} & F_{12} & F_{13} & F_{14} \\ \alpha F_{21} & \alpha F_{22} & \alpha F_{23} & \alpha F_{24} \\ 0 & 0 & -\gamma & 0 \\ 0 & \beta_1\varepsilon + \beta_2 & \beta_1 & 0 \end{bmatrix} \tag{56}$$

In this case the characteristic Eq. (49) becomes as follows:

$$\Delta_2(\lambda) \equiv |\lambda I - J_2| = |\lambda I - J_3|(\lambda + \gamma) = 0, \tag{57}$$

where

$$J_3 = \begin{bmatrix} F_{11} & F_{12} & F_{14} \\ \alpha F_{21} & \alpha F_{22} & \alpha F_{24} \\ 0 & \beta_1\varepsilon + \beta_2 & 0 \end{bmatrix} \tag{58}$$

and

$$|\lambda I - J_3| = \lambda^3 + w_1\lambda^2 + w_2\lambda + w_3 = 0, \tag{59}$$

$$w_1 = -\text{trace}J_3 = -\underbrace{F_{11}}_{(-)} - \alpha \underbrace{F_{22}}_{(+)}, \tag{60}$$

$$\begin{aligned} w_2 &= \text{sum of all principal second-order minors of } J_3 \\ &= \alpha\{\underbrace{(F_{11}F_{22} - F_{12}F_{21})}_{(+)} - (\beta_1\varepsilon + \beta_2)\underbrace{F_{24}}_{(-)}\} > 0, \end{aligned} \tag{61}$$

$$w_3 = -\det J_3 = \alpha(\beta_1\varepsilon + \beta_2)\underbrace{(F_{11}F_{24} - F_{14}F_{21})}_{(+)} > 0, \tag{62}$$

$$w_1 w_2 - w_3 = \alpha\{(\underbrace{F_{14}}_{(-)}\underbrace{F_{21}}_{(-)} - \alpha\underbrace{F_{22}}_{(+)}\underbrace{F_{24}}_{(-)})(\beta_1\varepsilon + \beta_2) + (-\underbrace{F_{11}}_{(-)} + \alpha\underbrace{F_{22}}_{(+)})\underbrace{(F_{11}F_{22} - F_{12}F_{21})}_{(+)}. \tag{63}$$

The characteristic Eq. (57) has a negative real root $\lambda_4 = -\gamma < 0$, and other three roots are determined by Eq. (59). If α is sufficiently small, we have

$$w_j > 0 \text{ for all } j \in \{1, 2, 3\} \text{ and } w_1 w_2 - w_3 > 0, \tag{64}$$

which means that all of the Routh-Hurwitz conditions for stable roots of Eq. (59) are satisfied (cf. Gandolfo 2009, Chap. 16). In this case, all roots of the characteristic Eq. (57) have negative real parts. This conclusion in case of $\xi = 1$ is unchanged even if $0 < \xi < 1$, as long as ξ is sufficiently closed to 1, by continuity. □

Propositions 2 and 3 imply that the increase (the decrease) of the "credibility" parameter of the central bank's inflation targeting (ξ) has a stabilising effect (a destabilising effect) of the macroeconomic system. Suppose that the equilibrium point of the dynamic system (28)–(32) is *locally unstable* in case of $\xi = 0$, and it becomes *locally stable* in case of $\xi = 1$. Then, there exists at least one "bifurcation point" $\xi_0 \in (0, 1)$ at which the switch between "unstable" region and the "stable" region occurs by continuity. It is clear that the real part of at least one characteristic root of Eq. (49) must become zero at the bifurcation point. On the other hand, it follows from Eqs. (49) and (53) that

$$\Delta_2(0) = |-J_2| = \det \quad J_2 = b_4 > 0, \tag{65}$$

which means that the characteristic Eq. (49) cannot have the real root such that $\lambda = 0$. This means that the characteristic Eq. (49) has at least a pair of *pure imaginary roots* at the bifurcation point $\xi = \xi_0$. This means that the *endogenous cyclical fluctuations* occur at some range of the parameter value ξ that is sufficiently close to ξ_0.

4 A Further Extension: Six-Dimensional Model of Monetary and Fiscal Stabilisation Policy Mix with Flexible Prices

In the models of the previous sections, it was assumed that the government expenditure capital ratio (v) is fixed. In this section, we relax this assumption, and study the effect of the monetary and fiscal stabilisation policy mix. In the Eqs. (28)–(32), v is no longer constant, and we add the following equations:

$$M/(pK) = m(\rho)H/K = l(y, \rho) = \phi(\rho)y; \; m_\rho = \mathrm{d}m/\mathrm{d}\rho > 0,$$

$$\varphi_\rho = d\varphi/d\rho < 0, \tag{66}$$

$$pT + \dot{B} + \dot{H} = pG + \rho B, \tag{67}$$

$$\dot{v} = \beta_3[\theta(\bar{y} - y) + (1-\theta)(\bar{b} - b)] = F_5(y, b);\ \beta_3 > 0,\ 0 < \theta < 1, \tag{68}$$

where $M = mH$ = nominal money stock, H = nominal high-powered money that is issued by the central bank, m = money multiplier> 1, $\bar{b}$ = the target value of b that is set by the government. We assume that the private firms and the government are the debtors and the households are the creditors. Equation (66) is the LM equation that describes the equilibrium condition for the money market. The function $l(y, \rho) = \varphi(\rho)y$ is a particular form of the standard Keynesian real money demand function. We can rewrite this equation as

$$h = \psi(\rho)y;\ h = H/(pK),\ \psi(\rho) = \varphi(\rho)/m(\rho),\ \psi'(\rho) = d\psi/d\rho < 0. \tag{69}$$

In our model which supposes that the central bank controls the nominal rate of interest (ρ), the high-powered money-capital ratio (h) becomes an endogenous variable that is determined by Eq. (69). Equation (67) is the budget constraint of the "consolidated government" that includes the central bank. This equation means that the government expenditure including the interest payment of the government bond ($pG + \rho B$) must be financed by (1) tax pT, (2) bond financing ($\dot{B}$), or (3) money financing by the central bank ($\dot{H}$).[13] Equation (68) formalises the government's fiscal policy rule. This equation means that the changes of the real government expenditure respond to both of the real national income (employment) and the level of the public debt. The parameter θ is the weight of the employment consideration rather than the public debt consideration in government's fiscal policy.

Differentiating the definitional equation $b = B/(pK)$ with respect to time and substituting Eq. (67) into it, we obtain[14]

$$\frac{\dot{b}}{b} = \frac{\dot{B}}{B} - \frac{\dot{p}}{p} - \frac{\dot{K}}{K} = \frac{pG + \rho B - pT - \dot{H}}{B} - \pi - g(\beta y, \rho - \pi^e, d). \tag{70}$$

We can rewrite this equation as

$$\dot{b} = v - \tau(y) - \frac{\dot{H}}{pK} + \{\rho - \pi - g(\beta y, \rho - \pi^e, d)\}b;\ \tau = T/K = \tau(y). \tag{71}$$

[13] Also in the models of the previous sections, the definitional Eq. (67) must be satisfied, but this equation has no impact on the dynamics of the main variables in the models of the previous sections as long as $s_3 = 1$.

[14] Note that we have $\dot{K}/K = g(\beta y, \rho - \pi^e, d)$ from the investment function that is formulated in Sect. 2.

This equation plays an important role in the dynamic of the public debt accumulation. If we neglect the impacts of the change of b on the changes of the variables such as $v,\ y,\ \dot{H}/(pK)$ etc., we have

$$\partial \dot{b}/\partial b = \rho - \pi - g. \tag{72}$$

Therefore, the inequality

$$\begin{aligned} &\text{real interest rate of government bond} = \rho - \pi < g \\ &= \text{real rate of capital accumulation} \end{aligned} \tag{73}$$

or equivalently,

$$\begin{aligned} &\text{nominal interest rate of government bond } \rho < g + \pi \\ &= \text{nominal rate of capital accumulation} \end{aligned} \tag{74}$$

is a *stabilizing factor* of the system, and the opposite inequality is a *destabilizing factor* of the system. The (partial) stabilising condition (73) or (74) is called the "Domar condition" after Domar's (1957).[15]

Next, differentiating the definitional expression $h = H/(pK)$ with respect to time, we obtain the following expression.

$$\frac{\dot{H}}{pK} = (\pi + \frac{\dot{K}}{K})h + \dot{h} = \{\pi + g(\beta y, \rho - \pi^e, d)\}h + \dot{h} \tag{75}$$

On the other hand, differentiating Eq. (69) with respect to time and substituting Eqs. (29) and (31) in Sect. 3, we obtain

$$\dot{h} = \underbrace{\psi'(\rho)}_{(-)} y\dot{\rho} + \psi(\rho)\dot{y} = \underbrace{\psi'(\rho)}_{(-)} F_4(y, \pi^e) + \psi(\rho)\alpha F_2(d, y, \pi^e, \rho, v, b). \tag{76}$$

Substituting Eqs. (25), (69), (75), and (76) into Eq. (71), we obtain the following equation that governs the dynamic of the variable b.[16]

$$\begin{aligned} \dot{b} &= v - \tau(y) - \{\varepsilon(y - \bar{y}) + \pi^e + g(\beta y, \rho - \pi^e, d)\}\psi(\rho)y - \underbrace{\psi'(\rho)}_{(-)} yF_4(y, \pi^e) \\ &\quad -\psi(\rho)\alpha F_2(d, y, \pi^e, \rho, v, b) + \{\rho - \varepsilon(y - \bar{y}) - \pi^e - g(\beta y, \rho - \pi^e, d)\}b \\ &= F_6(d, y, \pi^e, \rho, v, b) \end{aligned} \tag{77}$$

---

[15] There is a slight difference between the original "Domar condition" and our "Domar condition". In Domar's (1957) original model, the dynamic stability of the ratio $B/(pY)$ rather than the ratio $b = B/(pK)$ is studied, so that in original Domar model, g is not $\dot{K}/K$ but it is $\dot{Y}/Y$.

[16] This means that the equilibrium condition for the money market (69) affects other parts of the system through Eq. (77) so that the dynamic system in this section is no longer the decomposable system.

Equations (28)–(31) with variable v and b together with Eqs. (68) and (77) constitute a complete system of six-dimensional nonlinear differential equations.[17] We can summarise the system in this section as follows.[18]

$$\dot{d} = F_1(d, y, \pi^e, \rho) \tag{78}$$

$$\dot{y} = F_2(d, y, \pi^e, \rho, v, b) \tag{79}$$

$$\dot{\pi}^e = F_3(y, \pi^e) \tag{80}$$

$$\dot{\rho} = F_4(y, \pi^e) \tag{81}$$

$$\dot{v} = F_5(y, b) \tag{82}$$

$$\dot{b} = F_6(d, y, \pi^e, \rho, v, b) \tag{83}$$

The equilibrium solution of this system $(d^*, y^*, \pi^{e*}, \rho^*, v^*, b^*)$ that satisfies $\dot{d} = \dot{y} = \dot{\pi}^e = \dot{\rho} = \dot{v} = \dot{b} = 0$ can be expressed by the following system of equations.

$$F_1(d^*, \bar{y}, \bar{\pi}, \rho^*) = 0 \tag{84}$$

$$F_2(d^*, \bar{y}, \bar{\pi}, \rho^*, v^*, \bar{b}) = 0 \tag{85}$$

$$\pi^{e*} = \pi^* = \bar{\pi}, \ \ y^* = \bar{y}, \ \ b^* = \bar{b} \tag{86}$$

$$\begin{aligned} v^* = \tau(\bar{y}) + \{\bar{\pi} + g(\beta\bar{y}, \rho^* - \bar{\pi}, d^*)\}\psi(\rho^*)\bar{y} \\ + \{g(\beta\bar{y}, \rho^* - \bar{\pi}, d^*) + \bar{\pi} - \rho^*\}\bar{b} = v^*(d^*, \rho^*, \bar{y}, \bar{\pi}, \bar{b}) \end{aligned} \tag{87}$$

The system of simultaneous Eqs. (84), (85) and (87) determines the equilibrium values (d^*, ρ^*, v^*). We *assume* that there exists the unique equilibrium point that satisfies

$$d^* > 0, \ \ \rho^* > 0, \ \ v^* > 0. \tag{88}$$

In addition to **Assumptions 1** and **2** in the previous sections, let us assume as follows.

Assumption 3 Assume

$$0 < \rho^* - \bar{\pi} < g(\beta\bar{y}, \rho^* - \bar{\pi}, d^*).$$

This assumption implies that the equilibrium real interest rate of the government bond is positive and the "Domar condition" (73) is satisfied *at the equilibrium point*.

[17] Unlike the previous sections, we do *not* necessarily assume that $s_3 = 1$ in this section.

[18] This six-dimensional system is a generalised version of the five-dimensional system that is formulated by Asada (2013), which does not consider the explicit dynamic of the variable d.

Next, let us consider the local stability/instability of the equilibrium point. We can express the Jacobian matrix of the dynamic system (78)–(83) *at the equilibrium point* as

$$J_4 = \begin{bmatrix} F_{11} & F_{12} & F_{13} & F_{14} & 0 & 0 \\ \alpha F_{21} & \alpha F_{22} & \alpha F_{23} & \alpha F_{24} & \alpha & \alpha(1-s_3)\rho^* \\ 0 & \gamma\varepsilon(1-\xi) & -\gamma\xi & 0 & 0 & 0 \\ 0 & \beta_1\varepsilon+\beta_2 & \beta_1 & 0 & 0 & 0 \\ 0 & -\beta_3\theta & 0 & 0 & 0 & -\beta_3(1-\theta) \\ F_{61} & F_{62} & F_{63} & F_{64} & F_{65} & F_{66} \end{bmatrix}, \tag{89}$$

where F_{ij} $(i, j = 1, 2, 3, 4)$ are the same as those in the previous section, and other relevant partial derivatives at the equilibrium point become as follows.[19]

$$F_{61} = \partial F_6/\partial d = -\underbrace{g_d}_{(-)}\{\psi(d^*)\bar{y} + \bar{b}\} - \psi(\rho^*)\alpha\underbrace{F_{21}}_{(-)} > 0, \tag{90}$$

$$F_{64} = \partial F_6/\partial\rho = -\underbrace{g_{\rho-\pi^e}}_{(-)}\{\psi(\rho^*)\bar{y} + \bar{b}\} - \{\bar{\pi} + g(\beta\bar{y}, \rho^* - \bar{\pi}, d^*)\}\underbrace{\psi'(\rho^*)}_{(-)}\bar{y} - \psi(\rho^*)\alpha\underbrace{F_{24}}_{(-)} > 0, \tag{91}$$

$$F_{65} = \partial F_6/\partial v = 1 - \psi(\rho^*)\alpha, \tag{92}$$

$$F_{66} = \partial F_6/\partial b = \rho^* - \bar{\pi} - g(\beta\bar{y}, \rho^* - \bar{\pi}, d^*) < 0, \tag{93}$$

$$\begin{aligned} F_{14}F_{61} - F_{11}F_{64} = {} & -s_f d[\underbrace{g_d}_{(-)}\{\psi(d^*)\bar{y} + \bar{b} + \underbrace{\phi'(g)}_{(+)}\} + \psi(\rho^*)\alpha\{(1-s_2)(\underbrace{i_d}_{(+)}d + i)] \\ & + \{\underbrace{\phi'(g)}_{(+)} - d\}[-\underbrace{g_{\rho-\pi^e}}_{(-)}\psi(\rho^*)\{(1-s_2)(\underbrace{i_d}_{(+)}d + i) + \underbrace{\phi'(g)}_{(+)}\underbrace{g_d}_{(-)}\} \\ & + \underbrace{g_d}_{(-)}(g + \bar{\pi})\underbrace{\psi'(\rho^*)}_{(-)}\bar{y}] \\ & + \{-(g + \bar{\pi}) + s_f(\underbrace{i_d}_{(+)}d + i)\}[\underbrace{g_{\rho-\pi^e}}_{(-)}\{\psi(\rho^*)\bar{y} + \bar{b}\} + (g + \bar{\pi})\underbrace{\psi'(\rho^*)}_{(-)}\bar{y}]. \end{aligned} \tag{94}$$

Now, we shall assume that the following inequality is satisfied.

[19] The values of F_{62} and F_{63} are irrelevant for our purpose.

Assumption 4 Assume

$$F_{14}F_{61} - F_{11}F_{64} > 0$$

This inequality will be satisfied if $\phi'(g)$, $|g_{\rho-\pi^e}|$, $|g_d|$ and $|\psi'(\rho^*)|$ are sufficiently large *at the equilibrium point*. The characteristic equation of this system at the equilibrium point becomes

$$\Delta_4(\lambda) \equiv |\lambda I - J_4| = \lambda^6 + d_1\lambda^5 + d_2\lambda^4 + d_3\lambda^3 + d_4\lambda^2 + d_5\lambda + d_6 = 0, \quad (95)$$

$$d_1 = -\text{trace}J_4, \quad (96)$$

$$d_j = (-1)^j(\text{sum of all principal jth order minors of}J_4) \ (j = 2, 3, 4, 5), \quad (97)$$

$$d_6 = \det J_4. \quad (98)$$

It is worth noting that the conditions

$$d_j > 0 \text{ for all } j \in \{1, 2, \cdots, 6\} \quad (99)$$

are the *necessary* (but not sufficient) conditions for the local stability of the equilibrium point of the dynamic system (78)–(83) (cf. Gandolfo 2009, Chap. 16).

Under assumptions 1–4, we can prove the following two propositions.[20]

Proposition 4 (Instability Proposition) *Suppose that the following conditions are satisfied.*

(1) The credibility parameter of the central bank's inflation targeting (ξ) *is close to zero.*
(2) The adjustment speed of the inflation expectation (γ) *is sufficiently large.*
(3) The monetary policy parameters (β_1) *and* (β_2) *are close to zero.*
(4) The fiscal policy parameter that describes the weight of employment consideration (θ) *is close to zero.*

Then, the equilibrium point of the dynamic system (78)–(83) *becomes* locally unstable.

Proof See Appendix 1.

Proposition 5 (Stability Proposition) *Suppose that the following conditions are satisfied.*

(1) The adjustment speed of the goods market disequilibrium (α) *is sufficiently small.*

[20] It is worth noting that **Assumptions 3 and 4** are not necessary for the proof of **Proposition 4**, but it is only used for the proof of **Proposition 5**.

(2) *The credibility parameter of the central bank's inflation targeting* (ξ) *is close to 1 (including the case of* $\xi = 1$*).*
(3) *The monetary policy parameters* β_1 *and* β_2 *are non-negative and at least one of them is positive.*
(4) *The fiscal policy parameter* θ *is less than 1, but it is close to 1.*
(5) *The average propensity to save out of the interest on the public debt* (s_3) *is close to 1 (including the case of* $s_3 = 1$*).*

Then, the equilibrium point of the dynamic system (78)–(83) *becomes* locally stable.

Proof See Appendix 2.

Proposition 4 means that the "Domar condition" (**Assumption 3**) is by no means the sufficient condition for the local stability of the equilibrium point of the full six-dimensional system in this section, but it is only a partial stability condition.

5 Concluding Remarks: Economic Interpretation of the Analytical Results

In this final section, we shall provide an intuitive economic interpretation of the analytical results, which are presented in the previous section.

Proposition 4 means that the equilibrium point of the system (78)–(83) tends to become *dynamically unstable* if (1) the central bank's monetary policy is inactive and the central bank's inflation targeting is incredible, and (2) the real government expenditure responds sensitively to the amount of the outstanding public debt rather than the real national income (employment). This proposition characterises an *inappropriate* fiscal and monetary policy mix. We can illustrate this destabilising cumulative disequilibrium process by the following two coexisting positive feedback mechanisms $y \downarrow \Rightarrow y \downarrow$ and $b \uparrow \Rightarrow b \uparrow$.[21]

$$y \downarrow \Rightarrow \tau \downarrow \Rightarrow b \uparrow \Rightarrow v \downarrow \Rightarrow \text{(effective demand per capital stock)} \downarrow \Rightarrow y \downarrow, (FM_1)$$
$$b \uparrow \Rightarrow v \downarrow \Rightarrow \{y \downarrow, \tau \downarrow, H/(pK) \downarrow\} \Rightarrow b \uparrow . (FM_2)$$

In this depression process, the decrease of the government expenditure-capital ratio and the increase of the public debt-capital ratio coexist, and the actual and the expected rates of inflation continue to decline. In this process, the nominal interest rate of the government bond slowly declines and at last, it will reach to its lower

[21] Suppose that the central bank's monetary policy is inactive so that both of the monetary policy parameters β_1 and β_2 are sufficiently small. In this case, the movement of the nominal interest rate of the government bond ρ becomes so sluggish that $h = H/(pK)$ moves to the same direction as that of the movement of y like (FM_2) (see Eq. (69) in the text). This means that the central bank continues to reduce the high-powered money-capital ratio in the process of depression, which has the pro-cyclical destabilising effect.

bound. This theoretical scenario is quite consistent with the so-called "lost twenty years" of the Japanese economy that is characterised by the deflationary depression.[22]

On the other hand, **Proposition 5** means that the equilibrium point of the system (78)–(83) tends to be *dynamically stable* if (1) the central bank's inflation targeting is credible, and (2) the real government expenditure responds sensitively to the real national income(employment) rather than the amount of the outstanding public debt, under certain additional conditions. This proposition characterises an *appropriate* fiscal and monetary policy mix.

We can schematically represent the stabilising negative feedback mechanism of the government's fiscal policy $y \downarrow \Rightarrow y \uparrow$ that responds sensitively to the real national income(employment) rather than the amount of the outstanding public debt as follows.

$$y \downarrow \Rightarrow v \uparrow \Rightarrow (\text{effective demand per capital stock}) \uparrow \Rightarrow y \uparrow . \qquad (FM_3)$$

The central bank's active monetary policy that accompanies the "credible" inflation targeting will enhance this stabilising negative feedback mechanism. We can consider that this is the rationale of new macroeconomic policy in Japan called "Abenomics" that was initiated by Abe administration in 2013.[23]

Acknowledgments This research was financially supported by the Japan Society for the Promotion of Science (Grant-in Aid (C) 25380238) and the MEXT-Supported Program for the Strategic Research Foundation at Private Universities, 2013 - 2017.

Appendix 1: Proof of Proposition 4

Suppose that $\xi = \beta_1 = \beta_2 = \theta = 0$. In this case, we have

$$\begin{aligned} d_2 &= \text{sum of all principal second-order minors of } J_4 \\ &= -\gamma\alpha\varepsilon \underbrace{F_{23}}_{(+)} + A, \end{aligned} \qquad (100)$$

where A is independent of the value of γ. This means that we have $d_2 < 0$ for all sufficiently large values of γ, which violates one of the necessary conditions for local stability (99). By continuity, this conclusion applies even if the parameters ξ, β_1, β_2, and θ are positive, as long as they are sufficiently small. □

[22] For the "lost twenty years" of the Japanese economy, see Krugman (1998) and Asada (2013).

[23] For the detailed exposition of "Abenomics", see General Introduction of Asada (2013).

Appendix 2: Proof of Proposition 5

Step 1. Suppose that $\xi = s_3 = 1$. In this case, the characteristic Eq. (95) becomes

$$\Delta_4(\lambda) \equiv |\lambda I - J_4| = |\lambda I - J_5|(\lambda + \gamma) = 0, \tag{101}$$

$$J_5 = \begin{bmatrix} F_{11} & F_{12} & F_{14} & 0 & 0 \\ \alpha F_{21} & \alpha F_{22} & \alpha F_{24} & \alpha & 0 \\ 0 & \beta_1\varepsilon + \beta_2 & 0 & 0 & 0 \\ 0 & -\beta_3\theta & 0 & 0 & -\beta_3(1-\theta) \\ F_{61} & F_{62} & F_{64} & F_{65} & F_{66} \end{bmatrix} \tag{102}$$

Equation (101) has a negative real root $\lambda_6 = -\gamma$ and other five roots are determined by the equation

$$\Delta_5(\lambda) \equiv |\lambda I - J_5| = 0. \tag{103}$$

Step 2. Next, suppose that $\theta = 1$. In this case, Eq. (103) is reduced to

$$\Delta_5(\lambda) = |\lambda I - J_6|(\lambda - F_{66}) = 0, \tag{104}$$

$$J_6 = \begin{bmatrix} F_{11} & F_{12} & F_{14} & 0 \\ \alpha F_{21} & \alpha F_{22} & \alpha F_{24} & \alpha \\ 0 & \beta_1\varepsilon + \beta_2 & 0 & 0 \\ 0 & -\beta_3 & 0 & 0 \end{bmatrix}. \tag{105}$$

Equation (104) has a negative real root $\lambda_5 = F_{66}$ and other four roots are determined by the following equation.

$$\Delta_6(\lambda) = |\lambda I - J_6| = (\lambda^3 + z_1\lambda^2 + z_2\lambda + z_3)\lambda = 0, \tag{106}$$

$$z_1 = -\underbrace{F_{11}}_{(-)} - \alpha \underbrace{F_{22}}_{(+)}, \tag{107}$$

$$z_2 = \alpha\{\underbrace{(F_{11}F_{22} - F_{12}F_{21})}_{(+)} - \underbrace{F_{24}}_{(-)}(\beta_1\varepsilon + \beta_2) + \beta_3\} > 0, \tag{108}$$

$$z_3 = \alpha\{-\underbrace{F_{11}}_{(-)}\beta_3 + (\beta_1\varepsilon + \beta_2)\underbrace{(F_{11}F_{24} - F_{14}F_{21})}_{(+)}\} > 0, \tag{109}$$

$$\begin{aligned} z_1z_2 - z_3 = \alpha\{&(-\underbrace{F_{11}}_{(-)} - \alpha\underbrace{F_{22}}_{(+)})\underbrace{(F_{11}F_{22} - F_{12}F_{21})}_{(+)} - \alpha\beta_3\underbrace{F_{22}}_{(+)} \\ &+ (\beta_1\varepsilon + \beta_2)(\underbrace{F_{14}}_{(-)}\underbrace{F_{21}}_{(-)} + \alpha\underbrace{F_{22}}_{(+)}\underbrace{F_{24}}_{(-)})\}. \end{aligned} \tag{110}$$

Equation (106) has a real root $\lambda_4 = 0$, and other three roots are determined by the equation

$$\Delta_7(\lambda) \equiv \lambda^3 + z_1\lambda^2 + z_2\lambda + z_3 = 0. \tag{111}$$

Step 3. It is easy to see that all of the following Routh-Hurwitz conditions for stable roots of Eq. (111) are satisfied if α is sufficiently small (cf. Gandolfo 2009 Chap. 16).

$$z_j > 0 \ (j = 1, 2, 3), \ z_1 z_2 - z_3 > 0 \tag{112}$$

Hence, we have just proved the following result. "Suppose that $\theta = 1$. Then, the characteristic Eq. (103) has a real root $\lambda_4 = 0$ and other four roots of this equation have negative real parts under the conditions (1) and (3) of **Proposition 5**." This means that Eq. (103) has at least four roots with negative real parts under the conditions (1) and (3) of **Proposition 5** even if $0 < \theta < 1$, as long as θ is sufficiently close to 1 by continuity. On the other hand, in case of $0 < \theta < 1$, we have

$$\begin{aligned} \Delta_5(0) &= -\prod_{j=1}^{5} \lambda_j = |-J_5| = -\det J_5 \\ &= \alpha(\beta_1\varepsilon + \beta_2)\beta_3(1-\theta)\{F_{65}\underbrace{(F_{11}F_{24} - F_{14}F_{21})}_{(+)} + \underbrace{(F_{14}F_{61} - F_{11}F_{64})}_{(+)}\}, \end{aligned} \tag{113}$$

and F_{65} becomes positive if α is sufficiently small. Therefore, Eq. (113) becomes positive so that we have $\prod_{j=1}^{5} \lambda_j < 0$ if $0 < \theta < 1$ and α is sufficiently small. This means that all roots of Eq. (103) have negative real parts under the conditions (1), (3), and (4) of **Proposition 5**

Step 4. We have just proved the following result. "Suppose that $\xi = s_3 = 1$. Then, all of six characteristic roots of Eq. (95) in the text have negative real parts under the conditions (1), (3), and (4) of Proposition 5." By continuity, this conclusion applies even if $0 < \xi < 1$ and $0 < s_3 < 1$, as long as they are sufficiently close to 1. This proves **Proposition 5**. □

References

Asada, T. (2001). Nonlinear dynamics of debt and capital: A post Keynesian approach. In Y. Aruka & Japan Association for Evolutionary Economics (Eds.), *Evolutionary Controversies in Economics*, pp. 73–87. Tokyo: Springer.

Asada, T. (2004). Price flexibility and instability in a macrodynamic model with a debt effect. *Journal of International Economic Studies, 32*(2), 19–38. (Hosei University, Tokyo).

Asada, T. (2012). Modeling financial instability. *Intervention: European Journal of Economics and Economic Policies*, *9–2*, 215–232.

Asada, T. (2013). An analytical critique of 'New Keynesian' dynamic model. *Post Keynesian Review, 2*(1), 1–28. (http://ns.fujimori.cache.waseda.ac.jp/pkr.html)

Asada, T. (Ed.). (2014). *The development of economics in Japan: From the inter-war period to the 2000s*. London: Routledge.
Asada, T. (2014). Macrodynamics of deflationary depression: A Japanese perspective. In T. Asada, (Ed.) pp. 156–207.
Asada, T., Chen, P., Chiarella, C., & Flaschel, P. (2006). Keynesian dynamics and the wage-price spiral: A baseline disequilibrium model. *Journal of Macroeconomics*, *28*, 90–130.
Asada, T., Chiarella, C., Flaschel, P., & Franke, R. (2003). *Open economy macrodynamics: An integrated disequilibrium approach*. Berlin: Springer.
Asada, T., Chiarella, C., Flaschel, P., & Franke, R. (2010). *Monetary macrodynamics*. London: Routledge.
Asada, T., Chiarella, C., Flaschel, P., Mouakil, P., Proaño, C. R. & Semmler, W. (2010). Stabilizing an unstable economy: On the proper choice of policy measures. *Economics: The Open-Access, Open Assessment E-Journal*, *3–21*, 1–43.
Chiarella, C., Flaschel, P., & Franke, R. (2005). *Foundations for a disequilibrium theory of the business cycle: Qualitative analysis and quantitative assessment*. Cambridge: Cambridge University Press.
Chiarella, C., Flaschel, P., & Semmler, W. (2013). Keynes, dynamic stochastic general equilibrium model, and the business cycle. In R. Kuroki, (ed.) *Keynes and Modern Economics*, pp. 85–116. London: Routledge.
Domar, E. (1957). *Essays in the theory of economic growth*. Oxford: Oxford University Press.
Eggertsoon, G. B., & Krugman, P. (2012). Debt, deleveraging, and the liquidity trap: A Fisher-Minsky-Koo approach. *Quarterly Journal of Economics*, *127–3*, 1469–1513.
Fisher, I. (1933). The debt-deflation theory of great depression. *Econometrica*, *1*, 337–357.
Flaschel, P., Franke, R., & Proaño, C. R. (2008). On equilibrium determinacy in new Keynesian models with staggered wage and price setting. *B. E. Journal of Economics, 8*(31), 1–10.
Galí, J. (2008). *Monetary policy, inflation, and the business cycle: A new Keynesian framework*. Princeton: Princeton University Press.
Gandolfo, G. (2009). *Economic dynamics* (4th ed.). Berlin: Springer.
Kalecki, M. (1937). The principle of increasing risk. *Economica*, *4*, 440–447.
Kalecki, M. (1971). *Selected essays on the dynamics of the capitalist economy*. Cambridge: Cambridge University Press.
Keen, S. (2000). The nonlinear economics of debt deflation. In W. A. Barnett, C. Chiarella, S. Keen, R. Marks & H. Scnabl (eds.) *Commerce, complexity, and evolution*, pp. 83–110. Cambridge: Cambridge University Press.
Keynes, J. M. (1936). *The general theory of employment, Interest and Money*. Macmillan:London.
Krugman, P. (1998). It's Baaack: Japan's slump and the return of the liquidity trap. *Brookings Papers on Economic Activity*, *2*, 137–205.
Krugman, P. (2012). *End this depression now!*. New York: W. W. Norton.
Mankiw, G. (2001). The inexorable and mysterious tradeoff between inflation and unemployment. *Economic Journal, 111*, C45–C 61.
Minsky, H. P. (1975). *John Maynard Keynes*. New York: Columbia University Press.
Minsky, H. P. (1982). *Can 'It' happen again? Essays on instability and finance*. Armonk, New York: M. E. Sharpe.
Minsky, H. P. (1986). *Stabilizing an unstable economy*. New Haven: Yale University Press.
Nasica, E. (2000). *Finance, investment and economic fluctuations: An analysis in the tradition of Hyman Minsky*. Cheltenham, UK: Edward Elgar.
Pally, T. (1996). *Post-Keynesian economics: Debt, distribution and the macroeconomy*. London: Macmillan.
Semmler, W. (Ed.). (1989). *Financial dynamics and business cycles: New perspectives*. Armonk, New York: M. E. Sharpe.
Taylor, J. B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy*, *39*, 195–214.

Tobin, J. (1994). Price flexibility and output stability: An old Keynesian view. In W. Semmler (ed.), *Business cycles: Theory and empirical methods*, pp. 165–195. Boston: Kluwer Academic Publishers.

Uzawa, H. (1969). Time preference and the penrose effect in a two-class model of economic growth. *Journal of Political Economy*, *77*, 628–652.

Woodford, M. (2003). *Interest and prices: Foundations of a theory of monetary policy*. Princeton: Princeton University Press.

Bifurcation Structure in a Model of Monetary Dynamics with Two Kink Points

Anna Agliari, Laura Gardini and Iryna Sushko

1 Introduction

Carl Chiarella in his Ph.D. thesis (Chiarella, 1990) approached the so-called *dynamic instability problem* related to linear macro-economic models where the perfect foresight assumption leads to a saddle point equilibrium. To avoid that the economy ends up on a divergent path the early researchers proposed the adoption of the *jump-variable technique*, which relied upon the presumed full-knowledge by agents of their economic environment. It was assumed that, armed with this knowledge and realizing that the given initial values placed them on a divergent path, the agents would calculate the required change in initial values that would place the economy on the stable branch of the saddle point, from where it would move towards the equilibrium point. If there were some unanticipated changes in some underlying economic parameter (e.g. change in the money supply) that moved the economy to a new equilibrium point, then the agents would calculate the new jump required in order to arrive on the stable branch of the new saddle point.

Chiarella commented such a technique by saying that "*it is not possible to find any satisfactory economic justification for such jumps*"[1] and proposed to introduce some non-linearity in the models. In particular, he started from a historical continuous-time model proposed by Sargent and Wallace (1973) and assumed a non-linear money

[1] We can add that even mathematical justification for such jumps can not be found.

A. Agliari
DISES, Catholic University Piacenza, Piacenza, Italy
e-mail: anna.agliari@unicatt.it

L. Gardini (✉)
DESP, Università di Urbino Carlo Bo, Urbino, Italy
e-mail: laura.gardini@uniurb.it

I. Sushko
Institute of Mathematics NASU, Kiev, Ukraine
e-mail: ira_sushko@hotmail.com

R. Dieci et al. (eds.), *Nonlinear Economic Dynamics and Financial Modelling*,
DOI: 10.1007/978-3-319-07470-2_6,

demand function based on simple portfolio considerations. So doing, the economy is stabilised, with prices tending either to the steady-state point or to a relaxation cycle. He also proposed a discrete version of the same model in order to show that even more complex dynamics can be generated by the basic model of monetary dynamics with perfect foresight. A full analysis of the discrete time model is given in Agliari et al. (2004) where, working in a generic context, it is proved that relaxation cycles may appear and the existence of chaotic dynamics is conjectured.

In this paper, we consider again the discrete version of the Sargent and Wallace model with perfect foresight. We shall show that periodic or complex dynamics may appear even if the linearity of the money demand function is preserved, at least in a certain range. To obtain that, we consider a piecewise linear demand function having the same properties assumed in Chiarella (1990) and we obtain that the monetary dynamics are described by a one-dimensional (1D for short) map having two kink points.

The aim of the study is to describe the possible attractors of the map f and the parameter regions corresponding to their existence, so emphasizing once more the Chiarella's understanding. We also take the occasion to show how piecewise linear models offer a large variety of dynamic behaviours, quite simple to identify and describe analytically.

2 The Monetary Dynamics Model

Economic agents are assumed to allocate their wealth between a physical good and money. The good price adjusts with a lag to excess money demand. Consequently, the evolution of price over time is described by a discrete non-linear model given by

$$p_{t+1} = p_t + \alpha \left(m - p_t - D(E_t\left(\pi_{t+1}\right))\right), \tag{1}$$

where p is the logarithm of the price level, m is the logarithm of the money supply (here assumed constant) and $E_t\left(\pi_{t+1}\right)$ is the expected rate of inflation made at time t. The parameter $\alpha > 0$ is the speed at which prices respond to money market disequilibrium and the function $D\left(.\right)$ is the logarithm of the demand for real money balances.

To close the model given in (1), we shall consider the *perfect foresight* case, in which agents are assumed to be able to forecast at time t the exact value of the future inflation rate

$$E_t\left(\pi_{t+1}\right) = p_{t+1} - p_t. \tag{2}$$

The perfect foresight hypothesis leads to a 1D map in *implicit* form for p_{t+1}, namely

$$p_{t+1} = \alpha m + (1-\alpha)p_t - \alpha D(p_{t+1} - p_t), \tag{3}$$

obtained by substituting (2) in (1).

In Chiarella (1990), due to portfolio considerations, a non-linear money demand function satisfies

Assumption 1

(i) $D'(x) < 0$,
(ii) $\lim_{x\to-\infty} D(x) = D_1$ and $\lim_{x\to\infty} D(x) = -D_2$, where D_1 and D_2 are positive constants.

The existence of the two asymptotes ensures that agents shift their portfolio allocation between money and the physical good towards the physical good (money) as expected inflation tends to $+\infty$ ($-\infty$), while (i) is a standard assumption which conveys the losing of purchasing power of real money when the inflation rate increases. Without loss of generality, choosing a convenient unit of measure for price, it is also satisfied

Assumption 2

(iii) $D(0) = 0$.

This assumption implies that the equilibrium price is $p^* = m$.

The map (3) with Assumption 1 has been studied in Agliari et al. (2004), where it is shown that only if $1 + \alpha D'(0) \geq 0$ the price forward dynamics is univocally defined by the model. Under such an assumption, it is proved that, in the general framework, the equilibrium price can be destabilised, when $\alpha = 2\left(1 + \alpha D'(0)\right)$, and either cycles or complex dynamics can emerge.

In this paper, we assume a piecewise linear money demand, decreasing over a 'normal' range $(-a_R, a_L)$ and constant when the expected inflation rate is beyond these bounds. This means that D is a piecewise linear function given by

$$D(\pi) = \begin{cases} a_R \frac{1-\mu}{\alpha} & \text{if } \pi < -a_R, \\ -\frac{1-\mu}{\alpha}\pi & \text{if } -a_R \leq \pi \leq a_L, \\ -a_L \frac{1-\mu}{\alpha} & \text{if } \pi > a_L, \end{cases} \tag{4}$$

where $a_R > 0$, $a_L > 0$ and $\mu < 1$. The particular choice of the slope of the linear branch allows us to simplify the analysis, the parameter μ being in one-to-one correspondence with $D'(0)$. Substituting the money demand function (4) in the perfect foresight model (3) we obtain

$$p_{t+1} = \begin{cases} \alpha m + (1-\alpha) p_t - a_R (1-\mu) & \text{if } p_{t+1} - p_t < -a_R, \\ \alpha m + (1-\alpha) p_t + (1-\mu)(p_{t+1} - p_t) & \text{if } -a_R \leq p_{t+1} - p_t \leq a_L, \\ \alpha m + (1-\alpha) p_t + a_L (1-\mu) & \text{if } p_{t+1} - p_t > a_L. \end{cases}$$

Now, it is straightforward to observe that if $\mu \geq 0$ the dynamics of the price can be made explicit. We limit our analysis to such a case, considering the map

$$p_{t+1} = \begin{cases} (1-\alpha) p_t + \alpha m - a_R (1-\mu) & \text{if } p_t > m + \frac{a_R}{\alpha}\mu, \\ \left(1 - \frac{\alpha}{\mu}\right) p_t + \frac{\alpha m}{\mu} & \text{if } m + \frac{a_R}{\alpha}\mu \leq p_t \leq m - \frac{a_L}{\alpha}\mu, \\ (1-\alpha) p_t + \alpha m + a_L (1-\mu) & \text{if } p_t < m - \frac{a_L}{\alpha}\mu. \end{cases} \tag{5}$$

Before analyzing the monetary dynamics, we observe that the money supply parameter m only affects the values of price sequences but not their long run behaviours. Indeed we can prove, through the translation $x_t = p_t - m$, that the map (5) is topologically conjugated to the map

$$x_{t+1} = \begin{cases} (1-\alpha)\, x_t + a_L\,(1-\mu) & \text{if } x_t < -\frac{a_L}{\alpha}\mu, \\ \left(1-\frac{\alpha}{\mu}\right) x_t & \text{if } -\frac{a_L}{\alpha}\mu < x_t < \frac{a_R}{\alpha}\mu, \\ (1-\alpha)\, x_t - a_R\,(1-\mu) & \text{if } x_t > \frac{a_R}{\alpha}\mu. \end{cases} \tag{6}$$

It is easy to observe that for $\mu \geq \alpha$ the map (6) is increasing and $x^* = 0$ is its unique stable steady state. Hence, if the money demand function slowly declines with the inflation rate (with slope larger than $1 - \frac{1}{\alpha}$) the price monotonically converges to $p^* = m$. A second simple situation occurs when $\alpha \geq 1$. In such a case the price quickly adjusts to the market disequilibrium and the only possible attractors are either $x^* = 0$ or a cycle of period 2 ($\alpha > 2\mu$), since (6) is a decreasing map.

More interesting is the case $\mu < \alpha < 1$. Indeed, as we shall see in the following section, a rich variety of dynamics of the map (6) can be detected.

3 Bifurcation Structure of the Parameter Space

So, let us consider the family of 1D piecewise linear maps as defined in (6) which we rewrite, for our convenience, as a map $f : \mathbb{R} \to \mathbb{R}$ given by

$$f : x \mapsto f(x) = \begin{cases} f_L(x) = (1-\alpha)x + a_L(1-\mu), & x < d_L, \\ f_M(x) = \left(1-\frac{\alpha}{\mu}\right) x, & d_L \leq x \leq d_R, \\ f_R(x) = (1-\alpha)x - a_R(1-\mu), & x > d_R, \end{cases} \tag{7}$$

where

$$d_L = -a_L \frac{\mu}{\alpha}, \quad d_R = a_R \frac{\mu}{\alpha}$$

are the *border points,* and α, μ, a_L, a_R are real parameters. We restrict our analysis to the following parameter region:

$$P = \{p :\ 0 < \alpha < 1,\ 0 < \mu < \alpha,\ a_L > 0,\ a_R > 0\}, \tag{8}$$

where $p = (\alpha, \mu, a_L, a_R)$ denotes a point in the parameter space. Our aim is to study the bifurcation structure of the region P, that is, to describe possible attractors of the map f and the parameter regions corresponding to their existence. To this end, we apply to the map f the results stated in Panchuk et al. (2013), which are related to a generic 1D continuous piecewise linear map with two border points [see also Maistrenko (1995)].

3.1 Preliminaries

The map f is given by the linear functions f_L, f_M and f_R defined in intervals denoted I_L, I_M and I_R, respectively, where $I_L = (-\infty, d_L)$, $I_M = [d_L, d_R]$ and $I_R = (d_R, \infty)$. The simplest properties of the map f which are satisfied for $p \in P$, where P is given in (8), are the following:

- the map f is a *bimodal* map with maximum point $x = d_L < 0$ and minimum point $x = d_R > 0$ (see Fig. 1);
- the slopes of the outermost branches of f are equal, positive and less than 1, thus, all the orbits of f are *bounded*;
- the *fixed point* $x = x^* = 0$, associated with the middle branch f_M, is the *unique* fixed point of the map f;
- for $p \in P_M$ where
$$P_M = \{p \in P : \alpha < 2\mu\},$$
the fixed point x^* is *globally attracting*;
- for $\alpha = 2\mu$ the fixed point x^* undergoes degenerate flip bifurcation (DFB for short) at which eigenvalue of x^* equals -1 and there is an interval I around x^* such that any point of I, except for the fixed point x^*, is 2-periodic, where $I = [d_L, f_M(d_L)]$ for $a_L < a_R$, $I = [f_M(d_R), d_R]$ for $a_L > a_R$ and $I = I_M$ for $a_L = a_R$.

Considering the regime in which the fixed point x^* is repelling, that holds for $\alpha > 2\mu$, we can define an *invariant absorbing interval* J of map f, inside which the map may be defined either by two adjacent branches, or by all three branches. Namely, there are the following three possibilities:

1. If $p \in D_1$ where
$$D_1 = \left\{ p \in P : 2\mu < \alpha < \left(\frac{a_R}{a_L} + 1\right)\mu,\ a_L < a_R \right\}, \tag{9}$$
then $f_M(d_L) < d_R$, $f_M^2(d_L) < d_L$ so that $J = \left[f_M^2(d_L), f_M(d_L)\right]$. The map f in J is reduced to a skew tent map defined by the branches f_L and f_M (see Fig. 1a). The dynamics of the skew tent map has been studied by many authors, and the bifurcation structure of its parameter space is nowadays completely described (see, e.g. Takens (1987), Ito et al. (1979), Maistrenko et al. (1993), Avrutin et al. (2014)). With regard to map f these results are summarised in the next section.
2. If $p \in D_2$ where
$$D_2 = \left\{ p \in P : 2\mu < \alpha < \left(\frac{a_L}{a_R} + 1\right)\mu,\ a_R < a_L \right\}, \tag{10}$$
then $f_M(d_R) > d_L$, $f_M^2(d_R) > d_R$ so that $J = \left[f_M(d_R), f_M^2(d_R)\right]$. The map f in J is reduced to a skew tent map defined by the branches f_M and f_R (see Fig. 1b), so that the qualitative dynamics are the same of the previous case.

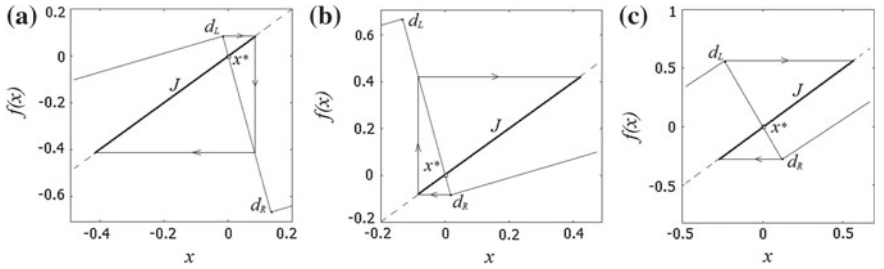

Fig. 1 Invariant absorbing interval J of the map f at $a_L = 0.1$, $a_R = 0.8$, $\alpha = 0.6$, $\mu = 0.1$ in (**a**), $a_L = 0.8$, $a_R = 0.1$, $\alpha = 0.6$, $\mu = 0.1$ in (**b**) and $a_L = 0.8$, $a_R = 0.4$, $\alpha = 0.4$, $\mu = 0.03$ in (**c**)

3. If $p \in D_3$ where

$$D_3 = \left\{ p \in P : \ \alpha > \mu \left(\max \left\{ \frac{a_L}{a_R}, \frac{a_R}{a_L} \right\} + 1 \right) \right\}, \tag{11}$$

then $f_M(d_L) > d_R$, $f_M(d_R) < d_L$ and, thus, $J = [f_M(d_R), f_M(d_L)]$, in which case all the three branches of map f are defined in J, that is, f is bimodal in J (see Fig. 1c). A piecewise linear bimodal map has been studied by other researchers (see, e.g. Maistrenko (1995), Panchuk et al. (2013)), however, its attractors and related parameter regions have not yet been completely described. The known results are applied to the map f given in (7).

Before the detailed description of the cases listed above, we note that for $a_L = a_R$ (in which a case map f is symmetric with respect to the origin) the dynamics of f are quite simple. It is easy to show that an attractor of f is either the attracting fixed point x^* (for $\alpha < 2\mu$) or an attracting 2-cycle (for $\alpha > 2\mu$) with one point belonging to the interval I_L and one point belonging to I_R.

In the case of different values $a_L \neq a_R$, interchanging a_L and a_R in (7) we get a map which is topologically conjugate to f via $x := -x$ (as, for example, the maps shown in Fig. 1a, b). From this fact, it follows that the bifurcation structure of the parameter space of map f is symmetric with respect to $a_L = a_R$, and we can restrict our investigation to the case $a_L < a_R$.

Let $\{x_i\}_{i=1}^n$ be the periodic points of a cycle of map f of period n. To denote this cycle, we use its symbolic representation $\sigma = s_1 s_2 ... s_n$, obtained associating to each point x_i the symbol $s_i \in \{L, M, R\}$ depending on the partition I_L, I_M or I_R which the point x_i belongs to.

Let P_σ denote the region in the parameter space related to existence and stability of the cycle with symbolic sequence σ. Our aim is to describe attracting cycles of map f and the related parameter regions which we call *periodicity regions*. The boundaries of a periodicity region can be related either to stability loss of the cycle due to its eigenvalue crossing ± 1 [we recall that such bifurcations are *degenerate* for a piecewise linear map, see Sushko and Gardini (2010)], or to appearance/disappearance of

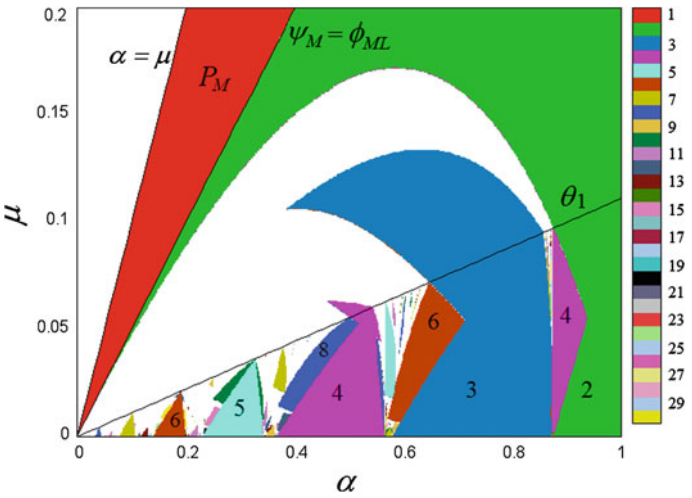

Fig. 2 2D bifurcation diagram of the map f in the (α, μ)-parameter plane for $a_L = 0.1$ and $a_R = 0.8$

the cycle due to a *border collision bifurcation* [see Nusse and Yorke (1995)], which we denote BCB for short. Recall that if some point of a cycle collides with a border point and neither period nor stability of the cycle changes after the collision, we say that this cycle undergoes *persistence border collision*, while a *border collision bifurcation* occurs when a qualitative change in the dynamics is observed after the collision.

Figure 2 shows a 2D bifurcation diagram in the (α, μ)-parameter plane for fixed values $a_L = 0.1$ and $a_R = 0.8$, where some periodicity regions corresponding to attracting cycles of different periods n, for $n \leq 30$, are shown by different colors according to the color bar. White region is related either to higher periodicity or to chaotic attractors. Note that the cycles with the same period but different symbolic sequences are shown by the same color. Given that $a_L < a_R$, the region D_1 defined in (9) is not empty, being confined by two boundaries. The first boundary of D_1 is

$$\psi_M : \qquad \alpha = 2\mu \tag{12}$$

which is defined by the condition $\lambda_M = (1 - \frac{\alpha}{\mu}) = -1$ and related to the DFB of the fixed point x^* already mentioned above (as we shall see, the equality $\psi_M = \phi_{\mathrm{ML}}$ holds where the boundary ϕ_{ML} is related to BCB of the 2-cycle ML). The second boundary of D_1 is

$$\theta_1 : \quad \alpha = \left(\frac{a_R}{a_L} + 1\right)\mu \tag{13}$$

which is defined by the condition $f_M(d_L) = d_R$ and corresponds to the contact of the absorbing interval $J = \left[f_M^2(d_L), f_M(d_L)\right]$ with the border point $x = d_R$.

The region D_1 has the bifurcation structure of a skew tent map, which is described in Sect. 3.2.

3.2 Region D_1 : Skew Tent Map Bifurcation Structure

Let $p \in D_1$ where D_1 is defined in (9). As we already mentioned, in such a case the map f in the absorbing interval $J = \left[f_M^2(d_L),\ f_M(d_L)\right]$ is reduced to the skew tent map defined by the linear branches f_L and f_M. Thus, the symbolic sequences of the cycles of f include only the symbols L and M. It is known that only *basic* cycles of the skew tent map can be attracting. Basic cycles of the map f in the considered case have symbolic sequences ML^{n-1}, $n \geq 2$.

To obtain the boundaries of the periodicity region $P_{\mathrm{ML}^{n-1}}$ first note that a cycle ML^{n-1} may exist not only for $p \in D_1$. It may happen that the map f is defined over the absorbing interval by all three branches (that occurs for $p \in D_3$), while the points of a n-cycle are located in I_L and I_M only. In fact, the region $P_{\mathrm{ML}^{n-1}}$ consists of two parts, $P^{I}_{\mathrm{ML}^{n-1}}$ and $P^{\mathrm{II}}_{\mathrm{ML}^{n-1}}$, where $P^{I}_{\mathrm{ML}^{n-1}} \in D_1$ and $P^{II}_{\mathrm{ML}^{n-1}} \in D_3$. The regions $P^{\mathrm{II}}_{\mathrm{ML}^{n-1}}$, $n \geq 2$, are considered in the next section, while now we consider the regions $P^{I}_{ML^{n-1}}$.

The bifurcation structure of the region D_1 is illustrated by means of 2D and 1D bifurcation diagrams in Figs. 3 and 4, respectively. Applying to the map f the results known for the skew tent map [see Maistrenko et al. (1993), Sushko and Gardini (2010), Avrutin et al. (2014)], we get that the periodicity region $P^{I}_{\mathrm{ML}^{n-1}}$, $n \geq 2$, is bounded from above by the curve $\phi_{\mathrm{ML}^{n-1}}$ and from below by the curve $\psi_{\mathrm{ML}^{n-1}}$ defined as

$$\phi_{\mathrm{ML}^{n-1}}: \qquad \mu = \frac{\alpha^2(1-\alpha)^{n-2}}{(1-\alpha)^{n-2}(2\alpha-1)+1}, \tag{14}$$

$$\psi_{\mathrm{ML}^{n-1}}: \qquad \mu = \frac{\alpha(1-\alpha)^{n-1}}{(1-\alpha)^{n-1}+1}, \tag{15}$$

that is,

$$P^{I}_{\mathrm{ML}^{n-1}} = \left\{ p \in D_1 : \frac{\alpha(1-\alpha)^{n-1}}{(1-\alpha)^{n-1}+1} < \mu < \frac{\alpha^2(1-\alpha)^{n-2}}{(1-\alpha)^{n-2}(2\alpha-1)+1} \right\}.$$

The curve $\phi_{ML^{n-1}}$ for $n \geq 3$ is related to the *fold BCB*[2] leading to the appearance of the basic cycle ML^{n-1} and its complementary cycle M^2L^{n-2}. The curve ϕ_{ML} ($n = 2$) is related to the birth of one cycle, the 2-cycle ML, moreover, $\phi_{\mathrm{ML}} = \psi_M$

[2] Border collision at which two fixed points (one attracting and one repelling, or both repelling) simultaneously collide with the border point (from its opposite sides) and disappear after the collision is called *fold BCB*. It is worth to emphasise that a fold BCB is not associated with an eigenvalue passing through 1.

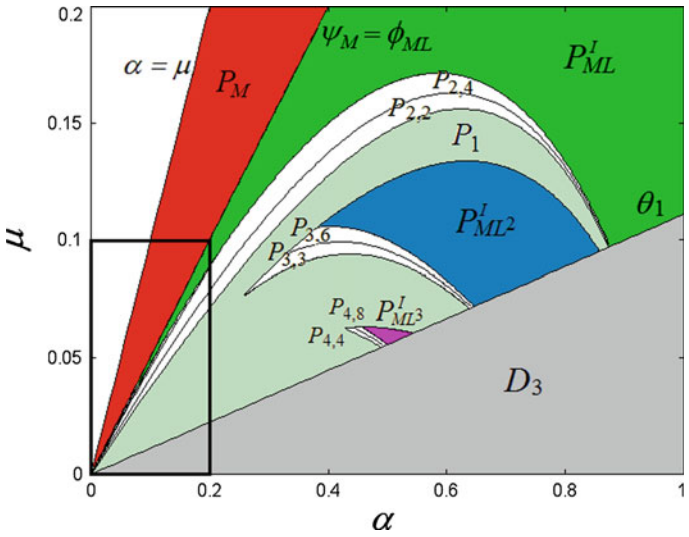

Fig. 3 Skew tent map bifurcation structure of the region D_1 defined in (9) in the (α, μ)-parameter plane at $a_L = 0.1$, $a_R = 0.8$

(see (12)), that is, the BCB of the 2-cycle ML occurs simultaneously with the DFB of the fixed point x^*. The curve $\psi_{\mathrm{ML}^{n-1}}$ is related to the DFB of the cycle ML^{n-1}, defined by the condition $\lambda_{\mathrm{ML}^{n-1}} = -1$. Clearly, the region $P^I_{\mathrm{ML}^{n-1}}$ for some fixed $n \geq 3$ is not empty if the intersection point of $\phi_{\mathrm{ML}^{n-1}}$ and $\psi_{\mathrm{ML}^{n-1}}$ belongs to D_1, so that for fixed values of a_L and a_R, $a_L < a_R$, only a finite number of such regions exist. For example, for $a_L = 0.1$, $a_R = 0.8$, as shown in Fig. 3 (see also Fig. 2), there are only three regions $P^I_{\mathrm{ML}^{n-1}}$, namely, for $n = 2, 3$ and 4 (it is easy to show that the region P^I_{ML} always exists in D_1).

Leaving the region $P^I_{\mathrm{ML}^{n-1}}$ through the boundary $\psi_{\mathrm{ML}^{n-1}}$, $n \geq 3$, the parameter point enters the region denoted $P_{n,2n}$ corresponding to $2n$-cyclic chaotic intervals denoted $Q_{n,2n}$ (here the first index n means that this chaotic attractor is born due to a DFB of the n-cycle, while $2n$ indicates that the chaotic intervals constituting the attractor are $2n$-cyclic). See, for example, the transitions $\mathrm{ML}^2 \overset{\psi_{\mathrm{ML}^2}}{\Rightarrow} Q_{3,6}$ and $\mathrm{ML}^3 \overset{\psi_{\mathrm{ML}^3}}{\Rightarrow} Q_{4,8}$ in Fig. 4. The boundary γ_n between the regions $P_{n,2n}$ and $P_{n,n}$, defined by

$$\gamma_n \ : \quad (1-\alpha)^{2(n-1)} \left(1 - \frac{\alpha}{\mu}\right)^3 + \frac{\alpha}{\mu} - \alpha = 0, \tag{16}$$

is related to the first homoclinic bifurcation of the cycle ML^{n-1}, causing the *merging bifurcation* $Q_{n,2n} \Rightarrow Q_{n,n}$. We have that

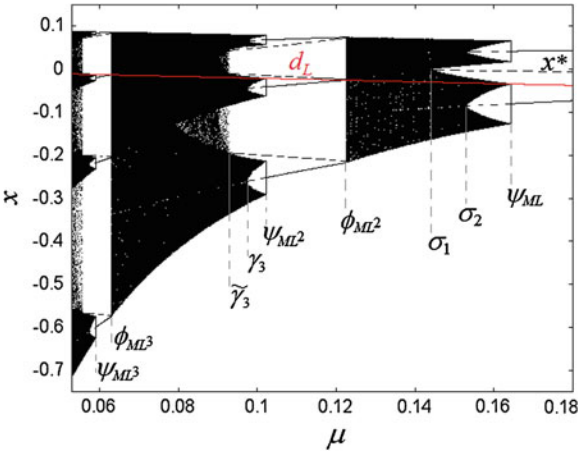

Fig. 4 1D bifurcation diagram of the map f at $a_L = 0.1$, $a_R = 0.8$, $\alpha = 0.48$ and $\mu \in [0.053, 0.18]$ related to the *straight line* with an arrow indicated in Fig. 3

$$P_{n,2n} = \left\{ p \in D_1 : \mu < \frac{\alpha^2(1-\alpha)^{n-2}}{(1-\alpha)^{n-2}(2\alpha-1)+1}, \right.$$
$$\left. \mu < \frac{\alpha(1-\alpha)^{n-1}}{(1-\alpha)^{n-1}+1},\ (1-\alpha)^{2(n-1)}\left(1-\frac{\alpha}{\mu}\right)^3 + \frac{\alpha}{\mu} - \alpha > 0 \right\}.$$

See Fig. 3 where the regions $P_{3,6}$ and $P_{4,8}$ are indicated, and Fig. 4 in which the transitions $Q_{3,6} \overset{\gamma_3}{\Rightarrow} Q_{3,3}$ and $Q_{4,8} \overset{\gamma_4}{\Rightarrow} Q_{4,4}$ can be observed. The boundary $\widetilde{\gamma}_n$ defined as

$$\widetilde{\gamma}_n : \quad (1-\alpha)^{n-1}\left(1-\frac{\alpha}{\mu}\right)^2 - \frac{\alpha}{\mu} + \alpha = 0, \tag{17}$$

is related to the first homoclinic bifurcation of the cycle M^2L^{n-2}, causing the *expansion bifurcation* $Q_{n,n} \Rightarrow Q_1$, where $Q_1 = \left[f_M^2(d_L), f_M(d_L)\right]$. That is, it is a bifurcation from n-cyclic chaotic intervals to a one-piece chaotic attractor. So,

$$P_{n,n} = \left\{ p \in D_1 : \mu < \frac{\alpha^2(1-\alpha)^{n-2}}{(1-\alpha)^{n-2}(2\alpha-1)+1}, \right.$$
$$(1-\alpha)^{2(n-1)}\left(1-\frac{\alpha}{\mu}\right)^3 + \frac{\alpha}{\mu} - \alpha < 0,$$
$$\left. (1-\alpha)^{n-1}\left(1-\frac{\alpha}{\mu}\right)^2 - \frac{\alpha}{\mu} + \alpha < 0 \right\}.$$

See the region $P_{3,3}$ and $P_{4,4}$ in Fig. 3, and the transitions $Q_{3,3} \overset{\widetilde{\gamma}_3}{\Rightarrow} Q_1$ and $Q_{4,4} \overset{\widetilde{\gamma}_4}{\Rightarrow} Q_1$ in Fig. 4.

Besides the bifurcation curves mentioned above, in the region D_1 there is also a set of curves denoted σ_{2^i}, $i \geq 0$, given by

$$\sigma_{2^i} : \left((1-\alpha)^{\delta_i}\left(1-\frac{\alpha}{\mu}\right)^{\delta_{i+1}}\right)^2 + \left((1-\alpha)/\left(1-\frac{\alpha}{\mu}\right)\right)^{(-1)^{i+1}} - 1 = 0, \quad (18)$$

where $\delta_i = (2^i - (-1)^i)/3$. The curve σ_{2^i} for $i \geq 1$ corresponds to the first homoclinic bifurcation of the *harmonic* 2^i-cycle (occurring when the image of the border point merges with the cycle for the first time), causing the merging bifurcation $Q_{2,2^{i+1}} \Rightarrow Q_{2,2^i}$, and the curve σ_1 ($i = 0$) is related to the first homoclinic bifurcation of the fixed point x^* leading to the merging bifurcation $Q_{2,2} \Rightarrow Q_1$. See, for example, the transitions $Q_{2,4} \overset{\sigma_2}{\Rightarrow} Q_{2,2}$ and $Q_{2,2} \overset{\sigma_1}{\Rightarrow} Q_1$ in Fig. 4. The region of existence of the chaotic attractor $Q_{2,2^i}$, $i \geq 1$, is defined by two consecutive homoclinic bifurcation curves and is given by

$$\begin{aligned} P_{2,2^i} = \Bigg\{ p \in D_1 : \; & \left((1-\alpha)^{\delta_{i-1}}\left(1-\frac{\alpha}{\mu}\right)^{\delta_i}\right)^2 \\ & + \left((1-\alpha)/\left(1-\frac{\alpha}{\mu}\right)\right)^{(-1)^i} - 1 < 0, \left((1-\alpha)^{\delta_i}\left(1-\frac{\alpha}{\mu}\right)^{\delta_{i+1}}\right)^2 \\ & + \left((1-\alpha)/\left(1-\frac{\alpha}{\mu}\right)\right)^{(-1)^{i+1}} - 1 > 0, \; \mu < \frac{\alpha(1-\alpha)}{2-\alpha} \Bigg\}. \end{aligned}$$

Figure 5 shows an enlarged window indicated in Fig. 3, where the regions $P_{2,2^i}$, $i = 1, 2, 3, 4$, are marked. The curves $\sigma_{2^i} \in D_1$ for $i \to \infty$ are accumulating to the point $(1-\alpha, 1-\frac{\alpha}{\mu}) = (1, -1)$, that is, to $(\alpha, \mu) = (0, 0)$, as it can be seen in Fig. 5.

If we extract from the region D_1 all the regions introduced above, that is, the regions $P^I_{\mathrm{ML}^{n-1}}$, $P_{n,2n}$, $P_{n,n}$, $n \geq 2$, and $Q_{2,2^i}$, $i \geq 1$, together with their boundaries, the rest of the parameter plane denoted P_1 is related to a one-piece chaotic attractor Q_1. Note that only the boundary θ_1 of the region D_1 (see (13)) depends on the parameters a_L and a_R, while all the other bifurcation curves in D_1 do not depend on these parameters.

In such a way, we have a complete description of the bifurcation structure of the parameter region D_1 given in (9). As remarked above, we have considered $a_L < a_R$. If $a_L > a_R$ a similar description can be given for the region D_2 defined in (10), with the only difference related to symbolic sequences of the cycles, in which the symbol L is to be substituted by the symbol R.

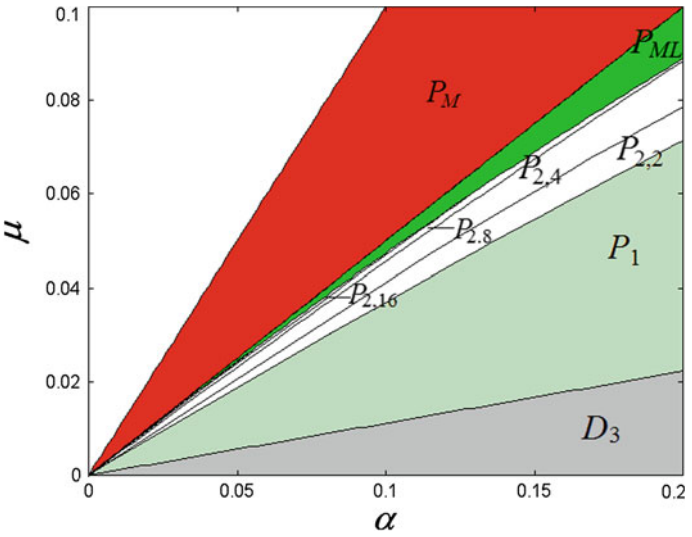

Fig. 5 Enlargement of the window indicated in Fig. 3

3.3 Bifurcation Structure of the Region D_3

Let $a_L < a_R$ and $p \in D_3$ where D_3 is defined in (11). It is not surprising that comparing with the regions D_1 or D_2 the bifurcation structure of the region D_3 is more richer (see, for example, Fig. 2), given that the map f is defined over the absorbing interval $J = [f_M(d_R), f_M(d_L)]$ by all the three branches. In this case, there is not yet such a complete description of the generic bifurcation structure of the parameter space as we have for the skew tent map, especially concerning the occurrence of chaotic attractors. As for the periodicity regions related to attracting cycles, several substructures can be recognised. In particular, as discussed in Panchuk et al. (2013), one can distinguish between *period adding* and *fin* structures. In the following, we limit our description to the first one, and give a simple sketch of the second one.

3.3.1 Period Adding Structure

Let us describe the period adding structure, also called *Arnold tongues* or *mode-locking tongues*, being characteristic for a certain class of circle maps, discontinuous maps defined by two increasing functions, etc. [see, e.g. Leonov (1959), Keener (1980), Boyland (1986)]. The periodicity regions constituting this structure are related to attracting n-cycles, $n \geq 2$, whose points belong only to the intervals I_L and I_R, that is, their symbolic sequences include only the symbols L and R. These

regions are ordered in the parameter space according to the Farey summation rule which is applied to the rotation numbers of the related cycles.

Following Leonov (1959) all the cycles associated with the period adding structure are grouped into certain families, called *complexity levels.* The *complexity level one* includes two families of basic cycles having the following symbolic sequences:

$$\Sigma_{1,1} = \left\{\mathrm{LR}^{n_1}\right\}_{n_1=1}^{\infty}, \quad \Sigma_{2,1} = \left\{\mathrm{RL}^{n_1}\right\}_{n_1=1}^{\infty}. \tag{19}$$

To get the symbolic sequences of the cycles of families of the *complexity level two* we apply to the families $\Sigma_{1,1}$ and $\Sigma_{2,1}$ the *symbolic replacements*

$$\kappa_m^L := \begin{cases} L \to \mathrm{LR}^m \\ R \to \mathrm{RLR}^m \end{cases}, \quad \kappa_m^R := \begin{cases} L \to \mathrm{LRL}^m \\ R \to \mathrm{RL}^m \end{cases}, \tag{20}$$

that is based on the *map replacement technique* [see Avrutin et al. (2010)]. Namely, substituting in $\Sigma_{1,1}$ at first each symbol L by LR^m and each symbol R by RLR^m (replacement κ_m^L), and then substituting in $\Sigma_{1,1}$ the symbol L by LRL^m and the symbol R by RL^m (replacement κ_m^R), where $m = n_2$, we get, respectively, two families of the complexity level two:

$$\Sigma_{1,2} = \left\{\mathrm{LR}^{n_2}\left(\mathrm{RLR}^{n_2}\right)^{n_1}\right\}_{n_1,n_2=1}^{\infty}, \quad \Sigma_{2,2} = \left\{\mathrm{LRL}^{n_2}\left(\mathrm{RL}^{n_2}\right)^{n_1}\right\}_{n_1,n_2=1}^{\infty}.$$

Similarly, applying the replacements κ_m^L and κ_m^R to $\Sigma_{2,1}$ we get the symbolic sequences of two more families:

$$\Sigma_{3,2} = \left\{\mathrm{RLR}^{n_2}\left(\mathrm{LR}^{n_2}\right)^{n_1}\right\}_{n_1,n_2=1}^{\infty}, \quad \Sigma_{4,2} = \left\{\mathrm{RL}^{n_2}\left(\mathrm{LRL}^{n_2}\right)^{n_1}\right\}_{n_1,n_2=1}^{\infty}.$$

In short, this procedure can be written as $\Sigma_{1,2} = \kappa_{n_2}^L(\Sigma_{1,1})$, $\Sigma_{2,2} = \kappa_{n_2}^R(\Sigma_{1,1})$, $\Sigma_{3,2} = \kappa_{n_2}^L(\Sigma_{2,1})$ and $\Sigma_{4,2} = \kappa_{n_2}^R(\Sigma_{2,1})$. In such a way we get four families of complexity level two. Further, applying the replacements (20) with $m = n_3$ to the families of complexity level two we obtain 2^3 families $\Sigma_{j,3}$, $j = 1, \ldots, 2^3$, of *complexity level three*, and so on. In this way, all the symbolic sequences of cycles associated with the period adding structure are obtained.

Now let us turn to the boundaries the periodicity regions related to the cycles of the map f associated with the period adding structure. Obviously, if the map f has such a n-cycle, it is always attracting, with multiplier $\lambda = (1-\alpha)^n < 1$. Thus, its periodicity region can be confined only by the boundaries related to the BCBs, at which a point of the cycle, the one which is nearest to the border point $x = d_L$, collides with it, or a point which is nearest to the border point $x = d_R$, collides with it.

Consider first the basic cycles LR^{n_1} and RL^{n_1} belonging to the families $\Sigma_{1,1}$ and $\Sigma_{2,1}$ of complexity level one. Using the results stated in Panchuk et al. (2013), we get that the periodicity regions $P_{\mathrm{LR}^{n_1}}$ and $P_{\mathrm{RL}^{n_1}}$ are defined as

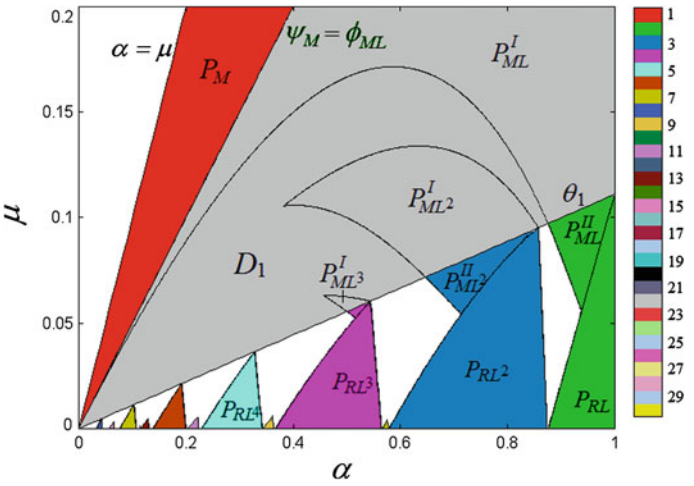

Fig. 6 Periodicity regions P_{RL^n}, $n = 1, ..., 7$, and $P^{II}_{\mathrm{ML}^m}$, $m = 1, 2, 3$, in the (α, μ)-parameter plane at $a_L = 0.1$, $a_R = 0.8$.

$$P_{\mathrm{LR}^{n_1}} = \left\{ p \in D_3 : \Psi_{1,1}(a, \mu_R, d_R, n_1) < \mu_L < \Phi_{1,1}(a, \mu_R, d_L, n_1) \right\},$$

$$P_{\mathrm{RL}^{n_1}} = \left\{ p \in D_3 : \Psi_{1,1}(a, \mu_L, d_L, n_1) > \mu_R > \Phi_{1,1}(a, \mu_L, d_R, n_1) \right\},$$

where

$$a = 1 - \alpha, \ \mu_L = a_L(1 - \mu), \ \mu_R = -a_R(1 - \mu), \tag{21}$$

$$\Phi_{1,1}(a, b, d, n) = -\psi(a, n)b + \varphi(a, n)d,$$

$$\Psi_{1,1}(a, b, d, n_1) = -\left(a + \psi(a, n - 1)\right) b + a\varphi(a, n)d,$$

$$\varphi(a, n) = \frac{1 - a^{n+1}}{a^n}, \quad \psi(a, n) = \frac{1 - a^n}{(1 - a)a^n}. \tag{22}$$

The regions $P_{RL^{n_1}}$ exist for $a_L < a_R$, and in the (α, μ)-parameter plane they originate from the envelope curve θ_1 given in (13) which bounds the region D_1 (see Fig. 6 in which the regions $P_{RL^{n_1}}$ are shown for $n_1 = 1, ..., 7$). The regions $P_{LR^{n_1}}$ exist for $a_L > a_R$, originating from the envelope curve

$$\theta_2 : \quad \alpha = \left(\frac{a_L}{a_R} + 1 \right) \mu \tag{23}$$

which bounds the region D_2.

To get the periodicity regions of complexity level two, we apply the map replacement technique to the periodicity regions of the complexity level one. For the cycles

belonging to the families $\Sigma_{1,2}$ and $\Sigma_{2,2}$ we obtain that

$$P_{\kappa^L_{n_2}(LR^{n_1})} = \{p \in D_3 : \Psi_{1,2}(a, \mu_R, d_R, n_1, n_2) < \mu_L < \Phi_{1,2}(a, \mu_R, d_L, n_1, n_2)\}, \tag{24}$$

$$P_{\kappa^R_{n_2}(LR^{n_1})} = \{p \in D_3 : \Psi_{2,2}(a, \mu_R, d_R, n_1, n_2) < \mu_L < \Phi_{2,2}(a, \mu_R, d_L, n_1, n_2)\}, \tag{25}$$

respectively, where $\kappa^L_{n_2}$ and $\kappa^R_{n_2}$ are defined in (20), a, μ_L, μ_R in (21), $\Phi_{1,2}$, $\Psi_{1,2}$, $\Phi_{2,2}$ and $\Psi_{2,2}$ are defined as

$$\Phi_{1,2}(a, \mu_R, d, n_1, n_2) = -\psi(a, n_2)\mu_R - \frac{a^{n_2+1}\psi(a^{n_2+2}, n_1)\mu_R - \varphi(a^{n_2+2}, a^{n_2+1}, n_1)d}{a^{n_2}(1 + \psi(a^{n_2+2}, n_1))},$$

$$\Psi_{1,2}(a, \mu_R, d, n_1, n_2) = -\frac{a^{n_2+1}\left(a^{n_2+1} + \psi(a^{n_2+2}, n_1 - 1)\right)\mu_R}{a^{n_2}(1 + a^{n_2+1} + \psi(a^{n_2+2}, n_1) - 1)} - \psi(a, n_2)\mu_R + \frac{a^{n_2+2}\varphi(a^{n_2+2}, a^{n_2+1}, n_1)d}{a^{n_2}(1 + a^{n_2+1} + \psi(a^{n_2+2}, n_1) - 1)},$$

$$\Phi_{2,2}(a, \mu_R, d, n_1, n_2) = \frac{\varphi(a^{n_2+1}, a^{n_2+2}, n_1)d - a^{n_2}\left(1 + \psi(a^{n_2+1}, n_1)\right)\mu_R}{a^{n_2}\left(a + \psi(a, n_2)(1 + \psi(a^{n_2+1}, n_1))\right)},$$

$$\Psi_{2,2}(a, \mu_R, d, n_1, n_2) = \frac{a\varphi(a^{n_2+1}, a^{n_2+2}, n_1)d - \left(1 + a^{n_2+2} + \psi(a^{n_2+1}, n_1 - 1)\right)\mu_R}{a + \psi(a, n_2)(1 + a^{n_2+2} + \psi(a^{n_2+1}, n_1 - 1))},$$

where

$$\varphi(a, c, n) = \frac{1 - a^n c}{a^n}, \quad \psi(a, n) = \frac{1 - a^n}{(1 - a)a^n}. \tag{26}$$

To get the periodicity regions for the cycles belonging to the families $\Sigma_{3,2}$ and $\Sigma_{4,2}$, one has to exchange the indices L and R, as well as to change the inequality signs to the opposite ones in (24) and (25), respectively, obtaining

$$P_{\kappa^L_{n_2}(RL^{n_1})} = \{p \in D_3 : \Psi_{1,2}(a, \mu_L, d_L, n_1, n_2) > \mu_R > \Phi_{1,2}(a, \mu_L, d_R, n_1, n_2)\},$$

$$P_{\kappa^R_{n_2}(RL^{n_1})} = \{p \in D_3 : \Psi_{2,2}(a, \mu_L, d_L, n_1, n_2) < \mu_R < \Phi_{2,2}(a, \mu_L, d_R, n_1, n_2)\},$$

In Fig. 6, besides the large periodicity regions P^n_{RL} of complexity level 1, a few small regions are visible between each pair, close to the axis, which are of complexity level two. Following similar procedure all the periodicity regions of the period adding structure can be obtained.

3.3.2 Fin Structure

Coming back to the 2D bifurcation diagram of the map f shown in Fig. 2, we can notice that there are particular periodicity regions which are attached to the regions belonging to the period adding structure described in the previous section. For example, one can clearly see a $2 \cdot 2$-periodicity region attached to the region P_{RL}, a $3 \cdot 2$-periodicity region attached to the region P_{RL^2}, two $4 \cdot 2$-periodicity regions attached to the region P_{RL^3} from its opposite sides, as well as $4 \cdot 3$-periodicity region, two $5 \cdot 2$-periodicity regions attached to the region P_{RL^4}, as well as $5 \cdot 3$-periodicity regions, and so on. These regions belong to so-called *fin structure,* which is formed by the periodicity regions called $n \cdot k$-fins, $k \geq 1$, related to attracting cycles having just one point in the interval I_M and all the other points are in I_L and I_R. The region of the period adding structure to which a fin is attached is called *trunk* region, and its fin has the same complexity level as the complexity level of the trunk. In fact, the periodicity region $P^{II}_{\mathrm{ML}^{n-1}}$, $n \geq 2$, mentioned in Sect. 3.2, is an $n \cdot 1$-fin of complexity level one of the trunk regions $P_{RL^{n-1}}$ (see Fig. 6).

As explained in Panchuk et al. (2013), for an $n \cdot k$-cycle whose periodicity region has the common boundary with the region $P_{\mathrm{LR}^{n-1}}$, the symbolic sequences of the cycles in the fins are either $\mathrm{MR}^{n-1}(\mathrm{LR}^{n-1})^{k-1}$ or $(\mathrm{LR}^{n-1})^{k-1}\mathrm{LR}^{n-2}M$, where $k = 1, 2, \ldots$. Interchanging L and R in these sequences we get the symbolic sequences of the cycles related to $n \cdot k$-fins whose trunks are $P_{\mathrm{RL}^{n-1}}$ regions. Each $n \cdot k$-fin region for $k \geq 2$, $n \geq 2$, has at most four boundaries, among which one is the common BCB boundary with the related trunk region, one boundary is related to DFB of the cycle (whose eigenvalue $\lambda_{n \cdot k} = (1-\alpha)^{nk-1}(1-\frac{\alpha}{\mu})$ is obviously negative) and two other boundaries are related to BCBs. Each $1 \cdot n$-fin region has only three boundaries, namely, one DFB boundary and two BCB boundaries. The DFB boundary is defined by the condition $\lambda_{n \cdot k} = -1$, while BCB boundaries are obtained using skew tent map as border collision normal form (for the details, we refer to Panchuk et al. (2013)).

In Fig. 2, one can see also periodicity regions issuing from the envelope curve θ_1 which are not related to the period adding structure. They are associated with the stabilised cycles whose symbolic sequences are ordered according to the well-known U-sequence (see Metropolis et al. (1973)). Complete description of the related bifurcation structure is still an open problem.

4 Conclusion

The purpose of this paper is twofold. First, coherently with an early Chiarella's work, we show that in a monetary growth model with perfect foresight bounded trajectories may appear even when the unique steady state is unstable. Second, we present quite recent analytical and numerical methods to study piecewise linear models with two kink points.

To achieve our goals we have considered a discrete version of the Sargent and Wallace model and assumed a demand function linear decreasing over a 'normal'

range and constant outside, in the spirit of Chiarella (1990). We have shown that when the slope of the demand function is sufficiently large in absolute value and the speed of adjustment of the price to the market disequilibrium is smaller than 1 either cycles of any period or chaotic dynamics may be generated by the model. In particular, we have described the bifurcation structure of the (α, μ) parameter plane.

Despite the simplicity of the model we consider, our study suggests that piecewise linear models are a very flexible tool to describe different economic phenomena and to obtain interesting analytical results.

References

Agliari, A., Chiarella, C., & Gardini, L. (2004). A stability analysis of the perfect foresight map in nonlinear models of monetary dynamics. *Chaos, Solitons and Fractals*, *21*, 371–386.

Avrutin, V., Gardini, L., Schanz, M., Sushko, I., & Tramontana, F. (2014). *Invariant sets and bifurcation structures*. Springer: Continuous and Discontinuous Piecewise-Smooth One-Dimensional Maps. (in progress).

Avrutin, V., Schanz, M., & Gardini, L. (2010). Calculation of bifurcation curves by map replacement. *International Journal of Bifurcation and Chaos*, *20*, 3105–3135.

Boyland, P. L. (1986). Bifurcations of circle maps: Arnold tongues, bistability and rotation intervals. *Communications in Mathematical Physics*, *106*(3), 353–381.

Chiarella, C. (1990). *The elements of a nonlinear theory of economic dynamics*. New York: Springer.

Homburg, A. J. (1996). Global aspects of homoclinic bifurcations of vector fields. *Memories of the American Mathematical Society* (Vol. 578). Heidelberg: Springer.

Ito, S., Tanaka, S., & Nakada, H. (1979). On unimodal transformations and chaos II. *Tokyo Journal of Mathematics*, *2*, 241–259.

Keener, J. P. (1980). Chaotic behavior in piecewise continuous difference equations. *Transactions of the American Mathematical Society*, *261*, 589–604.

Leonov, N. N. (1959). Map of the line onto itself. *Radiofisika*, *3*, 942–956.

Maistrenko, Y. L., Maistrenko, V. L., Vikul, S. I., & Chua, L. O. (1995). Bifurcations of attracting cycles from time-delayed Chua's circuit. *International Journal of Bifurcation and Chaos*, *5*, 653–671.

Maistrenko, Y. L., Maistrenko, V. L., & Chua, L. O. (1993). Cycles of chaotic intervals in a time-delayed Chua's circuit. *International Journal of Bifurcation and Chaos*, *3*, 1557–1572.

Metropolis, N., Stein, M. L., & Stein, P. R. (1973). On finite limit sets for transformations on the unit interval. *Journal of Combinatorial Theory*, *15*, 25–44.

Nusse, H. E., & Yorke, J. A. (1995). Border-collision bifurcations for piecewise smooth one-dimensional maps. *International Journal of Bifurcation and Chaos*, *5*, 189–207.

Panchuk, A., Sushko, I., Schenke, B., & Avrutin, V. (2013). Bifurcation structure in bimodal piecewise linear map. *International Journal of Bifurcation and Chaos*, *23*(12), doi:10.1142/S0218127413300401.

Sargent, T. J., & Wallace, W. (1973). The stability of models of money and growth with perfect foresight. *Econometrica*, *41*, 1043–1048.

Sushko, I., & Gardini, L. (2010). Degenerate bifurcations and border collisions in piecewise smooth 1D and 2D maps. *International Journal of Bifurcation and Chaos*, *20*, 2045–2070.

Takens, F. (1987). Transitions from periodic to strange attractors in constrained equations. In M. I. Camacho, M. J. Pacifico, & F. Takens (Eds.), *Dynamical systems and bifurcation theory* (pp. 399–421). Longman Scientific and Techical.

Boundedly Rational Monopoly with Single Continuously Distributed Time Delay

Akio Matsumoto and Ferenc Szidarovszky

JEL Classification C62 · C63 · D21 · D42

1 Introduction

In most traditional economic models, instantaneous and complete information has been assumed. Recent research on the dynamic behaviour of the economic agents, however, emphasises on their bounded rationality that arises when the agents have only limited information in making their decisions. In this paper, we build a dynamic monopoly model that accounts for such bounded rationality including partial and delayed knowledge on the price function. Its main purpose is to improve the monopoly theory by getting closer to the real world in which there are always uncertainty and delays in obtaining information and implementing decisions. To this end, we get rid of the questionable and unrealistic assumptions of the rational monopoly. Examining analytically and numerically the delay effects upon local and global dynamic behaviour of the boundedly rational monopoly, we show how cyclic dynamics can emerge from quite simple economic structures. Our analysis makes a sharp difference in the way the rational monopoly behaves. The boundedly rational monopoly becomes *dynamic* in nature because it gropes for its optimal choice by using data obtained in market experience. On the other hand, the rational monopoly is *static* in nature

A. Matsumoto (✉)
Department of Economics, Chuo University, 742-1, Higashi-Nakano,
Hachioji, Tokyo 192-0393, Japan
e-mail: akiom@tamacc.chuo-u.ac.jp

F. Szidarovszky
Department of Applied Mathematics, University of Pécs,
Ifjúság, u.6, Pécs H-7624, Hungary
e-mail: szidarka@gmail.com

R. Dieci et al. (eds.), *Nonlinear Economic Dynamics and Financial Modelling*,
DOI: 10.1007/978-3-319-07470-2_7,

because it can choose its optimal choices of price and quantity to maximise profit with one shot.

In the existing literature two types of delays are usually examined: fixed time delay and continuously distributed time delay (fixed delay and continuous delay henceforth). The former is applicable in economic situations in which an institutionally or socially determined fixed period of time delay is present for the agents involved. The latter is appropriate for economic situations in which different lengths of delays are distributed over the different agents. Uncertain delays can be modelled by continuous delays, and the same types of models describe the situation when firms want to react to the average past information instead of following a sudden market change. In this way firms might avoid fluctuating output trajectories and therefore the related additional cost. So the choice of the type of delays has situation-dependency and results in the use of different analytical tools. In the cases of fixed delays, the dynamic equations are delay differential equations where the characteristic equation is a mixed polynomial-exponential equation with infinitely many eigenvalues.[1] The classical book by Bellman and Cooke (1956) offers the theory of such dynamic models. Kuang (1993) gives good theoretical foundation and comprehensive summary of applications in population dynamics. In economic dynamics, Howroyd and Russel (1984) construct two linear continuous time dynamic oligopoly models and examine the effect of the delay on stability. Fixed delay dynamics has been investigated in various economic frameworks ranging from microeconomics (i.e. oligopoly dynamics) to macroeconomics (i.e. business cycle). In the case of continuous delays the dynamic equations are Volterra type integro-differential equations. Cushing (1977) discusses the mathematical methodology dealing with such dynamics. Invernizzi and Medio (1991) have introduced continuous delays into mathematical economics, and this methodology is later used to examine dynamic oligopolies by Chiarella and Khomin (1996) and Chiarella and Szidarovszky (2001, 2004). Recently, Matsumoto (2009) re-examined the classical Goodwin's accelerator business cycle by replacing fixed delay in the original model with continuous delay. Dynamics generated by fixed delay and continuous delay are compared in Matsumoto and Szidarovszky (2010) in which the Goodwin (2D) model, the Kaldor-Kalecki (3D) model and the Cournot oligopoly (4D) model are examined.

Puu (1995) and, more recently, Naimzada and Ricchiuti (2008) conduct the dynamic analysis of the boundedly rational monopoly with discrete timescale. Adopting a gradient rule in which production is increased if a change in profit is positive, decreased if negative and constant if zero, we show numerically that, under different forms of the demand function, stability of the monopoly equilibrium can be violated to chaos through the familiar period-doubling cascade. In our earlier paper, Matsumoto and Szidarovszky (2012a), the monopoly dynamics is examined in continuous-timescale with one and two fixed delays. A complete stability analysis is given and it is demonstrated that in the case of locally unstable monopoly equilibrium only simple dynamics (i.e., limit cycle) can be born when one fixed delay is

[1] A dynamic equation with fixed delays can be called a mixed difference-differential equation. However, Gandolfo (2009) points out that such terminology is somewhat dated.

involved while complex dynamics are reached through a period-doubling bifurcation when two fixed delays are involved. In this paper, the fixed delay is replaced with a continuous delay and in addition to complete stability analysis, the asymptotic behaviour of the equilibrium with fixed and continuous delays will be compared.

This paper is organised as follows. In the following section, we construct a gradient dynamic model of boundedly rational monopoly. Then in Sect. 3, we analytically examine local dynamics and numerically show that the continuous delay has a threshold value at which the monopoly equilibrium loses stability. In Sect. 4 we introduce an adaptive expectation formation and demonstrate that stability switch occurs twice, one switch to instability from stability for a small delay and the other switch to stability from instability for a large delay. Concluding remarks are given in Sect. 5.

2 Delay Monopoly

In this section, we construct a basic dynamic model of a boundedly rational monopoly which produces output q with marginal cost c. The price function is linear

$$f(q) = a - bq,\ a,\ b > 0.$$

When the monopoly has only limited information on the price function, there are several ways to deal with the behaviour under such circumstances. If it believes in a misspecified price function and chooses its decision accordingly, then a self-confirming stationary state may emerge which is different from the stationary state with full information. Or if it does not know certain parameters of the price function, although knowing that it is linear, then the monopoly uses a local linear approximation of the price function based on its past output data to update its estimate.[2] In this study, assuming that the monopoly does not want to react to sudden market changes, then instead of the most current marginal profit information, an average of past marginal profits is used in the adjustment process. Because of the linearity of the price function, it is equivalent to the use of an average of past output data q^a. Then the corresponding marginal profit is given as

$$\frac{\mathrm{d}\pi(q^a)}{\mathrm{d}q} = a - c - 2bq^a.$$

So the approximating gradient dynamics is

$$\frac{\dot{q}(t)}{q(t)} = \alpha \frac{\mathrm{d}\pi(q^a(t))}{\mathrm{d}q(t)} \tag{1}$$

[2] See Chap. 5 of Bischi et al. (2010) for stability/instability of economic models with misspecified and uncertain price functions.

with α being an adjustment coefficient. In (1), t denotes a point of continuous time and the dot over a variable means a time derivative. This adjustment equation implies that the growth rate of output is adjusted in proportion to the average past marginal profit. In adjustments towards best responses, global information is required about the price function, however, in applying gradient dynamics, only local information is needed. The gradient dynamic equation (1) is rewritten as

$$\dot{q}(t) = \alpha q(t)\left[a - c - 2bq^a(t)\right]. \tag{2}$$

Since $q(t) = q^a(t)$ for all t holds at a stationary point, Eq. (2) has two stationary points; a trivial point $q(t) = 0$ for all t and a non-trivial point

$$q^M = \frac{a-c}{2b}$$

where $a > c$ is assumed to ensure that the non-trivial point is positive. We call q^M a *monopoly equilibrium*. Dynamic behaviour of (2) depends on the formation of the average output. In a dynamic model with continuous timescales, time delays can be modelled with fixed delays or continuously delays. As mentioned in the Introduction, Matsumoto and Szidarovszky (2012a) examined dynamic monopoly with fixed delays. In this study, we adopt a single continuous delay and consider its delay effects on the dynamics.[3] Before proceeding, we briefly summarise the results obtained in the dynamic monopoly with one fixed delay.

We assume $q^a(t) = q(t-\tau)$ where $\tau > 0$ denotes a fixed delay. That is, $a-c-2bq(t-\tau)$ is the delayed information on the marginal profit. In many instances the firm does not have instantaneous price and profit information, and implementing output decisions usually needs time. So there is a time-gap between the time when information is obtained and the time when decision is actually implemented. Substituting $q(t-\tau)$ in equation (2) for $q^a(t)$ gives a nonlinear delay differential equation,

$$\dot{q}(t) = \alpha q(t)\left[a - c - 2bq(t-\tau)\right]. \tag{3}$$

Linearising equation (3) and introducing the new variable, $x(t) = q(t) - q^M$ yield the following linearised form:

$$\dot{x}(t) = -\gamma x(t-\tau) \text{ with } \gamma = \alpha(a-c) > 0.$$

Substituting the exponential solutions $x(t) = x_0 e^{\lambda t}$ into the linearised equation gives the characteristic equation

$$\lambda + \gamma e^{-\lambda\tau} = 0. \tag{4}$$

The sufficient condition for local asymptotic stability is that the real parts of the eigenvalues are negative. It can be shown that the monopoly equilibrium is locally

[3] Monopoly dynamics with multiple continuous delays is considered in Matsumoto and Szidarovszky (2012b).

asymptotically stable for $0 < \tau < \tau^*$, locally unstable for $\tau > \tau^*$ and undergoes a Hopf bifurcation at $\tau = \tau^*$ where a threshold value τ^* of delay is defined as

$$\tau^* = \frac{\pi}{2\alpha(a-c)}. \tag{5}$$

This curve, which is downward sloping with respect to $\gamma = \alpha(a-c)$, divides the parameter space into stable and unstable regions. We call it the *partition curve*. The monopoly equilibrium is locally stable below the partition curve, locally asymptotically unstable above and bifurcates to a limit cycle when it crosses the curve.

3 Dynamics with Continuous Delay

As mentioned earlier, continuous delay is an alternative approach to deal with delays. Continuously distributed delays are the appropriate approach if the delay is uncertain, or different lengths of delays are distributed over the different agents. A similar situation occurs when the firm wants to react to average past information instead of following sudden market changes which would lead to fluctuating output trajectories. The gradient dynamics with a single continuous delay is given by the following two equations:

$$\begin{aligned} \dot{q}(t) &= \alpha q(t)\left[a - c - 2bq^{\varepsilon}(t)\right], \\ q^{\varepsilon}(t) &= \int_0^t W(t-s,\tau,m)q(s)\mathrm{d}s, \end{aligned} \tag{6}$$

where the weighting function is defined as

$$W(t-s,\tau,m) = \begin{cases} \dfrac{1}{\tau}e^{-\frac{t-s}{\tau}} & \text{if } m = 0, \\ \dfrac{1}{m!}\left(\dfrac{m}{\tau}\right)^{m+1}(t-s)^m e^{-\frac{m(t-s)}{\tau}} & \text{if } m \geq 1, \end{cases} \tag{7}$$

in which m is a non-negative integer and τ is a positive real parameter, which is associated with the length of the delay. The first equation of (6) implies that the growth rate of output is proportional to the average past marginal profits. The second equation indicates that the average output at time t is the weighted average of the actual demand in the past. According to (7), the shape of the weighting function is determined by the value of the shape parameter, m. For $m = 0$, weights are exponentially declining with the most weight given to the most current data. For $m \geq 1$, zero weight is given to the most current data, rising to maximum at $s = t - \tau$ and declining exponentially thereafter. The weights take a bell-shaped form which becomes taller and thinner as m increases.

To investigate local dynamics of this system in a neighbourhood of the equilibrium point, we construct a linearised version. If the deviations of the actual and expected outputs from the equilibrium value are denoted by $x_\delta(t) = q(t) - q^M$ and $x_\delta^\varepsilon(t) = q^\varepsilon(t) - q^M$, then the linearised system with continuous delay can be written as

$$\begin{aligned} \dot{x}_\delta(t) &= -\gamma x_\delta^\varepsilon(t), \\ x_\delta^\varepsilon(t) &= \int_0^t W(t-s,\tau,m)x_\delta(s)\mathrm{d}s, \end{aligned} \tag{8}$$

where $\gamma = 2\alpha b q^M$. Substituting the second equation of (8) into the first yields the following Volterra-type integro-differential equation:

$$\dot{x}_\delta(t) + \gamma \int_0^t W(t-s,\tau,m)x_\delta(s)\mathrm{d}s = 0.$$

Looking for the solution in the usual exponential form $x_\delta(t) = x_0 e^{\lambda t}$ and substituting it into the above equation, we obtain

$$\lambda + \gamma \int_0^t W(t-s,\tau,m)e^{-\lambda(t-s)}\mathrm{d}s = 0.$$

Introducing the new variable $z = t - s$ simplifies the integral as

$$\int_0^t W(t-s,\tau,m)e^{-\lambda(t-s)}\mathrm{d}s = \int_0^t W(z,\tau,m)e^{-\lambda z}\mathrm{d}z.$$

By letting $t \to \infty$ and assuming that $\mathrm{Re}(\lambda) + \frac{m}{\tau} > 0$, we have

$$\int_0^\infty \frac{1}{\tau} e^{-\frac{z}{\tau}} e^{-\lambda z}\mathrm{d}z = (1+\lambda\tau)^{-1} \quad \text{if } m = 0$$

and

$$\int_0^\infty \frac{1}{m!}\left(\frac{m}{\tau}\right)^{m+1} z^m e^{-\frac{mz}{\tau}} e^{-\lambda z}\mathrm{d}z = \left(1+\frac{\lambda\tau}{m}\right)^{-(m+1)} \quad \text{if } m \geq 1.$$

That is,

$$\int_0^\infty W(z,\tau,m)e^{-\lambda z}\mathrm{d}z = \left(1+\frac{\lambda\tau}{\bar{m}}\right)^{-(m+1)}$$

with

$$\bar{m} = \begin{cases} 1 & \text{if } m = 0, \\ m & \text{if } m \geq 1. \end{cases}$$

Then the characteristic equation becomes

$$\lambda\left(1+\frac{\tau}{\bar{m}}\lambda\right)^{m+1} + \gamma = 0. \tag{9}$$

Expanding the characteristic equation presents the $(m+2)$-th order polynomial equation

$$a_0\lambda^{m+2} + a_1\lambda^{m+1} + \cdots + a_{m+1}\lambda + a_{m+2} = 0$$

where the coefficients a_k are given as

$$a_k = \left(\frac{\tau}{\bar{m}}\right)^{m+1-k}\binom{m+1}{k} \quad \text{for } 0 \leq k \leq m+1$$

and

$$a_{m+2} = \gamma.$$

In the case of the high-order polynomial equation, the Routh-Hurwitz theorem[4] provides the necessary and sufficient conditions for all the roots to have negative real parts. In order to apply the theorem, we first need to construct the Routh-Hurwitz determinant:

$$D_{m+2} = \det\begin{pmatrix} a_1 & a_0 & 0 & 0 & \cdots & 0 \\ a_3 & a_2 & a_1 & a_0 & \cdots & 0 \\ a_5 & a_4 & a_3 & a_2 & \cdots & 0 \\ a_7 & a_6 & a_5 & a_4 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & a_{m+2} \end{pmatrix}.$$

Then the stability conditions are as follows:

(1) all coefficients are positive, $a_k > 0$ for $k = 0, 1, 2, \ldots, m+2$,
(2) the principal minors of the Routh-Hurwitz determinant are all positive,

$$D_{m+2}^2 > 0,\ D_{m+2}^3 > 0, \ldots,\ D_{m+2}^{m+1} > 0$$

[4] See, for example, Gandolfo (2009) for this theorem.

where D^k_{m+2} is the kth order leading principal minor of D_{m+2}. Notice that $D^{m+1}_{m+2} > 0$ always leads to $D^{m+2}_{m+2} > 0$ since $a_{m+2} = \gamma > 0$.

Since it is difficult to obtain a general solution of Eq. (9), we draw attention to special cases with $m = 0, 1, 2, 3$ and $m \to \infty$ and examine stability of the monopoly equilibrium analytically as well as numerically.

Case I-0: $m = 0$
Substituting $m = 0$ reveals that the characteristic equation (9) is quadratic, $\tau\lambda^2 + \lambda + \gamma = 0$ where all coefficients are positive. Its real roots are negative and the real parts of the complex roots are also negative. Thus the monopoly equilibrium is locally asymptotically stable for all $\tau > 0$. Since the delay does not affect asymptotic behaviour of the monopoly equilibrium, such a delay is called *harmless*.

Case I-1: $m = 1$
The characteristic equation (9) with $m = 1$ becomes cubic and its coefficients are all positive

$$a_0 = \tau^2 > 0, \quad a_1 = 2\tau > 0, \quad a_2 = 1 > 0, \quad a_3 = \gamma > 0.$$

According to the Routh-Hurwitz criterion, the following leading minor of D_3 needs to be positive for preserving stability of the equilibrium:

$$D_3^2 = \begin{vmatrix} a_1 & a_0 \\ a_3 & a_2 \end{vmatrix}.$$

To obtain $D_3^2 = \tau(2 - \tau\gamma) > 0$, the delay τ should be less than the threshold value

$$\tau_1^* = \frac{2}{\gamma}. \tag{10}$$

There is a possibility of the emergence of a limit cycle when loss of stability occurs at $\tau = \tau_1^*$. The Hopf bifurcation theorem comes in to provide the sufficient conditions for it:

(H1) the characteristic equation of the dynamic system has a pair of pure imaginary roots and has no other roots with zero real parts;
(H2) the sign of the real parts of these roots vary with a bifurcation parameter.

The $D_3^2 = 0$ curve divides the parameter space into stable and unstable parts. Substituting $a_3 = a_1 a_2 / a_0$ into the characteristic equation gives the factored form

$$(a_1 + a_0\lambda)(a_2 + a_0\lambda^2) = 0.$$

We have therefore three characteristic roots, two purely imaginary roots and one real and negative root,

$$\lambda_{1,2} = \pm\sqrt{-\frac{a_2}{a_0}} = \pm i\frac{1}{\tau} \text{ and } \lambda_3 = -\frac{a_1}{a_0} = -\frac{2}{\tau} < 0.$$

The first condition (H1) of the Hopf theorem is satisfied.

Next, we select the delay τ as the bifurcation parameter and consider the roots of the characteristic equation as continuous functions of τ:

$$\tau^2\lambda(\tau)^3 + 2\tau\lambda(\tau)^2 + \lambda(\tau) + \gamma = 0$$

Differentiating it with respect to τ gives

$$\frac{\mathrm{d}\lambda}{\mathrm{d}\tau} = -\frac{2\tau\lambda^3 + 2\lambda^2}{3\tau^2\lambda^2 + 4\tau\lambda + 1}.$$

Substituting $\lambda = i/\tau$, rationalising the right-hand side and noticing that the terms with λ and λ^3 are imaginary while the constant and the term λ^2 are real yield the following form of the real part of the derivative of λ with respect to τ :

$$\mathrm{Re}\left[\left.\frac{\mathrm{d}\lambda}{\mathrm{d}\tau}\right|_{\lambda=\frac{i}{\tau}}\right] = \frac{1}{5\tau^2} > 0.$$

The last inequality indicates that the second condition (H2) is also satisfied. The real parts of the complex roots change to positive from negative value resulting in the loss of stability on the partition curve. Hence, the Hopf bifurcation theorem confirms the birth of a limit cycle when stability is lost.

We numerically examine the analytical result just obtained. The dynamic system under investigation is obtained by substituting $m = 1$ into (6),

$$\dot{q}(t) = \alpha q(t)\left[a - c - 2bq^{\varepsilon}(t)\right]$$

$$q^{\varepsilon}(t) = \int_0^t \left(\frac{1}{\tau}\right)^2 (t-s)e^{-\frac{t-s}{\tau}} q(s)ds.$$

Differentiating the second equation with respect to t and introducing a new variable

$$q_0(t) = \int_0^t \frac{1}{\tau} e^{-\frac{t-s}{\tau}} q(s)ds$$

transforms the dynamic system with continuous delay into a 3D system of ordinary differential equations,

$$\begin{aligned}
\dot{q}(t) &= \alpha q(t)\left[a - c - 2bq^{\varepsilon}(t)\right], \\
\dot{q}^{\varepsilon}(t) &= \frac{1}{\tau}\left(q_0(t) - q^{\varepsilon}(t)\right), \\
\dot{q}_0(t) &= \frac{1}{\tau}\left(q(t) - q_0(t)\right).
\end{aligned}$$

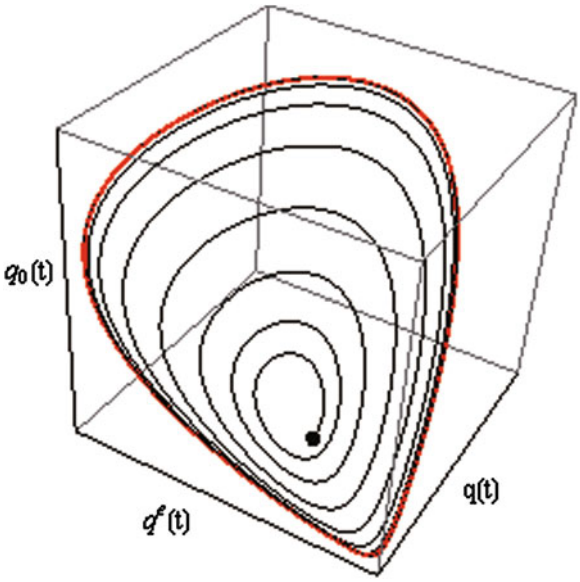

Fig. 1 Convergence to a limit cycle in the (q, q^e, q_0) space

We exploit the following parameter setting, $a = 2,\ b = 1,\ c = 1,\ \alpha = 1$,[5] and take $\tau = 3$ and the initial values of all variables to be $q^M - 0.1$. Then simulating the 3D system exhibits the birth of a limit cycle as shown in Fig. 1 where a black trajectory starting at the black dot (i.e., positive initial point) converges to a red cycle in the $(q,\ q^{\varepsilon},\ q_0)$ space. We then summarise this result as follows: the monopoly equilibrium point with $m = 1$ is destabilised through a Hopf bifurcation and, as it is numerically confirmed, converges to a limit cycle when the delay τ is larger than the critical value $2/\gamma$.

Case I-2: $m = 2$
The characteristic equation (9) with $m = 2$ becomes quartic

$$a_0\lambda^4 + a_1\lambda^3 + a_2\lambda^2 + a_3\lambda + a_4 = 0$$

and its coefficients are all positive,

$$a_0 = \tau^3, \quad a_1 = 6\tau^2, \quad a_2 = 12\tau, \quad a_3 = 8, \quad a_4 = 8\gamma.$$

The principal minors of the Routh-Hurwitz determinant are $D_4^2 = 64\tau^3 > 0$ and $D_4^3 = 32\tau^3(16 - 9\gamma\tau)$. Then to obtain $D_3^2 > 0$, the delay τ should be less than the threshold value

[5] This set of parameters is repeatedly used in the following numerical examples. Notice that $\gamma = 1$ under this set.

$$\tau_2^* = \frac{16}{9\gamma} \simeq \frac{1.78}{\gamma}. \tag{11}$$

Thus, the equilibrium is locally asymptotically stable if $\tau < \tau_2^*$ and locally unstable if $\tau > \tau_2^*$. In the same way as in Case I-1, we can show the existence of a limit cycle at the critical value τ_2^*. In particular, solving $D_4^3 = 0$ or $a_1a_2a_3 - (a_0a_3^2 + a_1^2a_4) = 0$ for a_4, substituting it into the characteristic equation and factoring the resultant equation yield

$$(a_3 + a_1\lambda^2)(a_1a_2 - a_0a_3 + a_1^2\lambda + a_0a_1\lambda^2) = 0.$$

The solutions of $a_3 + a_1\lambda^2 = 0$ are purely imaginary,

$$\lambda_{1,2} = \pm i\frac{2}{\tau\sqrt{3}},$$

and the other two characteristic roots are the solutions of the quadratic equation $(a_1a_2 - a_0a_3) + a_1^2\lambda + a_0a_1\lambda^2 = 0$,

$$\lambda_{3,4} = \frac{-9 \pm i\sqrt{15}}{3\tau}$$

whose real parts are negative. The first condition (H1) of the Hopf bifurcation theorem is satisfied.

To confirm the second condition, we choose τ as the bifurcation parameter again and differentiate the characteristic equation with respect to τ to have

$$\frac{\mathrm{d}\lambda}{\mathrm{d}\tau} = -\frac{3\tau^2\lambda^4 + 12\tau\lambda^3 + 12\lambda^2}{4\tau^3\lambda^3 + 18\tau^2\lambda^2 + 24\tau\lambda + 8}.$$

Substituting $\lambda = i\frac{2}{\tau\sqrt{3}}$ and taking the real part, we have

$$\mathrm{Re}\left[\left.\frac{\mathrm{d}\lambda}{\mathrm{d}\tau}\right|_{\lambda=i\frac{2}{\tau\sqrt{3}}}\right] = \frac{6}{19\tau^2} \simeq \frac{0.316}{\tau^2} > 0.$$

We thereby confirm that the monopoly equilibrium with $m = 2$ is destabilised through a Hopf bifurcation when the delay τ crosses the critical value τ_2^*.

Case I-3: $m = 3$
The characteristic equation (9) is quintic and its coefficients are all positive,

$$a_0 = \tau^4,\ a_1 = 12\tau^3,\ a_2 = 54\tau^2,\ a_3 = 108\tau,\ a_4 = 81,\ a_5 = 81\gamma.$$

The first two principal minors of the Routh-Hurwitz determinant are positive,

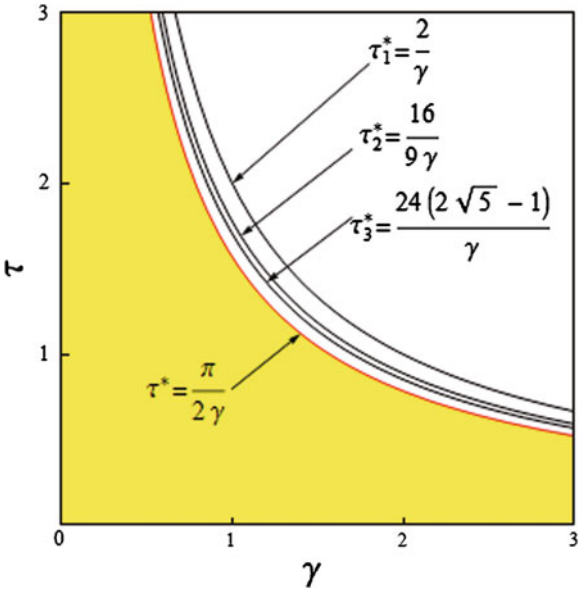

Fig. 2 Three partition curves and stability region under the fixed delay

$$D_5^2 = 540\tau^5 > 0 \quad \text{and} \quad D_5^3 = 972\tau^6(48 + \gamma\tau) > 0.$$

The sign of the fourth order principal minor $D_5^4 = -6561\tau^6(\gamma^2\tau^2 + 336\gamma\tau - 576)$ is ambiguous. To obtain $D_5^4 > 0$, the delay τ should be less than the threshold value

$$\tau_3^* = \frac{24(5\sqrt{2}-7)}{\gamma} \simeq \frac{1.71}{\gamma}. \tag{12}$$

Although we omit the detail, we can also show that the continuous delay system (6) with $m = 3$ can generate a limit cycle through a Hopf bifurcation in the same way as in the previous cases when the monopoly equilibrium loses stability.[6]

The relations (5), (10), (11), and (12) define the partition curves of (γ, τ) that divide the (γ, τ) space into stable and unstable parts. The three partition curves for $m = 1, 2, 3$ and the stability (yellow) region with fixed delay defined by $\tau\gamma < \pi/2$ are depicted in Fig. 2. It can be seen that all curves are hyperbolic and are approaching the red-coloured boundary of the stability region from above. In other words, the stable region with continuous delay becomes smaller as the value of m increases and converges to the region defined by the fixed delay when m tends to infinity. The second result is natural if we notice the properties of the weighting function. The weighting function for $m \geq 1$ is a bell-shaped and becomes more peaked around $t - s$ as m increases. Furthermore, it tends to the Dirac delta function if $m \to \infty$. In

[6] With tedious calculations, it may be possible to show the similar results in cases for $m \geq 4$.

consequence, for sufficiently large m, the weighting function may be regarded as very close to the Dirac delta function and the dynamic behaviour under the continuous delay is similar to that under the fixed delay. We can explain this phenomenon mathematically by noticing that the characteristic equation (9) of the continuously distributed case can be written as

$$\lambda + \gamma \left(1 + \frac{\tau\lambda}{m}\right)^{-m} \left(1 + \frac{\tau\lambda}{m}\right)^{-1} = 0,$$

and as $m \to \infty$, the left-hand side converges to

$$\lambda + \gamma e^{-\lambda\tau} = 0.$$

This is the characteristic equation of the delay differential equation with a single fixed delay and is identical to Eq. (4). In short, under continuous delay, although we comprehensively use all delayed or past output data, the stability domain is sensitive to the shape of the weighting function. Hence, we obtain the following two results:

Proposition 3.1 *(1) The monopoly equilibrium of the continuously distributed single delay model is always stable for any delays if $m = 0$ and is destabilised through a Hopf bifurcation if $m = 1, 2, 3$; (2) The stability region decreases as m increases (i.e., $\tau^* < \tau_3^* < \tau_2^* < \tau_1^*$) and converges to the stability region obtained under the fixed delay when m goes to infinity (i.e. $\tau_m^* \to \tau^*$ as $m \to \infty$).*

4 Delay Dynamics with Adaptive Expectation

In this section, we explore the effects caused by a different expectation formation on monopoly dynamics. For this reason, we use adaptive expectation formation defined as

$$\begin{aligned} q^{\varepsilon}(t) &= \omega q^{e}(t) + (1-\omega) q(t) \\ q^{e}(t) &= \int_0^t W(t-s, \tau, m) q(s) \mathrm{d}s \end{aligned} \tag{13}$$

with $0 < \omega \leq 1$. The expectation is formed with two steps: the weighted average of the past data is calculated at the first step and then the expected demand is selected at the second step, somewhere in between the current output level and the weighted average level.[7] If $m \geq 1$, then zero weight is given to the most current data in the

[7] Since the first equation of (13) can be rewritten as

$$q^{\varepsilon}(t) - q(t) = \omega(q^{e}(t) - q(t)),$$

it can be mentioned that the expected demand is formed in such a way that the expectation error is proportional to the difference between the weighted average level and the current level.

weighting function, and so the second term in the first equation of (13) gives a larger weight to it taking a certain learning procedure based on past data at the second stage. We examine in some detail the dynamic effects caused by a single delay with the adaptive expectation formation. Following the method we take in the previous section, the characteristic equation of the system (13) can be obtained as

$$\lambda\left(1+\frac{\lambda\tau}{\bar{m}}\right)^{m+1}+\gamma\left[\omega+(1-\omega)\left(1+\frac{\lambda\tau}{\bar{m}}\right)^{m+1}\right]=0 \tag{14}$$

which is reduced to Eq. (9) if $\omega = 1$. To find how the shape of the weighting function, $W(t-s,\tau,m)$, affects the dynamics of q^M, we sequentially increase the value of m from zero to five and then to infinity.

Case II-0: $m = 0$
Substituting $m = 0$ in Eq. (14) presents the form

$$\lambda(1+\lambda\tau)+\gamma\omega+\gamma(1-\omega)(1+\lambda\tau)=0$$

or

$$\tau\lambda^2+(1+\tau\gamma(1-\omega))\lambda+\gamma=0.$$

Since all coefficients are positive, there is no non-negative root and the real parts of the complex eigenvalues are negative,

$$\mathrm{Re}(\lambda_{\pm})=-\frac{1+(1-\omega)\tau\gamma}{2\tau}<0,$$

implying that the equilibrium is locally asymptotically stable for all $\tau > 0$. As in Case I-0, the continuous delay is again harmless when the weight exponentially declines (i.e. $m = 0$).

For $m \geq 1$, expanding the characteristic equation (14) yields the polynomial equation of degree $m + 2$

$$b_0\lambda^{m+2}+b_1\lambda^{m+1}+\cdots+b_m\lambda^2+b_{m+1}\lambda+b_{m+2}=0 \tag{15}$$

where the coefficients are defined as

$$b_0 = a_0,$$

$$b_k = a_k+\gamma(1-\omega)a_{k-1} \quad \text{for } 1\leq k\leq m,$$

$$b_{m+1} = a_{m+1}+\gamma(1-\omega)a_m,$$

$$b_{m+2} = \gamma$$

with

$$a_k = \left(\frac{\tau}{m}\right)^{m+1-k} \binom{m+1}{k} \quad \text{for } 0 \le k \le m+1. \tag{16}$$

Case II-1: $m = 1$
Equation (14) with $m = 1$ becomes cubic:

$$b_0\lambda^3 + b_1\lambda^2 + b_2\lambda + b_3 = 0 \tag{17}$$

where the coefficients are defined as

$$b_0 = \tau^2,\ b_1 = 2\tau + \gamma(1-\omega)\tau^2,\ b_2 = 1 + 2\tau\gamma(1-\omega) \text{ and } b_3 = \gamma.$$

All coefficients are positive, so the Routh-Hurwitz criterion implies that the monopoly equilibrium is locally asymptotically stable if

$$D_3^2 = \det \begin{pmatrix} b_1 & b_0 \\ b_3 & b_2 \end{pmatrix} > 0 \tag{18}$$

where the determinant is written as

$$\tau\left(2\gamma^2(1-\omega)^2\tau^2 + (4\gamma - 5\gamma\omega)\tau + 2\right).$$

Since $\tau > 0$, this expression is positive if and only if

$$f(\tau\gamma) = 2(1-\omega)^2(\tau\gamma)^2 + (4-5\omega)\tau\gamma + 2 > 0. \tag{19}$$

$f(\tau\gamma)$ is quadratic with respect to $\tau\gamma$ and its discriminant has the form

$$D = \omega(9\omega - 8).$$

If $\omega < 8/9$, then $D < 0$ so (19) always holds. If $\omega = 8/9$, then the right-hand side of (19) simplifies as

$$f(\tau\gamma) = 2\left(\frac{\tau\gamma}{9} - 1\right)^2$$

so (19) holds if $\omega = 8/9$ and $\tau\gamma \neq 9$. If $\omega > 8/9$, then $D > 0$ and $f(\tau\gamma)$ has two distinct real roots:

$$\tau\gamma_A = \frac{5\omega - 4 - \sqrt{\omega(9\omega-8)}}{4(1-\omega)^2} \text{ and } \tau\gamma_B = \frac{5\omega - 4 + \sqrt{\omega(9\omega-8)}}{4(1-\omega)^2}.$$

$f(0) > 0$ and $f'(0) < 0$ imply that both roots are positive. So the monopoly equilibrium is locally asymptotically stable if $\omega > 8/9$ and $\tau\gamma < \tau\gamma_A$ or $\tau\gamma > \tau\gamma_B$.

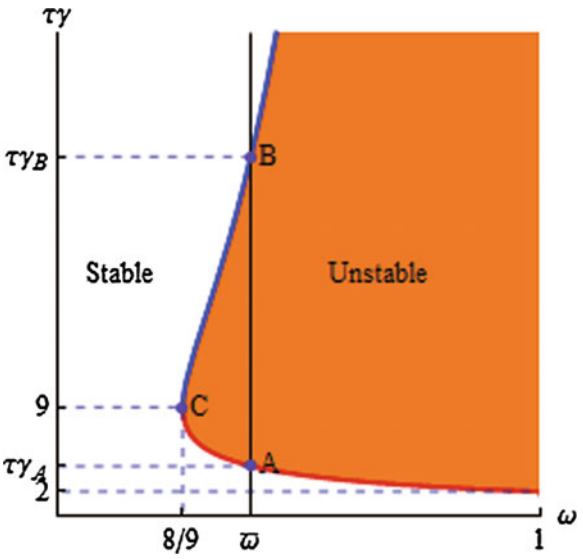

Fig. 3 Stability region with $m = 1$

These analytical results are graphically visualised in Fig. 3. The locus of $b_1b_2 - b_3b_0 = 0$ is depicted as the blue–red curve passing through points A, C and B and partitions the non-negative $(\omega,\ \tau\gamma)$ plane into a stable (white) region and an unstable (orange) region.[8] The monopoly equilibrium is locally asymptotically stable regardless of the value of $\tau\gamma$ if $\omega < 8/9$ or if $\omega = 8/9$ and $\tau\gamma \neq 9$. Stability-switch occurs twice if $8/9 < \omega < 1$. The vertical real line at $\omega = \bar{\omega} > 8/9$ crosses the partition curve at two points A and B whose ordinates are $\tau\gamma_A$ and $\tau\gamma_B$, respectively. The monopoly equilibrium loses stability at point A and regains stability at point B when $\tau\gamma$ increases from zero along the vertical line at $\omega = \bar{\omega}$. For $\omega = 1$, the stability condition (19) is reduced to

$$-\tau\gamma + 2 > 0.$$

The monopoly equilibrium is locally stable if $\tau\gamma < 2$ and locally unstable if $\tau\gamma > 2$, implying that the stability switch occurs only once at $\tau\gamma = 2$ for $\omega = 1$. This is the same as the result obtained in Case I-1 which is identical to Case II-1 when $\omega = 1$.

The local behaviour of the equilibrium at points A and B has been already examined. However, global dynamic behaviour of the locally unstable equilibrium between points A and B is still in question. To investigate such behaviour, we will show that at these critical values Hopf bifurcation occurs giving the possibility of the birth of limit cycles even under the adaptive expectation formation.

[8] Notice that $f(\tau\gamma) = 0$ generates equal roots at point $C = (8/9, 9)$.

We start with the first condition, (H1). The cubic characteristic equation (17) can be factored when $b_1 b_2 - b_0 b_3 = 0$,

$$(b_1 + b_0\lambda)(b_2 + b_0\lambda^2) = 0.$$

This factorisation implies that there are a pair of purely imaginary roots and one negative root. The two purely imaginary roots are given as

$$\lambda_{1,2} = \pm\sqrt{-\frac{b_2}{b_0}} = \pm i\beta$$

with

$$\beta = \frac{\sqrt{1 + 2\tau\gamma(1-w)}}{\tau}$$

and the negative root by

$$\lambda_3 = -\frac{b_1}{b_0} = -\frac{2 + (1-w)\tau\gamma}{\tau} < 0.$$

The fulfilment of (H1) is confirmed.

In turn, we verify (H2). Selecting τ as the bifurcation parameter we might treat the eigenvalue as a continuous function of τ, $\lambda = \lambda(\tau)$. Differentiating the characteristic equation (17) with $\lambda(\tau)$ implicitly with respect to τ and arranging terms, we have

$$\frac{\mathrm{d}\lambda}{\mathrm{d}\tau} = -\frac{2\lambda^3\tau + \lambda^2(2 + 2\tau\gamma(1-\omega)) + 2\lambda\gamma(1-\omega)}{3\lambda^2\tau^2 + 2\lambda(2\tau + \gamma(1-\omega)\tau^2) + (1 + 2\tau\gamma(1-\omega))}. \tag{20}$$

At $\lambda = i\beta$,

$$\frac{\mathrm{d}\lambda}{\mathrm{d}\tau} = \frac{(2\tau\beta^3 - 2\beta\gamma(1-\omega))i + \beta^2(2 + 2\tau\gamma(1-\omega))}{-2\beta^2\tau^2 + 2\beta i(2\tau + \gamma(1-\omega)\tau^2)},$$

where the relation $(\beta\tau)^2 = 1 + 2\tau\gamma(1-\omega)$ is used to simplify the denominator of the above equation. Then the real part becomes

$$\mathrm{Re}\left[\left.\frac{\mathrm{d}\lambda}{\mathrm{d}\tau}\right|_{\lambda=i\beta}\right] = \frac{1 - \gamma^2\tau^2(1-\omega)^2}{\beta^2\tau^4 + (2\tau + \gamma(1-\omega)\tau^2)^2},$$

with the positive denominator. It is easy to see that

$$\tau\gamma_A < \frac{1}{1-\omega} < \tau\gamma_B,$$

so

$$\mathrm{Re}\left[\left.\frac{\mathrm{d}\lambda}{\mathrm{d}\tau}\right|_{\lambda=i\beta}\right] > 0 \text{ at point } A \text{ and } \mathrm{Re}\left[\left.\frac{\mathrm{d}\lambda}{\mathrm{d}\tau}\right|_{\lambda=i\beta}\right] < 0 \text{ at point } B.$$

So at point A, the real part changes sign from negative to positive and at point B, from positive to negative. This demonstrates that (H2) of the Hopf theorem is satisfied. So Hopf bifurcation occurs at both points.

We next numerically examine the switching of stability and global behavior. The dynamic system under the investigation (i.e. $m = 1$) is obtained as

$$\begin{aligned} \dot{q}(t) &= \alpha q(t)\left[a - c - 2b\left(\omega q^{\varepsilon}(t) + (1-\omega)q(t)\right)\right] \\ q^{\varepsilon}(t) &= \int_0^t \frac{1}{\tau^2}(t-s)e^{-\frac{t-s}{\tau}} q(s)\mathrm{d}s. \end{aligned} \tag{21}$$

Differentiating the second equation with respect to t and introducing a new variable

$$q_0(t) = \int_0^t \frac{1}{\tau} e^{-\frac{t-s}{\tau}} q(s)\mathrm{d}s$$

transform the dynamic system with continuously distributed time delay into a 3D system of the ordinary differential equations

$$\begin{aligned} \dot{q}(t) &= \alpha q(t)\left[a - c - 2b\left(\omega q^{\varepsilon}(t) + (1-\omega)q(t)\right)\right], \\ \dot{q}^{\varepsilon}(t) &= \frac{1}{\tau}\left(q_0(t) - q^{\varepsilon}(t)\right), \\ \dot{q}_0(t) &= \frac{1}{\tau}\left(q(t) - q_0(t)\right). \end{aligned} \tag{22}$$

We use the same parameter setting (i.e. $a = 2,\ b = 1,\ c = 1, \alpha = 1$) and the same initial values (i.e., $q(0) = q^{\varepsilon}(0) = q_0(t) = q^M - 0.1$) as in Case I-1. Increasing the value of $\tau\gamma$ from $\tau\gamma_A$ to $\tau\gamma_B$ along the vertical line $\bar{\omega} = 0.91$ in Fig. 3, we obtain the bifurcation diagram shown in Fig. 4 where the local maximum and minimum are plotted against each value of $\tau\gamma$. It can be seen that the monopoly equilibrium loses stability bifurcating to a limit cycle when $\tau\gamma$ arrives ar $\tau\gamma_A$. It is further seen that the limit cycle expands, shrinks and then merges with the monopoly equilibrium when $\tau\gamma$ increases from $\tau\gamma_A$ to $\tau\gamma_B$. We then summarise the results obtained in Case II-1 as follows:

Proposition 4.1 *Under the cautious expectation formation, dynamics of the monopoly equilibrium* q^M *with* $\tau > 0$, $m = 1$ *and* $0 < \omega \leq 1$ *takes one of the following alternative behaviour:*

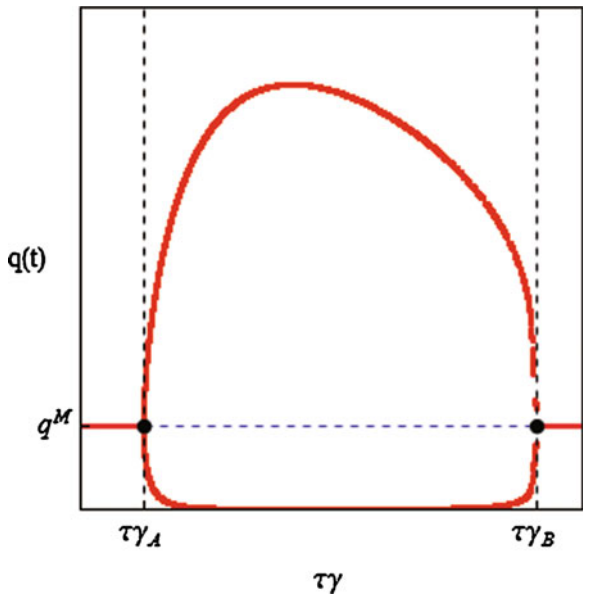

Fig. 4 Bifurcation diagram along the $\omega = 0.91$ line

(I) q^M *is locally asymptotically stable if* $0 < \omega < 8/9$ *regardless of the values of* $\tau\gamma$ *(i.e. the delay is harmless);*

(II) q^M *bifurcates to a limit cycle at* $\tau\gamma_A(\omega)$ *and the limit cycle merges with the monopoly equilibrium at* $\tau\gamma_B(\omega)$ *if* $8/9 < \omega < 1$ *where* $\tau\gamma_A(\omega)$ *and* $\tau\gamma_B(\omega)$ *are the bifurcation values depending on* ω *(i.e., stability switch occurs twice);*

(III) q^M *bifurcates to a limit cycle at* $\tau\gamma = 2$ *and never regains stability for* $\tau\gamma > 2$ *if* $\omega = 1$ *(i.e. stability switch occurs once).*

Case II-2: $m \geq 2$

As in the same way as in Case II-1, we can check the stability condition, the birth of a limit cycle and stability switch in the case of $m \geq 2$. For example, the characteristic equation (14) with $m = 2$ is quartic,

$$b_0\lambda^4 + b_1\lambda^3 + b_2\lambda^2 + b_3\lambda + b_4 = 0$$

and the stability conditions are given by

$$D_4^2 = \det\begin{pmatrix} b_1 & b_0 \\ b_3 & b_2 \end{pmatrix} > 0 \text{ and } D_4^3 = \begin{pmatrix} b_1 & b_0 & 0 \\ b_3 & b_2 & b_1 \\ 0 & b_4 & b_3 \end{pmatrix} > 0$$

where the coefficients are given by

$$b_0 = \frac{\tau^3}{8},\ b_1 = \frac{3\tau^2}{4} + \frac{\tau^3\gamma(1-\omega)}{8},\ b_2 = \frac{3\tau}{2} + \frac{3\tau^2\gamma(1-\omega)}{4},$$

$$b_3 = 1 + \frac{3}{2}\tau\gamma(1-\omega),\ b_4 = \gamma.$$

Since all coefficients are positive and $D_4^3 > 0$ implies $D_4^2 > 0$, it remains to check whether D_4^3 can be positive. Notice that

$$D_4^3 = \alpha_0(\tau\gamma)^3 + \alpha_1(\tau\lambda)^2 + \alpha_2(\tau\gamma) + \alpha_3 \tag{23}$$

where

$$\alpha_0 = (9\omega - 8)(1-\omega)^2, \quad \alpha_1 = -12(4 - 9\omega + 5\omega^2)$$

and

$$\alpha_2 = -12(8 - 11\omega), \quad \alpha_3 = -64.$$

The locus of $D_4^3 = 0$ divides the $(\omega, \tau\gamma)$ plane into two subregions as shown in Fig. 5 where the dark grey region is the unstable region for $m = 1$ and adding the light grey region to it gives the unstable region for $m = 2$. The discriminant of the cubic equation (23) has the form

$$\Delta = 442368(1-\omega)^3\omega^2(5\omega - 4)$$

which is obtained by substituting α_i into the definition of the discriminant,

$$-4\alpha_0\alpha_2^3 + \alpha_1^2\alpha_2^2 - 4\alpha_1^3\alpha_3 + 18\alpha_0\alpha_1\alpha_2\alpha_3 - 27\alpha_0^2\alpha_3^2.$$

If $\omega < 4/5$, then $\Delta < 0$ implying $D_4^3 > 0$ where $D_4^3 = 0$ has a pair of conjugate complex roots and one negative real root. If $\omega = 4/5$, then $\Delta = 0$ implying that $D_4^3 = (10 - \tau\gamma)^2(20 + \tau\gamma) > 0$ for $\tau\gamma \neq 10$ where (23) has equal roots at the red point $(4/5, 10)$ on the $D_4^3 = 0$ locus. If $\omega > 4/5$, then $\Delta > 0$ implying that $D_4^3 = 0$ has three distinct real roots. Since Eq. (23) is cubic, it is possible to derive explicit forms of the real roots. However, to simplify the analysis, we numerically obtain the roots. Taking $\omega_0 = 0.815(>4/5)$ in addition to the parametric setting, we have two positive real roots, $\tau\gamma_A \simeq 5.42$ at point A, $\tau\gamma_B \simeq 23.75$ at point B and one negative root. $D_4^3 \leq 0$ for $\tau\gamma \in [\tau\gamma_A, \tau\gamma_B]$ which is an unstable interval and $D_4^3 > 0$ for $0 \leq \tau\gamma < \tau\gamma_A$ or $\tau\gamma > \tau\gamma_B$. Stability switch occurs twice at $\tau\gamma = \tau\gamma_A$ and $\tau\gamma = \tau\gamma_B$. Further the cubic equation is reduced to a quadratic equation for $\omega = 8/9$ implying that the locus of $D_4^3 = 0$ is defined only for $\omega < 8/9$ when $\tau\gamma > 10$ and asymptotic to the vertical line at $\omega = 8/9$. If $8/9 < \omega < 1$, then the cubic equation (23) has one positive root $\tau\gamma_C$ for $\omega = \omega_1$. It is confirmed that $D_4^3 > 0$ for $\tau\gamma < \tau\gamma_C$ and $D_4^3 \leq 0$ otherwise. The equilibrium point switches to be unstable at point C on the $D_4^3 = 0$ locus when the delay is increased along the vertical line at $\omega = \omega_1$.

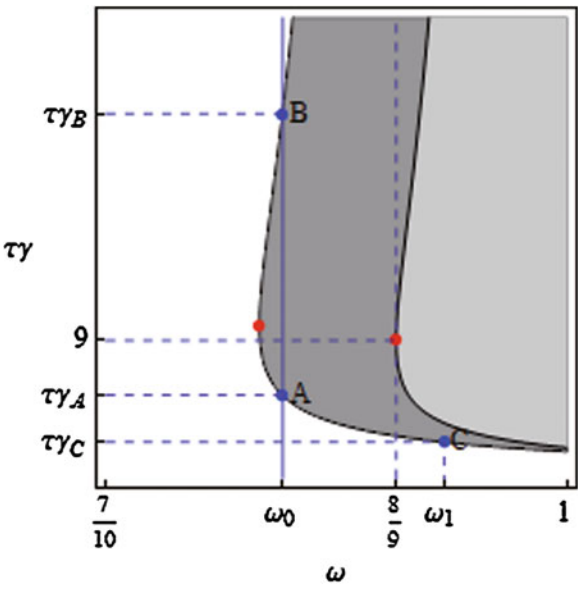

Fig. 5 The partition curves with $m = 1$ and $m = 2$

We turn now to show that Hopf bifurcation can occur on the curve $D_4^3 = 0$. When $D_4^3 = 0$ holds, the characteristic equation is factored as

$$(b_3 + b_1\lambda^2)(b_1b_2 - b_0b_3 + b_1^2\lambda + b_0b_1\lambda^2) = 0.$$

It is clear that the characteristic equation has a pair of purely imaginary roots,

$$\lambda_{1,2} = \pm\sqrt{-\frac{b_3}{b_1}} = \pm i\beta$$

with

$$\beta = \sqrt{\frac{4(2 + 3\gamma(1-\omega)\tau)}{6\tau^2 + \gamma(1-\omega)\tau^3}}$$

and the real parts of other two roots are not zero. So condition (H1) is satisfied.

Assuming that the characteristic root depends on τ and then differentiating the characteristic equation with respect to τ, we obtain the derivative:

$$\frac{\mathrm{d}\lambda}{\mathrm{d}\tau} = -\frac{\frac{3\tau^2}{8}\lambda^4 + \left(\frac{3\tau}{2} + \frac{3\gamma(1-\omega)\tau^2}{8}\right)\lambda^3 + \left(\frac{3}{2} + \frac{3\gamma(1-\omega)\tau}{2}\right)\lambda^2 + \frac{3\gamma(1-\omega)}{2}\lambda}{\frac{\tau^3}{2}\lambda^3 + 3\left(\frac{3\tau^2}{4} + \frac{\gamma(1-\omega)\tau^3}{8}\right)\lambda^2 + 2\left(\frac{3\tau}{2} + \frac{3\gamma(1-\omega)\tau^2}{4}\right)\lambda + \frac{3\gamma(1-\omega)}{2} + 1}.$$

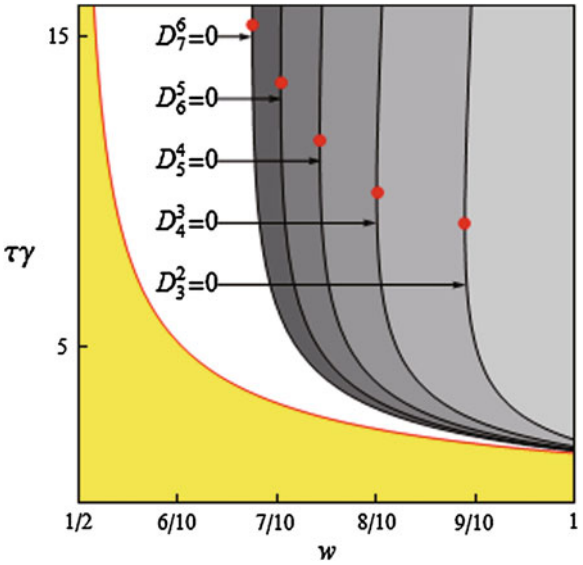

Fig. 6 Stability region and five partition curves

To determine stability switch, we evaluate the derivative at the purely imaginary solution, $\lambda = i\beta$. Thus

$$\mathrm{Re}\left(\left.\frac{\mathrm{d}\lambda}{\mathrm{d}\tau}\right|_{\lambda=i\beta}\right) = \frac{3(2-(1-\omega)\tau\gamma)}{4+\tau^2(16\beta^2+9\gamma^2(1-\omega)^2+12\gamma(1-\omega)\tau)}.$$

We numerically confirm that the real part of the derivative is positive at point $\tau\gamma_A$ and negative at point $\tau\gamma_B$. So the second condition (H2) is also satisfied. Therefore, Hopf bifurcation occurs at both critical points.

In the same way, it can be numerically confirmed that the stability condition is given by $D_{m+2}^{m+1} > 0$ for $2 \le m \le 5$. In Fig. 6, five partition curves $D_{m+2}^{m+1} = 0$ for $m = 1, 2, 3, 4, 5$ are depicted. The rightmost curve is the locus of $D_3^2 = 0$ (i.e., $m = 1$) and the leftmost curve is the locus of $D_7^6 = 0$ (i.e., $m = 5$). The partition curve shifts leftward with the increasing value of m. At the red dot on each curve, equation $D_{m+2}^{m+1} = 0$ has two real and equal roots.[9] As is shown in Case II-1, it is possible to show that for any $m \ge 3$, stability switch can occur twice for ω greater than the abscissa of the red point while the delay becomes harmless for smaller values of ω.

[9] The curves look like steeply shaped hypabolas. However, if we enlarge each curve in the neighbourhood of its red point, then it can be found that the curves take the C-shaped profiles as the curves in Figs. 3 or 5.

Case II-3: $m \to \infty$
As m increases, the weighting function becomes more peaked around $t-s$ and tends to the Dirac delta function. Notice that when m goes to infinity, the first equation of (21) is reduced to a delay differential equation,

$$\dot{q}(t) = \alpha q(t)\left[a - c - 2b\left(\omega q(t-\tau) + (1-\omega)q(t)\right)\right] \tag{24}$$

and the characteristic equation (14) converges to

$$\lambda + \gamma(1-\omega) + \gamma\omega e^{-\lambda\tau} = 0$$

which is the characteristic equation of the delay differential equation (24). To verify the possibility of stability switch for which the characteristic equation must have a pair of purely imaginary conjugate roots, we can assume without loss of generality that $\lambda = i\upsilon,\ \upsilon > 0$. By the real and imaginary parts, the characteristic equation is divided into two equations

$$\gamma(1-\omega) + \gamma\omega\cos\upsilon\tau = 0, \qquad \upsilon - \gamma\omega\sin\upsilon\tau = 0. \tag{25}$$

By moving $\gamma(1-\omega)$ and υ to the right-hand sides of equations in (25), squaring and adding them together, we obtain

$$\upsilon^2 = \gamma^2(2\omega - 1)$$

which is defined only for $\omega > 1/2$, otherwise no stability switch occurs. We then think of the roots as continuous functions in terms of τ and then differentiate the characteristic equation with respect to τ to obtain

$$\left(\frac{\mathrm{d}\lambda}{\mathrm{d}\tau}\right)^{-1} = \frac{e^{\tau\lambda}}{\gamma\omega\lambda} - \frac{\tau}{\lambda} \text{ and } e^{\tau\lambda} = -\frac{\gamma\omega}{\lambda + \gamma(1-\omega)}.$$

Thus

$$\left.\frac{\mathrm{d}(\mathrm{Re}\lambda)^{-1}}{\mathrm{d}\tau}\right|_{\lambda=i\upsilon} = \mathrm{Re}\left(\left.\frac{\mathrm{d}\lambda}{\mathrm{d}\tau}\right|_{\lambda=i\upsilon}\right)^{-1} = \frac{1}{\upsilon^2 + \gamma^2(1-\omega)^2} > 0$$

The last inequality implies that all the roots that cross the imaginary axis at $i\upsilon$ cross from left to right as τ increases.

From (25), we have

$$\gamma\omega\cos\upsilon\tau = -\gamma(1-\omega), \qquad \gamma\omega\sin\upsilon\tau = \upsilon. \tag{26}$$

Hence, there is a unique $\upsilon\tau$, $\pi/2 < \upsilon\tau < \pi$ such that $\upsilon\tau$ makes both equations in (26) hold. Using the first equation we derive the partition curve

$$\gamma\tau = \frac{\cos^{-1}\left(-\dfrac{1-\omega}{\omega}\right)}{\sqrt{2\omega-1}}$$

which is defined for $\omega > 1/2$. In Fig. 6, in addition to the five partition curves, the downward sloping hyperbolic red partition curve of the fixed delay case is illustrated. The red points have the same meaning as in Fig. 5. The monopoly equilibrium with the fixed delay is locally asymptotically stable in the yellow region. We can summarise three results obtained in Cases II-2 and II-3:

Proposition 4.2 *In the case with $\tau > 0$ and $m \geq 1$, (1) increasing m has a destabilising effect in the sense that it decreases the stability region; (2) the stability region with continuously distributed time delay is larger than the one with fixed time delay and the former converges to the latter as m goes to infinity; (3) the stability switch, if possible, occurs twice, implying that the equilibrium is locally stable for smaller or larger values of $\tau\gamma$ while it bifurcates to a limit cycle for medium values.*

5 Concluding Remarks

In this paper, a boundedly rational monopoly with a continuously distributed time delay is examined. Constructing a gradient dynamic system where the rate of the output change is proportional to the derivative of the expected profit, the following results are analytically and numerically demonstrated.

In the case of a single continuously distributed delay, the asymptotic properties of the monopoly equilibrium depend on the shape of the weighting function. If it is exponentially declining, then the delay is harmless and in the case of bell-shaped weighting functions, stability is lost when the delay parameter τ exceeds certain threshold. This value is decreasing in the shape parameter m, implying that the stability region decreases in m. It is also demonstrated that as m tends to infinity, the stability region converges to that with fixed delay. The cases of adaptive expectation were then examined, when the expectation is convex linear combination of the delayed and the instantaneous data. We have shown that the case of $m = 0$ is harmless again. If $m > 0$, then delay is harmless for small values of ω, and then the increasing value of m has a destabilising effect and stability switch occurs twice implying that the equilibrium is locally stable for small and large values of $\tau\gamma$ while it bifurcates to a limit cycle for medium values. These behavioural differences between the monopoly with continuous delay and the delay monopoly with adaptive expectation are found in the different forms of the partition curves as depicted in Figs. 2 and 3.

In addition to analytic investigation, numerical simulation was used to illustrate the theoretical findings and examining global asymptotic behavior.

Acknowledgments The authors are particularly grateful to Roberto Dieci and a referee for their comments and suggestions. They highly appreciate financial supports from the MEXT-Supported Programme for the Strategic Research Foundation at Private Universities 2013-2017, the Japan

Society for the Promotion of Science (Grant-in-Aid for Scientific Research (C) 24530201 and 25380238) and Chuo University (Joint Research Grant 0981 and Grant for Special Research). The paper was prepared when the second author visited the Department of Economics, Chuo University. He appreciates its hospitality during his stay. The usual disclaimer applies.

References

Bellman, R., & Cooke, K. L. (1956). *Differential-difference equations*. New York: Academic Press.

Bischi, G. -I., Chiarella, C., Kopel, M., & Szidarovszky, F. (2010). *Nonlinear oligopolies: Stability and bifurcations*. Berlin/Heidelberg/New York: Springer.

Chiarella, C., & Khomin, A. (1996). An analysis of the complex behavior of nonlinear oligopoly models with time delays. *Chaos, Solitons and Fractals*, *7*(12), 2049–2065.

Chiarella, C., & Szidarovszky, F. (2001). The birth of limit cycles in nonlinear oligopolies with continuously distributed information lags. In M. Dror, P. L'Ecuyer, & F. Szidarovszky (Eds.), *Modeling uncertainty* (pp. 249–268). Dordrecht: Kluwer.

Chiarella, C., & Szidarovszky, F. (2004). Dynamic oligopolies without full information and with continuously distributed time lags. *Journal of Economic Behavior and Organization*, *54*(4), 495–511.

Cushing, J. (1977). *Integro-differential equations and delay models in population dynamics*. Berlin/Heidelberg/New York: Springer.

Gandolfo, G. (2009). *Economic dynamics* (4th ed.). Berlin/Heidelberg/New York: Springer.

Howroyd, T., & Russel, A. (1984). Cournot oligopoly models with time delays. *Journal of Mathematical Economics*, *13*, 97–103.

Invernizzi, S., & Medio, A. (1991). On lags and chaos in economic dynamic models. *Journal of Mathematical Economics*, *20*, 521–550.

Kuang, Y. (1993). *Differential equations with applications in population dynamics*. Boston: Academic Press.

Matsumoto, A. (2009). Note on Goodwin's 1951 nonlinear accelerator model with an investment delay. *Journal of Economic Dynamics and Control*, *33*(4), 832–842.

Matsumoto, A., & Szidarovszky, F. (2012a). Nonlinear delay monopoly with bounded rationality. *Chaos, Solitons and Fractals*, *45*, 507–519.

Matsumoto, A., & Szidarovszky, F. (2012b). Dynamic monopoly with multiple continuously distributed time delays. IERCU Discussion Paper #184. Institute of Economic Research, Chuo University (http://www2.chuo-u.ac.jp/keizaiken/discuss.htm).

Matsumoto, A., & Szidarovszky, F. (2010). Delay differential nonlinear economic models. in G. I. Bischi, C. Chiarella, & L. Gardini (Eds.), *Nonlinear dynamics in economics, finance and social sciences* (pp. 195–214). Berlin/Heidelberg/New York: Springer.

Naimzada, A., & Ricchiuti, G. (2008). Complex dynamics in a monopoly with a rule of thumb. *Applied Mathematics and Computation*, *203*, 921–925.

Puu, T. (1995). The chaotic monopolist. *Chaos, Solitons and Fractals*, *5*, 35–44.

Learning and Macro-Economic Dynamics

Simone Landini, Mauro Gallegati, Joseph E. Stiglitz, Xihao Li and Corrado Di Guilmi

1 Introduction

According to a substantial and growing stream of literature (Chiarella and He 2002; Gaffeo and Delli Gatti 2008; Giansante et al. 2012; Gintis 2007, 2013; Kirman 2012; LeBaron 2002; Lengnick 2013; Markose 2005; Tesfatsion 2003; Tesfatsion and Judd 2006), which aims to go beyond the representative agent-based modelling (see Hartley 1997), the economy is conceived as an ensemble of many heterogeneous-interacting-learning agents.

The present study focuses on the macro effects induced by learning. Such an interest moves from Blume and Easley (1993): when learning is combined with interaction and heterogeneity, both individual perceptions of the circumstances and the circumstances change due to individual mutations. This feedback leads to periods of dominance of certain behavioural species, which alternates through time along a sequence of regime switching, i.e. changes in the configurations of the system over a space of behavioural states, superimposing with phase transitions. The system can therefore be understood as an ecology of learning strategies, coevolving and mix-

S. Landini (✉)
I.R.E.S. Piemonte, Turin, Italy
e-mail: landini@ires.piemonte.it

M. Gallegati · X. Li
DiSES, Università Politecnica delle Marche, Ancona, Italy
e-mail: mauro.gallegati@univpm.it

X. Li
e-mail: xihao.li@gmail.com

J. E. Stiglitz
Columbia Business School, Columbia University, New York, NY, USA
e-mail: jes322@columbia.edu

C. D. Guilmi
UTS Business School, University of Technology, Sydney,
Sydney, NSW, Australia
e-mail: Corrado.DiGuilmi@uts.edu.au

R. Dieci et al. (eds.), *Nonlinear Economic Dynamics and Financial Modelling*,
DOI: 10.1007/978-3-319-07470-2_8,

ing through time, reshaping the system as a well-stirred mixture of heterogeneous behavioural attitudes, realising states of dominance of given species as the economic state of the system changes. This phenomenon of "regenerative coordination" (see Landini et al. 2014) involves both individual and collective learning, together with intermediate variants (see Vriend 2000) by allowing agents for learning in decision-making, obeying a vital impulse for surviving and improving, regenerative coordination determines regime switching and phase transitions consistently with temporary equilibria.

The research questions in the present chapter are the following: Allowing agents for learning in choosing one among a set of differentiated behaviours, is there a possibility for the emergence of a dominant behaviour? If yes, does this behaviour dominate in the long run or can different regimes can alternate through time? What determines such transitions? Are dominant behavioural regimes consistent with specific phases of the system? Are system's phase transitions a cause or an effect (or both) for the micro-learning behaviours superimposition?

The chapter is structured as follows. Section 2 introduces some ontological notes on system's constituents. Section 3 provides a detailed description of the microeconomic model as the DGP of aggregate data. Section 4 introduces the behavioural rules that agents can choose while learning. Section 5 deals with comments and interpretation of the ABM-DGP outcomes. Section 6 concerns the inferential techniques to describe the dynamics of system's level observables providing insights regarding the effect of learning in the financially constrained economy. Section 7 summarises the main findings.

2 Few Ontological Notes

An agent is a "social atom" Buchanan (2007), a minimal (i.e. elementary) constituent of the system able to learn how to behave at its best given the constraints in making decisions. A social atom does not behave in the only way it can, as atoms do, but mainly in the way it wants, when possible, to satisfy its vital impulse to survive and improve. Social atoms have both anticipatory capabilities and backward looking instruments; they are endowed with a set of information to operate "just in case", just like memory. Atoms do not decide whether to interact or not: if they are allowed to interact the outcome is almost predictable. On the contrary, usually, a social atom decides whether or not to interact and it can even choose the counterparts in interactions, often with unpredictable outcomes. A social atom is an active (see Schweitzer 2003) and lively minimal cell with consciousness of its capabilities, feelings, desires, preferences, needs, free will, regrets and a vital impulse to behave at the best it can in order to stay alive and, possibly, improve its state.

Basically, the difference between complex adaptive agents in the ABM literature and social atoms is that the latter does not solely behave *adaptively* by *updating* their state, but they mainly behave *proactively* and evolve while upgrading their state. Conscious of their capabilities, needs and limits, they involve *learning as an*

anticipatory resource, which might also, but not only, involve adaptation: that is, they do their best "today" to realise "tomorrow" what they want or needed, as firmly as possible. This does not mean that they only forecast or play according to decoded rationality. Of course they can, but the main character is that, due to a vital impulse to survive and improve, learning is involved to reasonably choose among a wide set of solutions, in order to progressively approach a (hopefully) long-lasting desirable state. Obviously, the scenario can change and uncertainty is still a distinctive trait. Nevertheless, the social atom aims at a certain degree of stability once a comfortable condition realises, i.e. making revisions only when necessary, as the new basis to evolve and upgrade. The "social" attribute for such atoms is therefore due to both the intrinsic indispensable propensity to interact, in order to gather what they need from the other heterogeneous atoms, and the learning activity which marks a relevant difference with respect to (natural) atoms beyond a mere analogy with mechanical systems of particles.

Heterogeneity and interaction are entangled categories: due to heterogeneity agents are pushed to interaction, due to interaction outcomes agents increase their degree of heterogeneity. Agents give rise to organised structures lumping about a stereotyped agent or behaviour. A system is therefore a structure of within-homogeneous and between-heterogeneous sub-systems. Two kinds of heterogeneity and interaction are isolated. Heterogeneity can be weak (in endowments) or strong (in behaviours). Interaction can be direct (between two or more agents) or indirect (i.e. *mean-field*) when an agent interacts with a sub-system or when sub-systems interact with each other.

The system is a stylised economy as in Greenwald and Stiglitz (1993) where firms produce the same perishable good in a homogeneous market by using only labour supplied at a fixed wage rate, with a financially constrained production function as in Delli Gatti et al. (2010). When financial resources are not enough to pay the wage bill, they borrow money from a bank, which charges a fixed rate of interest. The firm is characterised by a few elements: a set of information Ω that it uses and produces at the same time; a set of parameters Θ; a set of functions $\Im$. In this simplified economy, there is only one state observable, which is the equity base $A(i,t) \in \Re_+^A$ as a measure of net worth: it characterises the agent's degree of (weak) heterogeneity in endowments.

The functions in the set $\Im$ take the values of the equity and combine them by means of parameters and other information from the environment: each of such combinations defines a quantity and its functional.

The subclass Λ within the set $\Im$ defines the behavioural rules. In a step-by-step learning mechanism as in Landini et al. (2014), each agent tests all the members $\lambda \in \Lambda$ and chooses the optimal one λ^* according to an evolutionary criterion or vital-impulse, coupled with anticipatory capability: *do the most reasonable choice, first to stay alive and, then, to sensibly improve your well-being the way you want, if you can.* The chosen λ^* is a mapping $\lambda^*(.|\Theta):\Omega \rightarrow \Re_+$ for the control parameter $\alpha(i,t) = \alpha(i,t)[\lambda^*]$ characterising the agent's degree of (strong) behavioural heterogeneity since, time by time, in the same or in a different state (way of being), an agent can change its way of behaving. In short, agents can learn how to behave the best they

can. Due to this, $\alpha(i,t)$ is called the *learning parameter*. The agent learns how to properly set it to schedule its financially constrained level of desired output; due to this, $\alpha(i,t)$ is defined as the *scheduling function*.[1]

The complement $H = \Im \backslash \Lambda$ defines all the other observables $h \in H$ subclass, each of these functions is a map $h_m(.,.|\Theta) \in H : \Re_+^A \times \Lambda \to H_m \subset \Re$. Once the control parameter $\alpha(i,t) = \alpha(i,t)[\lambda^*]$ is set, given (i.e. conditioned on) financial resources (i.e. equity $A(i,t)$), the firm schedules its output ($Q(i,t) = h_Q(\alpha(i,t)|A(i,t),\beta)$), which is the control variable. All the other quantities are instrumental to the update of the equity for the next period. The firm's activity is represented by the sequence $\{h_m|\lambda^*\}$ of actions it performs given the decisions it makes: the outcomes represent both *ways of being and of behaving*.

The parameter set $\Theta = \Theta^S \cup \Theta^B$ contains the systemic parameters $\Theta^S = (w,r) \in \Re_+^2$, namely the wage rate and the rate of interest, and the behavioural parameters $\Theta^B = (\beta,\gamma,\delta,\varphi) \in \Re^5$: basic elasticities (β,δ,φ) or scaling parameters (γ,ζ).

3 The Micro-Model

Assume that the control parameter $\alpha(i,t)$ is set according to $\lambda \in \Lambda$, and that the present period equity $A(i,t)$ is updated from the previous period, which could be characterised by a different $\lambda' \in \Lambda$.[2]

The firm schedules the desired output[3] level according to its financial resources

$$Q(i,t) = \alpha(i,t)A(i,t)^{\beta}, \quad \beta \in (0,1) \tag{1}$$

to be realised according for the production function

$$Q(i,t) = (N(i,t)^{1/\delta})/\gamma, \quad \gamma \in (0,1), \ \delta > 0 \tag{2}$$

which gives labour demand as a function of the scheduling (learning) function conditioned on the current level of net worth (endowments)

---

[1] This set-up looks similar to the classifier approach discussed in LeBaron (2002). However, as it will be clearer from Sect. 4 onwards, the involved rules are not "current conditions in the market" as states of the world, but agents' ways of behaving characterising different species. Dynamic conditions in the market are due to the realisations of the aggregate observables, to which the agents contribute with differentiated behaviours. The change in such conditions motivates agents to maintain or switch the actual way of being. Hence, the proposed modelling is less procedural and more phenomenological. However, as the classifier approach of LeBaron (2002), this provides the modeller a "large amount of power design".

[2] The present model follows Landini et al. (2014) introducing nonlinear bankruptcy costs.

[3] Since, given financial resources, this level of output depends on the temporarily chosen rule λ, it should more properly be written as $Q(i,t)[\lambda] = \alpha(i,t)[\lambda].A(i,t)^{\beta}$. This notation is a short hand for the fact that, conditioned on the same equity (state variable), different rules provide different levels of desired output (control variable). The learning mechanism will allow for the final choice.

$$N(i,t) = (\gamma Q(i,t))^{\delta} = \chi \alpha(i,t)^{\delta} A(i,t)^{\phi}, \;\; \chi = \gamma^{\delta}, \; \phi = \beta\delta \tag{3}$$

This is how learning and endowments (i.e. strong and weak heterogeneity) enter the model. Being $w > 0$ the constant wage rate, the wage bill is $W(i,t) = wN(i,t)$. The difference $L(i,t) = W(i,t) - A(i,t)$ gives a credit flow revealing the state of financial soundness: $X(i,t) = 1$ if $L(i,t) \geq 0$. This is how learning and endowments (i.e. strong and weak heterogeneity) enter the model. The difference gives a credit flow revealing the state of financial soundness: if means the firm is not self-financing (NSF), otherwise it is self-financing (SF). The SF firm deposits the amount $|L(i,t)|$ in the bank, without earning interest, the NSF firm borrows $|L(i,t)|$ from the bank, charged by a fixed interest rate $r > 0$, with financial commitments $F(i,t) = X(i,t) r L(i,t)$. Production costs are: means the firm is NSF, otherwise it is SF. The SF firm deposits the amount in the bank without earning interest, the NSF firm borrows from the bank charged by a fixed interest rate with financial commitments. Production costs are: $C(i,t) = W(i,t) + F(i,t)$.

In general, the firm faces bankruptcy costs,

$$B(i,t) = X(i,t)\zeta Q(i,t)^{\varphi} = X(i,t)\zeta \alpha(i,t)^{\varphi} A(i,t)^{\eta}, \;\; \zeta > 0, \;\; \varphi > 0, \;\; \eta = \beta\varphi \tag{4}$$

but, if the firm is SF bankruptcy costs are null: $X(i,t) = 0$. Revenues are the product of output and the individual price $Y(i,t+1) = p(i,t+1)Q(i,t)$. For the sake of simplicity, the market price is approximated as $P(t+1) = W(t)/Q(t)$, being $W(t) = w\sum_i N(i,t)$ the realised total demand, i.e. the aggregate value of wage bills, and $Q(t) = \sum_i Q(i,t)$ the output. The individual price is calculated as a multiplicative shock on the market price $p(i,t+1) = u(i,t)P(t+1)$, where $u(i,t) \longrightarrow^{i.i.d.} U(u, 2-u)$ is an idiosyncratic shock such that $E[p(i,t+1)] = P(t+1)$.

Profit is the difference between revenues and production costs. The profit curve, which accounts for bankruptcy costs only for NSF firms, is

$$\begin{aligned} \tilde{\Pi}(i,t+1) &= \Pi(i,t+1) - B(i,t) \\ &= p(i,t+1)\alpha(i,t)A(i,t)^{\beta} - (1 + X(i,t)r)\theta\alpha(i,t)^{\delta} A(i,t)^{\varphi} \\ &\quad - X(i,t)[\zeta\alpha(i,t)^{\phi} A(i,t)^{\eta} - rA(i,t)] \end{aligned} \tag{5}$$

in case of SF firms $X(i,t) = 0$ implies $\tilde{\Pi}(i,t+1) = \Pi(i,t+1)$.

Finally, equity updates by accumulating profits with the present period equity: $A(i,t+1) = A(i,t) + \Pi(i,t+1)$.

Profit maximization. According to the shorthand notation ($\alpha_t = \alpha(i,t), a_{t+k} = A(i,t+k), \; \tilde{\pi}_{t+1} = \tilde{\Pi}(i,t+1), \; x_t = X(i,t), \; p_{t+1} = p(i,t+1)$) set the following implicit optimization problem (IOP) on the scheduling (learning) parameter:

$$\text{IOP:} \begin{cases} \alpha_t^* = \arg\max O(\alpha_t) = E\{\tilde{\pi}(a_t, \alpha_t, p_{t+1}|x_t)\} s.t. \\ c(\alpha_t) = a(a_t, E\{\tilde{\pi}(a_t, \alpha_t, p_{t+1}|x_t)\} = a_t + O(\alpha_t) = a_{t+1} \geq 0 \\ \alpha_t \geq 0 \end{cases} \tag{6}$$

For the IOP to be well-posed, the objective function $O(\alpha_t)$ must be concave and the constraint $c(\alpha_t)$ convex. To show the unconstrained first order conditions (uFOC) are necessary and sufficient, apply the expectation on the profit function in (5). Hence the uFOC are

$$\partial_\alpha O(\alpha_t) > 0 \Rightarrow P_{t+1}(\alpha_t a_t^\beta) > (1 + x_t r)\delta\theta(\alpha_t a_t^\beta)^\delta + x_t \varphi\zeta(\alpha_t a_t^\beta)^\varphi \tag{7}$$

which does not allow for a closed-form solution w.r.t. α_t unless, beyond a suitable *ad hoc* polynomial expression w.r.t. α_t is set, a parametric constraint $0 < \omega \equiv (\delta = \varphi) \neq 1$ is assumed allowing for non-linear bankruptcy costs in (4). By using this assumption in (7) it can be found that

$$\begin{aligned} 0 < \alpha_t^* = \alpha^*(a_t, P_{t+1}|x_t) = \left[\frac{P_{t+1}}{\omega(\theta(1 + x_t r) + x_t \zeta)}\right]^{\frac{1}{\omega - 1}} a_t^{-\beta} \\ \text{s.t.} \quad 0 < \omega \equiv (\delta = \varphi) \neq 1 \end{aligned} \tag{8}$$

which fulfils the third constraint in the IOP, while

$$\partial_\alpha O(\alpha_t) \begin{cases} > & 0 \quad \text{iff } \alpha_t \in (0, \alpha_t^*) \\ = & 0 \quad \text{iff } \alpha_t = \alpha_t^* \\ < & 0 \quad \text{iff } \alpha_t > \alpha_t^* \end{cases} \tag{9}$$

Equation (9) ensures that $\alpha_t^* = \alpha^*(a_t, P_{t+1}|x_t)$ is the always existing unique profit maximizing global optimum point, which dynamically changes as the equity (a_t) and the state of financial soundness (x_t) change under the effect of the market price (P_{t+1}).

The IOP is a constrained problem, hence the constraint $c(\alpha_t^*) \geq 0$ must be fulfilled to avoid next-period equity being negative, which implies bankruptcy. Due to the vital-impulse of firms to stay alive and make profits to improve, this cannot be rationally accepted. Therefore, even though the optimum point certainly exists, the feasibility condition $c(\alpha_t^*) \geq 0$ must be met. Therefore, consider the constraint in explicit form

$$c(\alpha_t) = P_{t+1}\alpha_t a_t^\beta - (1 + x_t r)[\theta\alpha_t^\delta a_t^\phi - a_t] - x_t(\zeta\alpha_t^\varphi a_t^\eta) \geq 0 \tag{10}$$

and substitute for $0 < \omega \equiv (\delta = \varphi) \neq 1$ to get

$$c(\alpha_t) = P_{t+1}\alpha_t a_t^{\beta} - (1 + x_t r)[\theta \alpha_t^{\omega} a_t^{\beta\omega} - a_t] - x_t(\zeta \alpha_t^{\omega} a_t^{\beta\omega}) \geq 0 \quad (11)$$

which gives

$$c(\alpha_t) = (\alpha_t a_t^{\beta})\Big[P_{t+1} - (\theta(1 + x_t r) + x_t \zeta)(\alpha_t a_t^{\beta})^{\omega-1}\Big] + (1 + x_t r)a_t \geq 0 \quad (12)$$

Simplifying for a_t gives

$$c(\alpha_t) = (\alpha_t a_t^{\beta-1})\Big[P_{t+1} - (\theta(1 + x_t r) + x_t \zeta)(\alpha_t a_t^{\beta})^{\omega-1}\Big] + (1 + x_t r) \geq 0 \quad (13)$$

Substitution of (8) and multiplication by a_t leads to

$$c(\alpha_t^*) = \left[\frac{P_{t+1}}{\omega(\theta(1 + x_t r) + x_t \zeta)}\right]^{\frac{1}{\omega-1}} \left(\frac{\omega - 1}{\omega}\right) P_{t+1} + (1 + x_t r)a_t \geq 0 \quad (14)$$

If $\omega > 1$ then $c(\alpha_t^*) \geq 0$ is always fulfilled, $\omega = 1$ is not feasible. A solution can be identified by relating equity and market price but it is not generic, although it considers the specific situation of the firm (i.e. by means of a_t) within its environment (i.e. by means of P_{t+1}) both for SF and NSF (x_t). As discussed so far, the parabolic hypothesis is needed in order to obtain a generic solution: as shown in Landini et al. (2014), a sufficient condition for the profit curve to be parabolic is $\delta \in (0, 1/\beta) : \beta \in (0, 1)$. The assumption therefore becomes stronger in order to obtain a closed-form feasible solution: $\omega > 1/\beta : \beta \in (0, 1)$ ensures $\omega > 1$ and $c(\alpha_t^*) \geq 0$.

The economic implication of the parabolic assumption $0 < 1/\beta < \omega \equiv (\delta = \varphi) \neq 1 : \beta \in (0, 1)$. Consider $E_N(\alpha_t) = \delta > 0$ in (2) is the elasticity of labour demand (3) w.r.t. the scheduling parameter. Consider $E_B(\alpha_t) = \varphi > 0$ as the elasticity of bankruptcy costs (4) w.r.t. the scheduling parameter. Since bankruptcy costs depend on labour demand through output scheduling (4), (1) and since labour demand depends on scheduled output (3), (1), the assumption states that, for any given level of equity and for any state of financial soundness, for a profit maximizing firm a change in the scheduling parameter affects labour demand and this effect transmits to bankruptcy costs with the same magnitude $0 < E_N(\alpha_t) = E_B(\alpha_t) \equiv \omega \neq 1$. As far as $\omega > 1/\beta$, and $E_Q(a_t) = \beta \in (0, 1)$ then $E_N(\alpha_t) = E_B(\alpha_t) > 1/E_Q(a_t)$. Accordingly, the elasticities of labour demand and bankruptcy costs w.r.t the scheduling parameter are inversely related to that of output w.r.t. equity: by tuning the values of $0 < 1/\beta < \omega \equiv (\delta = \varphi) \neq 1 : \beta \in (0, 1)$, different scenarios follow.

Table 1 Stereotyped behaviours taxonomy

Learning	Interaction			
	None	Micro	Local	Global
None	λ_1 elementary			
Individual	λ_2 trend followers λ_3 profit maximizers	λ_4 micro-imitators		
Collective			λ_5 local imitators λ_6 best performers followers	λ_7 average followers

4 Behavioural Rules and Species in the Economy

The set $\Lambda = \{\lambda_k : k \leq K\}$ of scheduling rules[4] for $\alpha(i,t) = \alpha(i,t)[\lambda_k]$ is represented in Table 1. The classification criteria follow two dimensions, making explicit the strong heterogeneity for the possible behaviours.

Interaction can be indirect or direct. If it is *indirect (i.e. mean-field)* it is concerned with the interaction of a single agent with a system. If it is *global*, the firm interacts with the system as a whole. If it is *local*, the firm interacts with a sub-system characterised by a given scheduling rule. If it is direct, it takes place between two agents. As regards the learning side, if it is present in the specific rule, it can be *collective*, so coupling with interaction, or *individual*, i.e. without interaction.

- Rule λ_1—*not interactive without learning: elementary.* The learning parameter is constant through time and homogeneous across firms using it.
- Rule λ_2—*not interactive with self-referential backward looking learning: trend follower.* It models the self-referential firms looking at their past periods without care for the environment. This rule introduces some memory degree: the higher the value of τ the longer the memory.
- Rule λ_3—*not interactive with self-referential, profit-maximizing learning: profit maximizer.* The firm looks at itself and, given its endowments, sets the output scheduling parameter by choosing the value maximizing its expected profit.
- Rule λ_4—*micro-interaction with individual learning: micro-imitator.* A firm i directly interacts with a randomly sampled firm j. The firm i adopts the previous period scheduling parameter value of the firm j if i's profit growth rate was lower than that of j. This introduces direct interaction as imitation. The imitator does not look at the way the scheduling parameter is computed by its counterpart.

[4] Some of the rules are specified in Landini et al. (2014). Rules 1 and 3 are preserved, rule 2 takes the place of rule 4, rule 5 that of 6, rule 6 that of 7 and rule 7 that of 5. The change in positions is due to the classifications criteria here involved, Table 1. Rule 4 is new and what was rule 2 has been dropped. Rule 3 always concerns profit maximization but, here, it introduces nonlinear bankruptcy costs. All rules obey the spanning–learning as described in the following.

- Rule λ_5—*local interaction with collective learning: local-imitator.* A random sample of M firms in the same state of financial soundness is associated to a firm i, $s_i = (i_1, \ldots, i_M | X(i,t))$. The firm sets its output scheduling parameter value as the average value of parameters in its sample. Learning is collective in the sense that information is drawn from a collection of firms and interaction is indirect because the firm chooses by looking at the sub-system value of the output scheduling parameter.
- Rule λ_6—*local interaction with collective learning: best performers follower.* Firms in the economy are split into NSF and SF. A firm looks at its own group, uses the ratio of profit to equity to measure other firms' performance and considers as best performers those with a ratio higher than its own. The sample of reference for a firm i is defined as $s_i^* = (i_1, \ldots, i_{M_i^*} | X(i,t) = X(i_j, t))$ where $[\Pi(j,t)/A(j,t)] > [\Pi(i,t)/A(i,t)]\ \forall j : X(j,t) = X(i,t)$. The firm calculates its scheduling parameter value as the average value in its reference sample of best performers. Differently from rule λ_5, the dimension M_i^* of the reference sample in λ_6 can change through time and firm by firm.
- Rule λ_7—*global interaction with collective learning: average follower.* A firm i sets its own output scheduling parameter at the average value in the system.

The learning mechanism considers the time spanning for learning: once the firm has chosen the scheduling parameter according to a given rule, it keeps it unchanged for more than one period. Before this, the firm tries all the rules at $t + k$ and chooses the one with the highest value of expected profit to increase the next period equity in the attempt to improve the state of financial soundness. Since the scheduling parameter is used to programme the production in the next period, it has been considered the firms are using this parameter value for $T = 5$ periods. The spanning–learning test holds on $[t + k, t + k + T)$ and involves two activities. At $t + k$ a rule is chosen to be used from the beginning till the end of the period. At the end of the period, a new decision is made and its outcome can be to maintain the rule previously chosen or to adopt a different one. In any case, the scheduling parameter value will change since the internal state of the firm has changed according to its financial resources and to the state of the economy.

The credit flow values $L = \{L(i,t) : i \leq I\}$ of firms in the economy can be partitioned into several (H) intervals to determine states of financial soundness $\Sigma = \{\varsigma_h : h \leq H\}$. For the sake of simplicity $H = 2$ states are considered, that is, $\varsigma_1 = NSF$ if $X(i,t) = 1$ and $\varsigma_0 = SF$ if $X(i,t) = 0$.

The space $\Xi = \Lambda \times \Sigma = \{\zeta_j = (\lambda_k \wedge \varsigma_h) : j = H(k-1) + h\}$ defines the species in the economy, both according to financial soundness and behavioural attitudes. Therefore, being $K = 7$ and $H = 2$ there are $J = KH = 14$ species in the system. With $x_i(\lambda_k \wedge \varsigma_h; t) \equiv X_j$ it is meant that, at time t, a randomly drawn agent i is a firm of the species j since it is found in the h-th state of financial fragility. The occupation number $I_j(t) = \sharp\{x_i(\lambda_k \wedge \varsigma_h;) : i \leq I\}$ counts how many firms belong to the j-th state on Ξ. The total number of firms never changes, and the vector $\mathbf{I}(t) = (I_i(t), \ldots, I_J(t)) \in N \subset \mathbb{N}_0^J$ is the configuration of the system.

5 Simulation Results

The following results are the outcomes of a 50 runs Monte Carlo simulation of the ABM model with parameter setting: $\beta = 0.5, \gamma = 1.9, \delta = 1.4, w = 1, r = 0.05, \varphi = 1.4, \zeta = 0.1$ and $\tau = 3$. Firms' initial equity level is randomly drawn from a uniform distribution with support between 0 and 20, and the shock in price u is drawn from a uniform distribution with support between 0 and 2.

Figure 1 shows the aggregate time series from the ABM-DGP with a population of $N = 1{,}000$ firms for $T = 1{,}000$ periods. On average, the economy is populated with a share of 39 % of NSF firms and 61 % SF firms. NSF firms concentrate 26 % of total equity while they realise 79 % of total output, since, on average, their output scheduling parameter α is 10 times the SF one. As a consequence, NSF firms realise 62 % of total profit on average. On the other hand, SF firms possess 74 % of total equity but they realise only 21 % of total output with 38 % of total profit on average. As 39 % population of NSF firms own only 26 % of total equity, whereas 61 % of SF firms own 74 % of total equity, NSF firms on average have smaller per-capita equity than SF firms.

A *dominant configuration* is a ranking of the behavioural rules at period t, according to the decreasing order of concentration of a given quantity Z on each learning rule $\lambda \in \Lambda$.[5] By considering the quantity to be the number of firms, $Z = I$, the dynamics of the dominant configuration defines the *diffusion-dominance*. On the other hand, by considering the quantity to be the equity, the output, the wage bill or the profit ($Z = A, Q, W, \Pi$) it defines the *effects-dominance*. Figure 2 shows that the diffusion-dominance does not overlap with the effect-dominance. Moreover, firms with different financial fragility (NSF/SF) follow different patterns of diffusion-dominance and effect-dominance.

Figure 1 also shows that the NSF scheduling parameter α[6] is almost 10 times of SF firms. NSF firms are more "aggressive" than SF ones because the vital impulse of NSF firms pushes them mainly to recover their financial fragility by seeking for profit-improving behavioural rules while SF firms are more prudent: the sounder the financial health, the less the incentive to change.

A *regime* is a sub-period characterised by a given dominant configuration, either in diffusion-dominance or in effect-dominance. A *phase* is a sub-period during which a quantity Z, or its growth rate, is in a certain state with specific qualitative characteristics such as expansion or contraction. Therefore, the system can go through both regime and phase transitions. Due to the intrinsic complexity in the microscopic behaviour (heterogeneity, interaction and learning) and the external effects from the environment, it may happen that phases of the same type in different sub-periods are synchronised with different types of regimes. For instance, as regards the SF bottom

[5] For example, if the quantity $Z = I$ is the number of firms, a dominant configuration like [7132645] at period t shows that firms are most concentrated in rule 7 (average followers) at that period, the second most concentrated is rule 1 (elementary) and so on.

[6] The aggregate estimate of α has been computed using (1) with aggregate data solving the equation for α , that is $\alpha(t) = Q(t)/A(t)^{\beta}$.

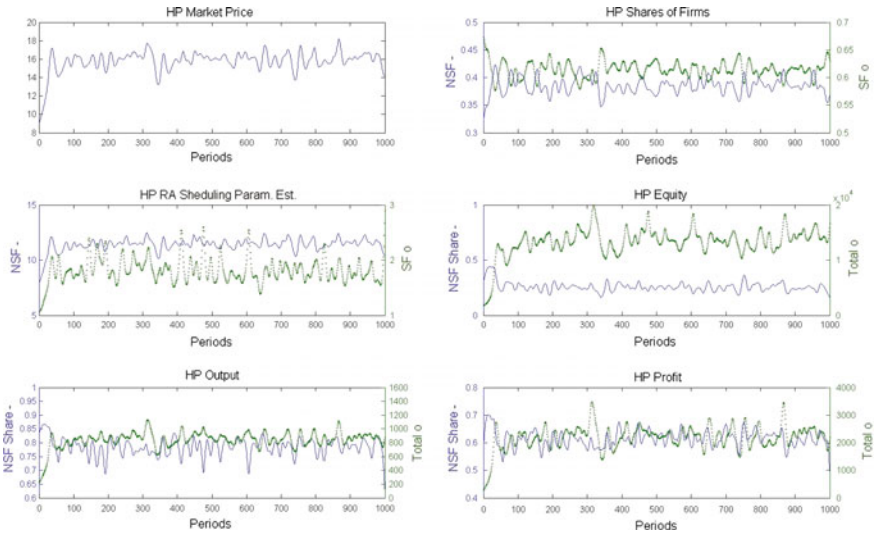

Fig. 1 Monte Carlo simulation (50 runs of the ABM-DGP): HP-filtered aggregate quantities. The time series for the scheduling parameter is found by solving (1) for aggregate quantities (output and equity)

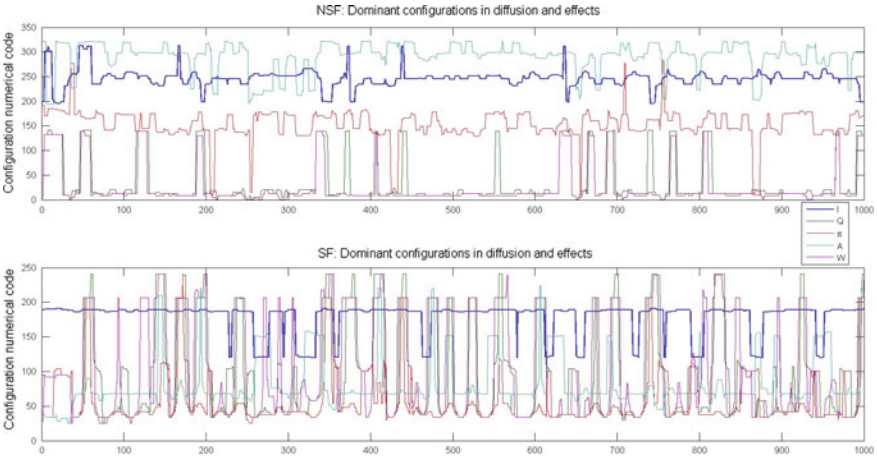

Fig. 2 Dominant configurations for *NSF* and *SF*: (I) is the diffusion-dominance

panel, Fig. 3 shows two expansion phases in sub-periods τ_1 and τ_2. Sub-period τ_1 is synchronised with the dominant configuration of [3654271] and [3652471], whereas sub-period τ_2 is synchronised with the configuration [6354271]. A possible reason for this is that the micro-level structure changes with time. Therefore, a dominant configuration in a high-performance phase in a given period can become dominated in another high performance period because the conditions in the system are different.

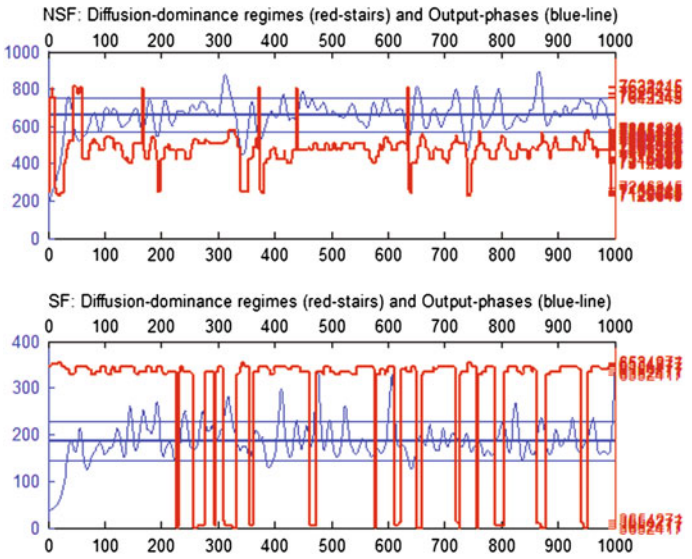

Fig. 3 *NSF* and *SF* diffusion-dominance (*red-stairs*) and output-phases (*blue-line*). *Horizontal lines* the time average and confidence bands about the mean (+/− standard deviation)

Further investigation in Fig. 3 indicates that NSF firms behave differently from SF ones in learning activities. NSF firms involve 48 dominant configurations, whereas SF firms only 10. Moreover, NSF firms and SF firms do not share any dominant configuration in common. Therefore, as social atoms, firms do not only change the way of behaving but they also change the way of being.

Moreover, as Fig. 3 shows, the diffusion dominance regimes follow the output phases trajectory with some delay at certain points in time. Basically, at the turning point of the output dynamics, one can observe jumps or breaks in the configuration dominance, after which the output phase keeps on growing or decreasing. Switches in the behavioural rules still occur but they basically consist in recombination of some dominant configuration. This is the effect of the regenerative-coordination: when some regularity induced by agents behaviour faces a criticality, then there is change in agents' behaviour that destroys the previous setting and creates a new one [see Landini et al. (2014)].

6 Macro-Dynamics

Agents' behaviours allow for transitions in the two dimensions of the state space $\Xi = \Lambda \times \Sigma$: along the Σ-dim it concerns transitions over degrees of financial fragility, along the Λ-dim it concerns the change in behaviour. As regards the Σ-dim, between-transitions realise when a firm jumps from the NSF state to the SF state

and vice versa. As regards the Λ-dim behavioural rule changes can be observed. Consider then the case of a firm $x_i(\lambda_k \wedge \varsigma_h; t) \equiv X_j$.[7] It then can happen that: (a) $x_i(\lambda_k \wedge \varsigma_{h'}; t+1) \equiv X_{jh}$ if it maintains the same behavioural rule while assuming a different state of financial soundness; (b) $x_i(\lambda_{k'} \wedge \varsigma_h; t+1) \equiv X_{jk}$ if it persists in the same state of financial soundness while changing the behavioural rule; (c) $x_i(\lambda_{k'} \wedge \varsigma_{h'}; t+1) \equiv X_{j'}$ if it changes both the state of financial soundness and the behavioural rule; (d) $x_i(\lambda_k \wedge \varsigma_h; t+1) \equiv X_j$ if nothing has changed. The first three cases concern movers, the last case concerns *stayers*.

Conditional transitions are analysed by means of the master equations techniques discussed in Landini et al. (2014). Transitions are of two kinds. Financial fragility state transitions (*between-transitions*) are $x_i(* \wedge \varsigma_a; t) \rightarrow x_i(* \wedge \varsigma_b; t+1)$, without care for the learning behaviour: such transitions are analysed by means of first master equation to determine the evolution of occupation numbers in NSF and SF states. This kind of master equation has been introduced in the economic literature by Aoki (1996), Aoki (2002) and Aoki and Yoshikawa (2006), and applied to different economic problems,[8] none of them involving learning agents.

The second kind of transitions are the so-called *within-transitions* and is concerned with transitions like $x_i(\lambda_h|\varsigma; t) \equiv X_h \rightarrow x_i(\lambda_k|\varsigma; t+1) \equiv X_k$. Transitions on the rules space Λ are conditioned on the state of financial soundness. These kinds of transitions have been introduced by developing combinatorial master equations drawing techniques from stochastic kinetics of chemical reactions (see Gardiner and Chaturvedi 1977; Gardiner 1985).

6.1 The Financial Soundness State Master Equation

Between-transitions $x_i(* \wedge \varsigma_a; t) \rightarrow x_i(* \wedge \varsigma_b; t+1)$ consider NSF and SF firm shares dynamics in the economy. Since only two alternative states of financial fragility are present, the target density is assumed to be the NSF one, the SF concentration is found by complement,

$$S_1(t) = I(\varsigma_1; t) := \sharp_t\{X(i,t) = 1 : i \leq I\} : S_0(t) = I - S_1(t) \tag{15}$$

The basic hypothesis is that of a regenerative-coordination: firms' behaviours are influenced by those macro-dynamic effects to which they contributed through the channel of the market price (see 11).

A master equation in this case is a time-continuous, space-discrete differential equation for the dynamics of the distribution of $S_1(t)$

[7] Differently from the ABM of Sect. 3, from hereon the symbol X refers to species and not to the state of financial soundness, which is $\varsigma \in \Sigma$.

[8] Other references on this topic are Aoki (1996), Aoki (2002), Aoki and Yoshikawa (2006), Delli Gatti et al. (2012), Guilmi (2008), Di Guilmi et al. (2010), Di Guilmi et al. (2012) and Landini and Uberti (2008).

$$\frac{dP(n,t)}{dt} = [b(n-1,t)P(n-1,t) + d(n+1,t)P(n+1,t)] - [(b(n,t) + d(n,t)P(n,t)] \tag{16}$$

Equation (16) is a balance equation between inflows and outflows into and out from the state value $S_1(t) = n$. The underlying transitory mechanism is consistent with a birth-and-death stochastic process. Except for very few cases, Equation (16) does not admit a closed-form solution, hence some approximation methods[9] are needed. Following Delli Gatti et al. (2012), the van Kampen systematic method (see Kampen 2007) is applied by assuming the following ansatz:

$$S_1(t) = I\phi(t) + \sqrt{I}\varepsilon(t) \tag{17}$$

Equation (17) transforms (16) into a system of coupled equations: namely a Fokker-Planck equation w.r.t. to the distribution $\tilde{P}(\varepsilon, t)$ of the spreading fluctuation component $\varepsilon(t)$ about the drifting trajectory $\phi(t)$, which is the expected share of NSF firms, driven by an ordinary differential equation known as the macroscopic equation. The solution of the macroscopic equation provides the dynamics of the expected value of $S_1(t)$ while the solution of the Fokker-Planck equation gives the distribution of fluctuations around this drift.[10] The specification of the transition rates according to the economic phenomenology of the ABM allows for an endogenous description of fluctuations around the drift as the macroscopic outcomes of the unobservable microscopic interactions.

The higher the market price growth rate the higher the probability of becoming NSF, due to a decrease in output or due to a stimulus to increase profits, which increases credit demand to increase output. Accordingly, it is assumed that the market price plays the role of a pilot-quantity driving the dynamics of financial soundness configurations. Therefore, being γ_t the change of the present period market price with respect to its long-run dynamics, the death-and-birth rates are

$$\zeta(t) = (1 + Erf(\gamma_t))/2 \;\; \iota(t) = 1 - \zeta(t): \;\; \gamma_t = \hat{P}(t)/\bar{P} - 1 \tag{18}$$

where $Erf(\gamma_t)$ is the error function: the death rate ζ is therefore the standard normal c.d.f. at γ_t. The transition rates are then specified as

[9] References Aoki (1996), Aoki (2002) and Aoki and Yoshikawa (2006) provide an introduction to such methods. More advanced references are Gardiner (1985), Risken (1989) and Kampen (2007). Analytic details on the method involved in the present chapter can be found in Landini et al. (2014), Delli Gatti et al. (2010) and Di Guilmi et al. (2013) to which the reader is referred.

[10] In general, this method suffers from some limitations, among others the relevant are: (a) the Fokker-Planck equation is equivalent to that of a second order local approximation in the Kramers-Moyal expansion if, according to Pawula's theorem, no closed-form solution is feasible (see Delli Gatti et al. (2012) and Risken (1989)), (b) the ansatz (17) is suitable for weak-noise unimodal processes. In dealing with a macroscopic observable in large systems, such as the present case, the consequences of such limitations are negligible.

$$\begin{cases} b(S_1(t) - \vartheta, t) = \zeta(t) \cdot [1 - (S_1(t) - \vartheta)/I] \\ d(S_1(t) + \vartheta, t) = \iota(t) \cdot [S_1(t) + \vartheta]/I \end{cases} \tag{19}$$

where terms in square brackets are introduced to account for the fact that an inflow or an outflow depends on the size of the initial state.

Substituting (17) into (19), plugging the result into (16) and then computing the needed derivatives, two main equations are found. The first is for the solution of the macroscopic equation and is given by

$$\phi(t) = \left\{1 + \left(\frac{1}{\phi_0} - 1\right) \exp[-t.B.\Delta(t)]\right\}^{-1} : \Delta(t) = \zeta(t) - \iota(t),\ B = 1 \tag{20}$$

The second is the following Gaussian distribution with zero mean and variance:

$$\varsigma_\varepsilon^2(t) = \varepsilon_{eq}^2(t)[1 - \exp(2.t.\Delta'(t))] : \Delta'(t) = \frac{\partial \Delta(t)}{\partial \phi} \tag{21}$$

for the spreading fluctuation distribution, being $\varepsilon_{eq}^2(t)$ the temporary equilibrium for the spread in the Fokker-Planck equation.

Figure 4 shows the dynamics of the share of NSF firms in the economy. This series is highly coordinated with the dynamics of the market price (correlation: 0.9415). This result confirms that the market price can be qualified as a pilot-quantity for the master equation. The bottom panel shows the dynamics of the master equation solution compared with the ABM outcome. In this case, the overlapping is evident (correlation: 0.6770).

Figure 5 contrasts the ABM outcome with the master equation estimate of the NSF firms share time series. A perfect correspondence is represented by a straight line bisector of the plane, such that the picture represents the deviation from the optimal result. The range of variation of the ABM data is from 32.68 to 42.31 % while the master equation gives a wider interval, from 2.48 to 54.94 %. This is basically due to the initial part of the estimated series. However, there are also intermediate points in time where the estimate deviates from data. The small box represents the relationship between the two series. By classifying periods for both time series in dates within and outside their confidence bands and then computing the joint frequency, it has been found that in the 57.7 % of the dates data points are within the inner box and that the 13.9 % of the dates are outside the inner box. This proves the fitting capability of the master equation: note that no adjustment parameters are introduced in the master equation, which is just one equation while the outcome of the ABM-DGP is due to $I = 1{,}000$ systems of equations like those presented in Sect. 3.

Despite this equation not embedding the learning mechanism, it accounts for its effects in interpolating the ABM outcome. The solution of this master equation is introduced into the second master equation for learning, determining a so-called nested model. Moreover, even though ABM micro-data are available, they are not used except for their aggregation which represents the benchmark. Therefore, the

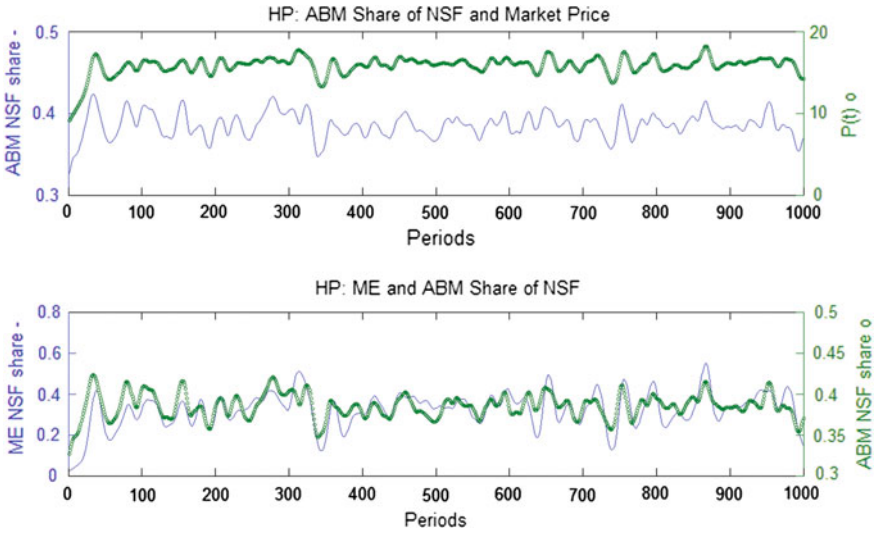

Fig. 4 Dynamics of the estimated share of NSF firms compared with ABM results. Time series have been filtered using the Hodrick-Prescott filter to show the underlying trend components

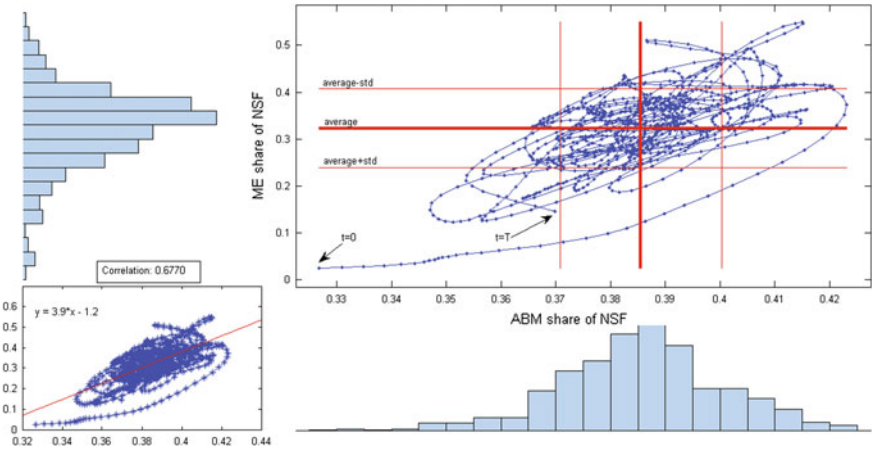

Fig. 5 Superimposition of the ABM outcome with the master equation solution

method extends to situations in which only macro-data are available. Finally, no assumption is made about the distribution of the stochastic process, but it is inferred to be Gaussian and its dynamics is completely determined by transition rates. These rates are phenomenologically specified on the ground of the consequences of the economic theory implemented in the ABM. This stream of modelling is promising in developing new micro-foundation approaches, joining numerical computation and analytics.

6.2 The Learning Master Equation

This section develops the learning phenomenology, described in Sect. 4 and implemented in the ABM model, at a mean-field level to infer aggregate dynamics of transitions over the space Λ. The aim is to model within-transitions over the space of behavioural rules conditioned on a state of financial fragility: $x_i(\lambda_h|\varsigma;t) \equiv X_h \rightarrow x_i(\lambda_k|\varsigma;t+1) \equiv X_k$.

Consider a sample firm randomly drawn from the $\varsigma \in \Sigma$ state (NSF or SF). The firm schedules output according to (1) by setting $\alpha(i,t)[\lambda]$ according to some behavioural rule $\lambda \in \Lambda$: consider the sample firm as a stereotyped or mean-field agent, i.e. a sub-system characterised by $(\lambda|\varsigma)$. To make the final choice, it faces $K = |\Lambda|$ rules to test: each scheduling parameter is due to a single behavioural rule and it leads to a given level of profit $\{\Pi(i,t)[\lambda] : \lambda \in \Lambda\}$. A measure of profitability is needed to mimic micro-decision making at macro-level, to compare rules' profitability and to specify a probability measure for each possible choice to be made.

Total profit of the sub-system $\varsigma \in \Sigma$ is $\Pi_+(\varsigma;\tau_k) = \sum_{\lambda\in\Lambda} |\Pi(\lambda|\varsigma;\tau_k)|$, where $\Pi(\lambda|\varsigma;\tau_k)$ is the aggregate profit of state $\varsigma \in \Sigma$ behaving as $\lambda \in \Lambda$. Accordingly, $u(\lambda|\varsigma;\tau_k) = \Pi(\lambda|\varsigma;\tau_k)/\Pi_+(\varsigma;\tau_k) := u_\lambda$ is the profit share associated to a given rule. To estimate the switching probability, the sub-system characterised by that rule is assumed to interact with all the others in a sequence of interactions $X_h + \{X_k\}$ each associated to a specific vector $\mathbf{u}_{hk} = (u_h, u_k)$, where u_h is the profit share of the so-called *effective agent* X_h and u_k is that of the so- called *virtual agent* X_k: in the sequence $X_h + \{X_k\}$ of interactions an initial effective agent is fixed while virtual ones progressively change while testing all the rules. Since $\mathbf{u}_{hk} = (u_h = ||\mathbf{u}_{hk}|| \cos\theta, u_k = ||\mathbf{u}_{hk}|| \sin\theta_{hk})$ then an isomorphic unit vector $\mathbf{v}_{hk} = \mathbf{u}_{hk}/||\mathbf{u}_{hk}|| = (v_h = \cos\theta_{hk}, v_k = \sin\theta_{hk})$ is associated on the unit circle, where v_h is the normalised profit of the effective agent and v_k is the normalised profit of the virtual agent in the specific interaction $X_h + X_k$. The specific interaction is now uniquely characterised by the interaction angle θ_{hk}. According to a clockwise rotation $(\vartheta = \theta - \pi/4)$ an equivalent interaction vector is found $\mathbf{V}_{hk} = (V_h = \cos\theta_{hk}, V_k = \sin\theta_{hk})$, and $V_h = \cos\theta_{hk} = (v_h - v_k)/\sqrt{2}$ defines the profitability indicator. It compares the effective and virtual profits on the unit circle and is employed in order to estimate the probability for the sample firm in state $(\lambda_h|\varsigma)$ to switch into state $(\lambda_k|\varsigma)$. This is the transition probability $\Pr\{X_h \rightarrow X_k\} = r_{hk|k}$ associated to the interaction represented by the stenographic equation[11] $X_h + X_k = 2X_k$:

[11] A stenographic equation is a writing drawn from stochastic kinetics of chemical reaction (among others see McQuarrie (1967), Nicolis and Prigogine (1977) to represent the interaction/reaction of reactants into products in a chemical reaction). These processes are usually described in terms of birth-and-death stochastic processes and this writing has been found very suitable to represent learning at mean-field level as switching-maintenance mechanism. Further details can be found in Landini et al. (2014).

$$r_{hk|k} = Z_h \int_{-\pi/2}^{\vartheta_{hk}} \frac{\cos\vartheta}{2} d\vartheta|_{\vartheta_{hk}=\theta_{hk}-\pi/4} = \frac{Z_h}{4}(1+\sin(\theta_{hk}-\pi/4)) = r(\theta_{hk})$$

$$\text{s.t. } Z_h = 4\left(K + \sum_{k\leq K} \sin(\theta_{hk}-\pi/4)\right)^{-1} \tag{22}$$

Its complement is the maintenance probability $\Pr\{X_h \to X_h\} = r_{hh|k}$ associated to $X_h + X_k = X_h + X_k$. Hence, a specific interaction is a Bernoulli event

$$X_h + X_k \to \begin{cases} 2X_k: & \Pr\{(\lambda_h|\varsigma) \to (\lambda_k|\varsigma)\} = r_{hk|k} \\ X_h + X_k: & \Pr\{(\lambda_h|\varsigma) \to (\lambda_h|\varsigma)\} = r_{hh|k} \end{cases} \tag{23}$$

Since all the virtual-rules have to be tested while learning, the sequence of interactions $X_h + \{X_k\}$ determines two vectors, one for switching and the other for maintenance probabilities for each specific interaction. Changing the effective agent two other vectors are found. Once all the effective agents have been made interacting with all the virtual agents, two matrices are quantified at each point in time: $\mathbf{W}^s = \{r_{hk|k} : h, k \leq K\}$ and $\mathbf{W}^m = \{r_{hh|k} : h, k \leq K\}$.

The aim is to evaluate the probability for an effective agent— $x_i(\lambda_h|\varsigma; t) \equiv X_h$—becoming a different agent—$x_i(\lambda_k|\varsigma; t+1) \equiv X_k$—one period ahead. While learning, i.e. along the sequence $X_h + \{X_k\}$ of interactions, the sample effective agent can temporarily switch or maintain its own rule in testing all the single rules step-by-step, that is, starting as $(\lambda_p|\varsigma)$ it follows a learning path on to end up with its final choice into being $(\lambda_q|\varsigma)$. Landini et al. (2014) develops in details this problem and obtains analytic formulae to compute such probabilities[12] in order to define a time-dependent transition matrix $\mathbf{W}_\varsigma(t+1) = \{w_\varsigma(h,k;t+1):h,k \leq K\}$ where $w_\varsigma(h,k;t+1) = \Pr\{x_i((\lambda_h|\varsigma);t) \equiv X_h \to x_i((\lambda_k|\varsigma);t+1) \equiv X_k\}$. The matrix $\mathbf{W}_\varsigma(t+1)$ provides local dynamic transition rates from $(\lambda_h|\varsigma)$ to $(\lambda_k|\varsigma)$.

What happens at $t+1$ is determined by what happens at t, therefore the transition mechanism is consistent with a Markov scheme. Accordingly, by defining $z_\varsigma(h,t) = \sum_{k\neq h} w_\varsigma(h,k,t)$, the holding time τ_k is exponentially distributed, $\Pr\{\tau_h < \Delta\} = 1 - \exp(-z_\varsigma(h,t)\Delta) = z_\varsigma(h,t)\Delta + o(\Delta)$, hence

$$\begin{cases} \Pr\{x_i(\lambda_k|\varsigma) = X_k, t+\Delta | x_i(\lambda_h|\varsigma) = X_h, t\} = w_\varsigma(h,k,t)\Delta + o(\Delta) \\ \Pr\{x_i(\lambda_h|\varsigma) = X_h, t+\Delta | x_i(\lambda_h|\varsigma) = X_h, t\} = 1 - z_\varsigma(h,t)\Delta + o(\Delta) \end{cases} \tag{24}$$

The probability of finding a sample agent in state $(\lambda_k|\varsigma)$ at time $t > 0$ with initial condition at $(\lambda_h|\varsigma)$ is $h_\varsigma(h,k,t) = \Pr\{x_i(\lambda_k|\varsigma) = X_k, t | x_i(\lambda_h|\varsigma) = X_h, 0\}$. According to Feller (1966), a dynamic model obeys a Kolmogorov equation

[12] The interested reader is invited to follow Sect. 4 and Appendix C in Landini et al. (2014), for which the present section is a synthesis of results.

$$\frac{dh_{\varsigma}(h,k,t)}{dt} = -z_{\varsigma}(h,t)h_{\varsigma}(h,k,t) + \sum_{k \neq h} w_{\varsigma}(k,h,t)h_{\varsigma}(h,k,t) \tag{25}$$

which is a master equation representation (see Aoki (1996)) whose local transition rates have the following generator square matrix:

$$\mathbf{G}_{\varsigma}(t) = \{g_{\varsigma}(h,k,t)\}: g_{\varsigma}(h,k,t) = \begin{cases} -z_{\varsigma}(h,t) = -\sum_{k \neq h} w_{\varsigma}(h,k,t): h = k \\ w_{\varsigma}(h,k,t): h \neq k \end{cases} \tag{26}$$

allowing for a matrix representation of (25)

$$\dot{\mathbf{H}}_{\varsigma}(t) = \mathbf{H}_{\varsigma}(t)\mathbf{G}_{\varsigma}(t): \mathbf{H}_{\varsigma}(0) = \mathbf{H}_{\varsigma}^{0} \tag{27}$$

where $\mathbf{H}_{\varsigma}(t) = \{h_{\varsigma}(h,k,t)\}$. ABM-DGP data make possible the estimation of the time series of $\mathbf{W}_{\varsigma}(t)$ and $\mathbf{G}_{\varsigma}(t)$. The matrix $\mathbf{G}_{\varsigma}(t)$ is very stable about the time average, therefore $\lim_{T \to \infty} \sum_{t \leq T} \mathbf{G}_{\varsigma}(t)/T = \hat{\mathbf{G}}_{\varsigma}$ (see Fig. 6) gives

$$\dot{\mathbf{H}}_{\varsigma}(t) = \mathbf{H}_{\varsigma}(t)\hat{\mathbf{G}}_{\varsigma}: \mathbf{H}_{\varsigma}(0) = \mathbf{H}_{\varsigma}^{0} \Rightarrow \hat{\mathbf{H}}_{\varsigma}(t) = \mathbf{H}_{\varsigma}^{0}.\exp(t.\hat{\mathbf{G}}_{\varsigma}) \tag{28}$$

Assume then the economy is large enough to state the asymptotic convergence

$$p(\lambda_h|\varsigma,t) = \frac{n_h}{\hat{S}_{\varsigma}(t)} \longrightarrow^{I \to \infty} \lim_{\Delta \to 0} \Pr\{I(\lambda_h, t+\Delta|\varsigma)\} \; \forall h \leq K \tag{29}$$

where n_h is the h-th species concentration in the configuration vector $\mathbf{I}_{\varsigma}(t) = \{I(\lambda_h,t|\varsigma) = n_h{:}\lambda_h \in \Lambda\} = \mathbf{n} = (n_1,\ldots,n_h,\ldots,n_K)$ of the sub-system in the state $\varsigma \in \Sigma$ of financial soundness and $\hat{S}_{\varsigma}(t)$ is the solution of the financial fragility master equation obtained by substituting (20) and (21) into (17). Accordingly, (29) nests[13] the previous master equation into the solution of the following combinatorial one.

The dynamic equation for the state probability in (29) is defined as follows:

$$\begin{aligned} &\hat{\mathbf{p}}_{\varsigma}(t) = \hat{\mathbf{H}}_{\varsigma}(t).\mathbf{p}_{\varsigma}^{0} \; s.t. \\ &\mathbf{p}_{\varsigma}^{0} = \mathbf{n}_0/\hat{S}_{\varsigma}(0); \; \mathbf{I}(0,\varsigma) = \mathbf{n}_0 = (n_{1,0},\ldots,n_{h,0},\ldots,n_{K,0})' \end{aligned} \tag{30}$$

[13] This nesting approach moves a step further from Landini et al. (2014), which involves the total number of NSF and SF firms as aggregate from the ABM-DGP. By substituting the solution macroeconomic equation (20) into the ansatz (17), together with volatility from (21), both derived from the solution of the master equation (16) for the distribution of agents on the space Σ, the analytic estimators for the expected value of NSF/SF firms is at hand. By plugging this into the estimator for expected concentrations of species on the space Λ conditioned on Σ, as found by means of the Poisson representation in the following (35), the result is an analytic-endogenous nested model for the distribution of agents on the space $\Xi = \Lambda \times \Sigma$.

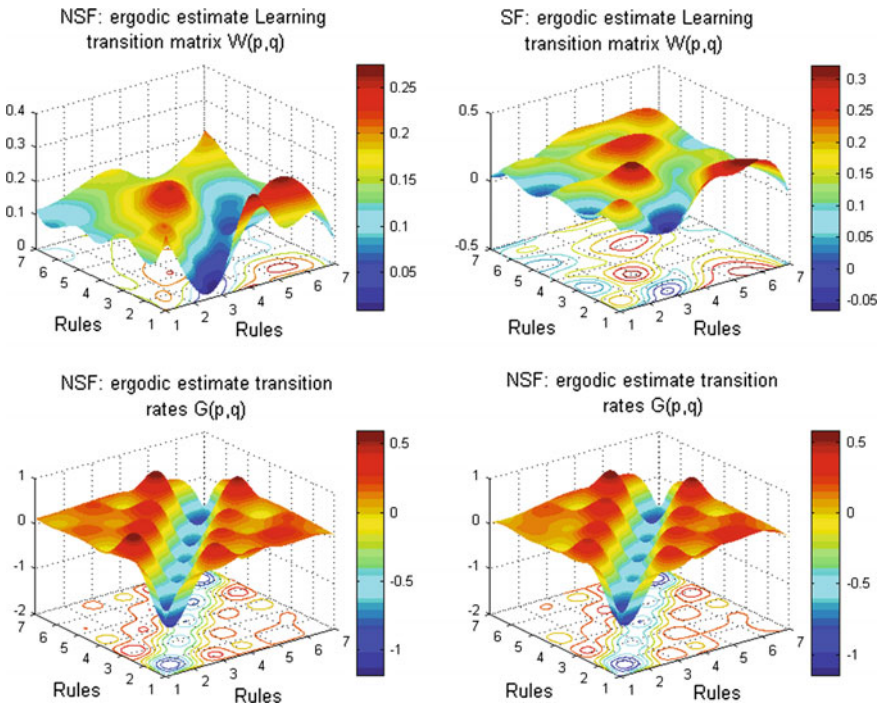

Fig. 6 Estimates of the time averages for matrices $\mathbf{W}_{\varsigma}(t)$ and $\mathbf{G}_{\varsigma}(t)$

hence

$$\hat{\mathbf{m}}_{\varsigma}(t) = \hat{\mathbf{p}}_{\varsigma}(t).\hat{S}_{\varsigma}(t) = \mathbf{H}^{0}_{\varsigma}.\exp(t.\hat{\mathbf{G}}_{\varsigma}).\mathbf{p}^{0}_{\varsigma}(\varsigma).\hat{S}_{\varsigma}(t) \tag{31}$$

defines the expected concentration of firms in each behavioural state $\lambda \in \Lambda$ conditioned on the state $\varsigma \in \Sigma$ financial soundness. In (Fig. 7) an aggregation of (31) as regarding the NSF state is compared with the outcome of the aggregation from the ABM simulation as given by (15) and with the outcome of (17) for the first ME by using (20) and (21).

The trajectory (31) describes the expected changes on the rules space within the sub-system characterised by $\varsigma \in \Sigma$. Landini et al. (2014), by developing a combinatorial master equation, find that $\hat{\mathbf{m}}_{\varsigma}(t)$ is the estimator of the expected value of a K-dim Poisson distribution as the solution of the combinatorial master equation. It accounts for all the possible interactions to which the different species can give rise, that is,

$$\sum_{k \leq K} s^{c}_{k} X_{k} \longleftrightarrow^{H^{+}_{c}}_{H^{-}_{c}} \sum_{k \leq K} r^{c}_{k} X_{k} \text{ or } \mathbf{s}^{c}.\mathbf{L} \longleftrightarrow \mathbf{r}^{c}.\mathbf{L} \; \forall c \leq C \tag{32}$$

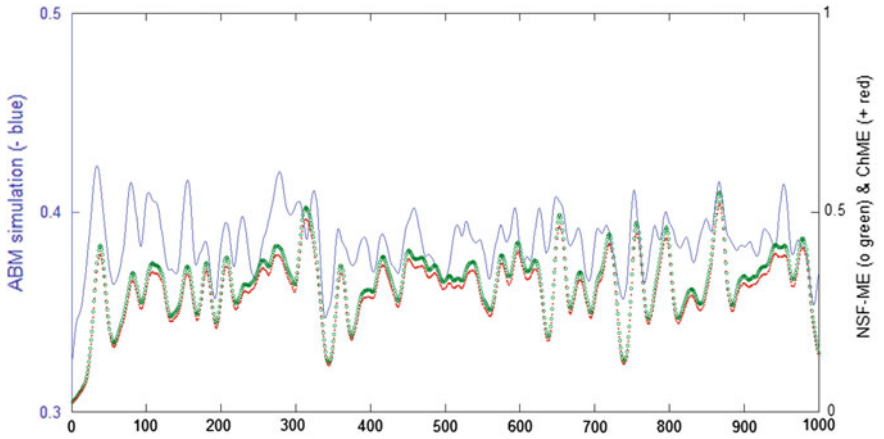

Fig. 7 ABM concentration of NSF firms and aggregate expected values according to the nested model of master equations

where $\mathbf{s}^c = (s_1^c, \ldots, s_k^c, \ldots, s_K^c)$ and $\mathbf{r}^c = (r_1^c, \ldots, r_k^c, \ldots, r_K^c)$ are stoichiometric vectors and $\mathbf{X} = (X_1, \ldots, X_k, \ldots, X_K)$ being s_k^c and r_k^c portions of the X_k-concentration, i.e. portions of the n_k agents, activated as reactants and products in the c-th interaction; the $H_k^{\pm}$ are called rate constants and they represent direct interaction "$\longrightarrow$", allowing for "switching", and an inverse interaction "$\longleftarrow$", allowing for "maintenance", in the frame of the learning mechanism developed so far.

The master equation for (32) is

$$\frac{\partial P(\mathbf{n}, t)}{\partial t} = \sum_{k \le K} \sum_{\vartheta = 0,1} (2\vartheta - 1)[T_k^-(\mathbf{n} + \vartheta \mathbf{u}^k; t) P(\mathbf{n} + \vartheta \mathbf{u}^k, t) + T_k^+(\mathbf{n} - \vartheta \mathbf{u}^k; t) P(\mathbf{n} - \vartheta \mathbf{u}^k, t)] \tag{33}$$

where $\vartheta = 0$ means outflows and $\vartheta = 1$ inflows, $\mathbf{u}^c = \mathbf{r}^c - \mathbf{s}^c$ determines forward jumps $\mathbf{n} \longrightarrow \mathbf{n} + \mathbf{u}^c$ and backward ones $\mathbf{n} \longrightarrow \mathbf{n} - \mathbf{u}^c$. Transition rates are

$$\begin{aligned} T_k^+(\mathbf{n} - \vartheta \mathbf{u}^k; t) &= H_k^+ \prod_{h \le K} \frac{(n_h - \vartheta u_h^k)!}{(n_h - \vartheta u_h^k - s_h^k)!} \\ T_k^-(\mathbf{n} + \vartheta \mathbf{u}^k; t) &= H_k^- \prod_{h \le K} \frac{(n_h + \vartheta u_h^k)!}{(n_h + \vartheta u_h^k - r_h^k)!} \end{aligned} \tag{34}$$

They account for the combinatorial nature of the possible interactions the different species can set up. Gardiner and Chaturvedi (1977) shows that the stationary solution for this combinatorial master equation is (Fig. 8)

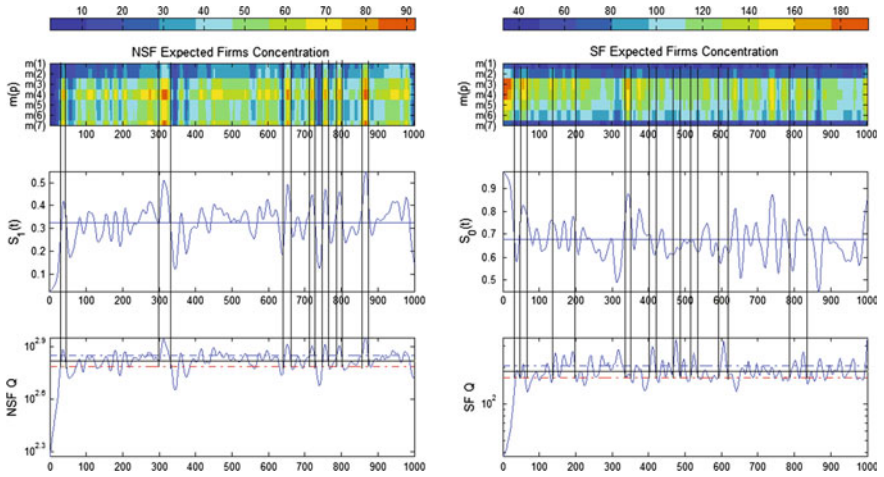

Fig. 8 Expected configuration of the system, dynamics of NSF and SF shares estimated by the master equation and output

Table 2 Expected concentrations time average and standard deviation in behavioural states

Time average (standard deviation)	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	Total
NSF	9.51 (8.67)	10.16 (9.49)	15.41 (12.56)	18.36 (14.01)	14.75 (11.39)	14.76 (11.53)	17.05 (13.40)	100
SF	7.67 (5.89)	11.27 (9.49)	19.40 (17.75)	19.25 (15.97)	17.84 (13.84)	15.96 (11.99)	8.61 (6.45)	100

$$P_e(\mathbf{n}) = \prod_{h \leq K} e^{-m_h} \frac{m_h^{n_h}}{n_h!} : m_h =< n_h > \tag{35}$$

Table 2 shows the dominance of both among NSF and SF firms. In case of NSF it is evident that $\lambda_4 \succ \lambda_7 \succ \lambda_3 \succ \lambda_6 \approx \lambda_5 \succ \lambda_2 \succ \lambda_1$ while for SF it follows that $\lambda_3 \approx \lambda_4 \succ \lambda_5 \succ \lambda_6 \succ \lambda_2 \succ \lambda_7 \succ \lambda_1$. In both cases, the elementary rule is expected to be the least preferred one. It is worth noting that NSF firms are likely to be average followers, while SF firms seldom use this rule. In both states of financial soundness, micro-imitation, due to direct interaction, is one of the mostly expected behaviours: on expectation it is the most diffused rule for NSF and the second choice for SF. SF firms are expected to be profit-maximisers or micro-imitators while NSF are expected to be micro-imitators or average followers. This is interesting because micro-imitation is concerned with direct interaction at micro-level while following the average implies setting the scheduling parameter at the system level, that is the most extensive kind of interaction: in a sense, these are the *extrema* in Table 1.

This is the prediction of the mean-field nested model. For explanation, one should consider the dynamics of these expected values compared with some system observable, such as output considered as the control variable: indeed, output is under the control of firms in learning how to set their output scheduling parameter.

Given the results detailed in Sect. 5 for output and profit for the different types of firms, one can conclude that NSF firms are significantly more active and reactive than the SF ones. In other words, their vital impulse in surviving and improving their financial soundness is stronger than the one of SF firms, which seems to behave much more prudentially in order to not harm their condition. More accurately, it should be noted that, due to the positive equity constraint in profit maximisation, NSF firms are closer to a lower bound rather than SF ones, which rationally use this rule as first choice. Hence, when the standard assumption of rational profit maximisation is not feasible because it is too costly, such firms reasonably choose other solutions due to the bankruptcy risk. Nonetheless, NSF firms do not completely exclude profit maximisation; indeed they are expected to involve it as a third choice.

Figure 8 shows the expected concentrations on the rules in the two states of financial soundness (top panels), the financial soundness trajectory (middle panels) and output dynamics in states of financial soundness (bottom panels). As regarding the NSF sub-system, the left column of panels in Fig. 8 demonstrates that, for most of the time, the output is within its confidence bands although with some fluctuations. When the time series is above the upper limit (see around $t = 300$) the concentration of NSF share of firms shows some peaks and it transmits to a higher concentration of firms in the first two dominant rules, which are λ_4 for micro-imitators and λ_7 for average followers. In periods where the time series is below the lower limit, the concentration of firms is almost uniform: this means that during such periods some of the NSF jump to the SF state (see around $t = 350$). The expected improvement for the system is not so evident. Even though it is clear that SF firms increase in volume and that they concentrate on the SF dominant rule λ_3 of profit maximisation, the value of output is below the SF mean reference and within the bands.

This is because, as noted above, SF firms have a lower output performance due to their prudential attitude and, moreover, the scheduling parameter of SF profit maximisers is lower than the NSF counterpart. It cannot be therefore ascertained that the system performs better with a higher concentration of SF firms (note that around $t = 350$ there is more than 80 % of SF but the same observation also holds around $t = 75$ or, less evidently, in other points in time). It can be concluded that the learning activity of social atoms is important since it allows the financially *weaker* firms to survive and improve more than the *stronger* ones.

The vital impulse of NSF firms is stronger than the SF ones. Therefore, if the banking system meets the credit demand from NSF firms, the learning activity can improve the performance of the system despite the increase in financial fragility. Given the very simple banking system of this model the conclusions are not general but nevertheless make possible an assessment of the effects of learning on macro-dynamics.

Finally, there is evidence that the learning activity can be considered to be the cause of phase transitions characterising the system (*regenerative-coordination*).

Being imitation the most likely strategy, once some regularity stabilises the system into a certain phase there can happen critical events pushing the more active firms to change their behaviour and destroying that regularity they contributed to define, suggesting other firms to follow them through direct interaction.

7 Concluding Remarks

The chapter introduces an original inferential methodology, providing some theoretical results. This work can be a step towards applying complex system theory for a new micro-foundation of macro models when heterogeneity and interaction are present. From the analytic point of view, some results are applicable for the study of financial fragility. In this respect, it has been found that the more financially fragile firms are the most active ones as they are pushed by the vital impulse to survive and improve. On the contrary, financially robust firms behave in a very prudential way. In particular, being the most active ones in learning, financially fragile firms are those contributing more to growth. Their behaviours determine periods of expansion when their financial needs are accommodated by the credit supply. Accordingly, a higher share of financially fragile firms does not always compromise the system stability. On the other hand, when almost all firms are financially robust, the outcome is sub-optimal. Indeed, in these periods growth is lower than during periods with high financial fragility.

As an additional activity to the standard management of the firm, learning improves the system performance because it offers more tools than the standard profit maximisation perspective, which is rarely the dominant behaviour. Social atoms have ambitions and stimuli, they try different possible paths to reach their target avoiding failures by reasonably making decisions.

The macroscopic dynamics of the system is not predictable by only looking at micro-behaviours. This is because, as typical for complex systems, aggregate emergent phenomena occur at system level without a direct relationship with micro-events, as in the case of “regenerative-coordination”, which manifests in the cyclical succession of regimes and phase transitions.

There are still open questions to answer in this research stream. Many improvements are possible, either in modelling the learning activity or in the model for approaching a more realistic or empirically based set-up. The main aim of the present study is to set a theoretical and methodological basis to develop a new approach to complex systems with “social atoms”.

Acknowledgments The authors thank an anonymous referee for suggestions and remarks; the discussants and the participants to the EEA conference in NYC in May 2013, PRIN Bologna, June 2013, for suggestions. The financial support of the Institute for New Economic Thinking grant INO1200022, EFP7, MATHEMACS and NESS are gratefully acknowledged.

References

Aoki, M. (1996). *New approaches to macroeconomic modeling*. New York: Cambridge University Press.

Aoki, M. (2002). *Modeling aggregate behavior and fluctuations in economics*. New York: Cambridge University Press.

Aoki, M., & Yoshikawa, H. (2006). *Reconstructing macroeconomics*. Cambridge: Cambridge University Press.

Blume, L. E., & Easley, D. (1993). Rational expectations and rational learning. RePEc: wps:wuwpga:9307003.

Buchanan, M. (2007). *The social atom: Why the rich get richer, cheaters get caught, and your neighbor usually looks like you*. New York: Bloomsbury.

Chiarella, C., & He, X. (2002). Heterogeneous beliefs, risk and learning in a simple asset pricing model. *Computational Economics*, *19*(1), 95–132.

Delli Gatti, D., Di Guilmi, C., Gallegati, M., & Landini, S. (2012). Reconstructing aggregate dynamics in heterogeneous agents models: A markovian approach. in J. L. Gaffard & M. Napoletano (Eds.), *Agent-based models and economic policy, Revue de l'OFCE* (Vol. 124, pp. 117–146).

Delli Gatti, D., Gallegati, M., Greenwald, B., Russo, A., & Stiglitz, J. E. (2010). The financial accelerator in an evolving credit network. *Journal of Economic Dynamics and Control*, *34*, 1627–1650.

Di Guilmi, C. (2008). *The generation of business fluctuations: Financial fragility and mean-field interactions*. Frankfurt: Peter Lang.

Di Guilmi, C., Gallegati, M., & Landini, S. (2010). Financial fragility, meanfield interaction and macroeconomic dynamics: A stochastic model. in N. Salvadori (Ed.), *Institutional and social dynamics of growth and distribution* (pp. 323–351). Cheltenham: Edward Elgar Publising.

Di Guilmi, C., Gallegati, M., & Landini, S. (2012). Towards an analytical solution for agent based models: An application to a credit network economy. in M. Aoki, K. Binmore, S. Deakin, & Gintis H. (Eds.), *Complexity and institutions: Markets, norms and corporations* (Vol. 150, pp. 61–78). London: Palgrave Macmillan, IEA Conference.

Di Guilmi, C., Gallegati, M., Landini, S., & Stiglitz, J. E. (2013). *Dynamic aggregation of heterogeneous interacting agents and network: An analytical approach for agent based models*. New York: Mimeo.

Feller, W. (1966). *An introduction to probability theory and its applications* (Vol. I). New York: Wiley.

Gaffeo, E., Delli Gatti, D., Desiderio, S., & Gallegati, M. (2008). Adaptive micro-foundations for emergent macroeconomics. *Eastern Economic Journal*, *34*(4), 441–463.

Gardiner, C. W., & Chaturvedi, S. (1977). The Poisson representation I. A new technique for chemical master equations. *Journal of Statistical Physics*, *17*(6), 429–468.

Gardiner, C.W. (1985). *Handbook of stochastic methods*. Berlin: Springer.

Giansante, S., Chiarella, C., Sordi, S., & Vercelli, A. (2012). Structural contagion and vulnerability to unexpected liquidity shortfall. *Journal of Economic Behavior and Organization*, *83*(3), 558–569.

Gintis, H. (2007). The dynamics of general equilibrium. *Economic Journal*, *117*, 1289–1309.

Gintis, H. (2013). Markov models of social dynamics: Theory and applications. *ACM Transactions on Intelligent Systems and Technology*, *4*(3), 53:1–53:19.

Greenwald, B., & Stiglitz, J. E. (1993). Financial markets imperfections and business cycles. *Quarterly Journal of Economics*, *108*(1), 77–114.

Hartley, J. E. (1997). *The representative agent in macroeconomics*. London: Routledge.

Kirman, A. (2012). *Can artificial economies help us understand real economies. Debates and policies: Revue de l'OFCE*, *124*(5), 15–41.

Landini, S., & Uberti, M. (2008). A statistical mechanic view of macro-dynamics in economics. *Computational Economics*, *32*(1), 121–146.

Landini, S., Gallegati, M., & Stiglitz, J. E. (2014). Economies with heterogeneous interacting learning agents. *Journal of Economic Interaction and Coordination*, 1–28. DOI:10.1007/s11403-013-0121-1.

LeBaron, B. (2002). Building the Santa Fe artificial stock market. Working Paper, Brandeis University.

Lengnick, M. (2013). Agent-based macroeconomics: A baseline model. *Journal of Economic Behavior and Organization*, *86*, 102–120.

Markose, S. M. (2005). Computability and evolutionary complexity: Markets as complex adaptive systems (CAS). *Economic Journal, Royal Economic Society*, *115*(504), F159–F192.

McQuarrie, D. A. (1967). Stochastic approach to chemical kinetics. *Journal of Applied Probability*, *4*(3), 413–478.

Nicolis, G., & Prigogine, I. (1977). *Self-organization in non-equilibrium systems*. New York: Wiley.

Risken, H. (1989). *The Fokker-Planck equation: Methods of solution and applications*. Berlin: Springer.

Schweitzer, F. (2003). *Brownian agents and active particles. Collective dynamics in the natural and social sciences*. Berlin: Springer.

Tesfatsion, L. (2003). Agent-based computational economics: Modeling economies as complex adaptive systems. *Information Sciences*, *149*(4), 262–268.

Tesfatsion, L., & Judd, K.L. (Ed.) (2006). *Handbook of computational economics*. Amsterdam: North-Holland Company.

van Kampen, N. G. (2007). *Stochastic processes in physics and chemistry*. Amsterdam: North-Holland.

Vriend, N. J. (2000). An illustration of the essential difference between individual and social learning, and its consequences for computational analyses. *Journal of Economic Dynamics and Control*, *24*, 1–19.

How Non-normal Is US Output?

Reiner Franke

1 Introduction

Macroeconomic variables or the shock processes that drive them are commonly considered to be largely compatible with normal distributions. This supposition is not just a semantic issue, for example in assessing whether or not our economic system, occasional crises notwithstanding, is mainly well-behaved and the risk from the random perturbations in it can be satisfactorily controlled. It is also an important question in the academic field, where the flourishing business of the estimation of DSGE models with its present predominance of likelihood techniques heavily rests on the assumption of normally distributed innovations.

There have nevertheless been always some doubts about the prevalent view that takes normality for granted, in the first instance with respect to output and employment variables (as far as the empirical side is concerned, the following will refer to US data). Independently of discussions among heterodox theorists (like Blatt 1983), concerns about asymmetry were already raised some 30 years ago. The possibly best-known contribution at that time was Delong and Summers (1986) who, however, did not get much evidence for this. More recently, the tests by Bai and Ng (2005) failed to establish significant support of skewness or excess kurtosis. In contrast to these negative results, Christiano (2007) obtained significant excess kurtosis in the residuals of an unconstrained estimated vector autoregression. Interestingly, on the other hand, this did not prove sufficient to distort the Bayesian analyses that use the normal likelihood.

Other work finds indications of non-normality, too, and tries to take them into account in their theoretical models. De Grauwe (2012) concentrates on the kurtosis of output and emphasizes that, despite the normal shock processes, the nonlinear

Comments by an anonymous referee are gratefully acknowledged.

R. Franke (✉)
University of Kiel (GER), Kiel, Germany
e-mail: franke@uni-bremen.de

R. Dieci et al. (eds.), *Nonlinear Economic Dynamics and Financial Modelling*,
DOI: 10.1007/978-3-319-07470-2_9,

mechanisms in his model are able to reproduce this feature with great success.[1] While De Grauwe's discussion remains more informal, Ruge-Murcia (2012) rigorously estimates the third-order approximation of a DSGE model by the simulated method of moments, an approach that need not rely on normality assumptions. Within this framework he can reject the null that productivity innovations are normally distributed in favour of the alternative that they follow an asymmetric Skew normal distribution.[2]

Besides studying skewness and kurtosis, there is another approach that assesses deviations of the empirical distribution in its entirety from normality. To this end, the analysis refers to the class of the exponential power (or Subbotin) distributions. Their general shape is governed by a parameter b, where $b = 2$ yields the normal distribution and lower values give rise to progressively fatter tails. An alternative benchmark is then $b = 1$, at which the Laplace distribution prevails.[3] Several papers by Fagiolo et al. (2007, 2008, 2009) claim it a universal phenomenon for output growth rates that estimates of b are so low that normality has to be rejected. Ascari et al. (2012) even raise this into the category of a stylized fact, i.e., a standard that macroeconomic models should seek to meet. Checking this with calibrated versions of a Real Business Cycle and a New-Keynesian model, they find out that the former can replicate this type of fat tails exogenously but not endogenously, and the latter neither endogenously nor exogenously. Thus, this work raises a serious criticism against the current practice of DSGE modelling.

The present paper sets in with the observation for US quarterly output data that all of the studies revealing non-normality are based on samples covering 40 years and more. While such a sample size is certainly desirable from an econometric point of view, so many things have changed historically over this span of time that also the economy may be suspected of having shifted from one regime to another. Somewhat strangely, this possible problem is not touched upon in any of the studies. After all, the distinction between the two periods of the Great Inflation (GI) and the Great Moderation is a well-known topic in macroeconomics. In addition and more specifically, McConnell and Perez-Quiros (2000) provided firm statistical evidence of a structural break in the volatility of output growth around 1984.[4]

An obvious question thus arising is whether the features of non-normality can also be detected in the two subsamples before and after the structural break, even though with weaker rejections of normality because of the smaller samples. Regarding the long sample period, one may furthermore ask if the non-normality, if it is found to prevail at all, is perhaps spurious in the sense that it could be alternatively explained

[1] The strong nonlinearities originate with the agents' switching between different types of boundedly rational expectations, which he puts forward as a pronounced alternative to DSGE modelling.

[2] The moments themselves on which the model is estimated include the third-order moments of hours worked and of the consumption and investment growth rates; see Ruge-Murcia (2012, p. 931, Table 10).

[3] It plays a prominent role in cross-sectional distributions of firm characteristics such as growth rates and profit rates; see, e.g., Alfarano et al. (2012).

[4] Which from a theoretical point of view might also be due to multiple equilibria.

by the pooling of two samples, which are characterized by two normal distributions with—according to the structural break—different dispersion. Apart from this, it may be checked whether the different test statistics used in the aforementioned literature all come to similar conclusions. In short, it is time that the issue of non-normality, even if confined to the levels and growth rates of US output, was subjected to a careful reconsideration. This is the purpose of the present paper.

The investigation is organized as follows. Section 2 presents five test statistics to detect non-normality. On the one hand, these are the popular Jarque-Bera test and two generalizations that take account of the serial correlation in the data. The other two tests are concerned with the overall shape of the distributions, namely, the Anderson-Darling test and the shape parameter of the exponential power distribution, which has already been mentioned. The remaining sections deal with the quarterly data of US output, where we focus on the growth rates of GDP and the output of the firm sector after the often employed CBO output gap will have been found to offer less prospects for non-normality.[5] Three different samples are moreover considered: the full sample of 47 years from 1960 on, and its decomposition into two almost equally long subsamples.

Section 3 applies the five tests to the empirical data and to the residuals from suitable autoregressions. While the conclusions drawn here are based on asymptotic theory, Section 4 uses Monte Carlo (MC) experiments to learn more about the small sample properties. Section 5 takes up the idea from above and indeed finds out that the non-normality results in the full sample could be satisfactorily explained by two estimated AR(p) processes over the two subsamples, the two normally distributed innovations of which have distinctly different variances.

As we will learn that the only feature of possible non-normality in a subsample is a low estimate of the shape parameter of the exponential power distribution, Sect. 6 puts forward the alternative hypothesis that the growth rates follow a Laplace distribution. This gives rise to the most pronounced result of our work: the Laplace can be safely rejected in favour of normality in one subsample, and it cannot be rejected in the other.

Lastly, Sect. 7 is concerned with the precision with which the shape parameter can be estimated. This issue is of particular relevance for models that may have the ambition to reproduce this aspect of normality or non-normality. Section 8 concludes, and an appendix contains several technical details.

2 Test Statistics to Detect Non-normality

There are various fields in applied macroeconomic research where it is of interest whether a given realization of a stochastic process could have been obtained from a normal distribution. The most common approach to checking the normality of the marginal distribution of the data are procedures that test whether the third and

[5] The CBO output gap is based on the Congressional Budget Office's estimates of potential output. See the appendix for the data source.

fourth moments coincide with those of the normal distribution.[6] Accordingly, let a stationary univariate time series $\{x_t\}_{t=1}^T$ of length T be given with mean $\bar{x}$ and estimated standard deviation $\hat{\sigma}$. Its skewness S and kurtosis K are estimated as

$$\begin{aligned} \widehat{S} &= \hat{\mu}_3 / \hat{\sigma}^3 \\ \widehat{K} &= \hat{\mu}_4 / \hat{\sigma}^4 \\ \hat{\mu}_k &= (1/T) \sum_{t=1}^{T} (x_t - \bar{x})^k \,, \quad k = 2, 3, 4 \qquad (\text{hence } \hat{\mu}_2 = \hat{\sigma}^2) \end{aligned} \tag{1}$$

The normal distribution has $S = 0$ and $K = 3$. When $\widehat{K}$ is "sufficiently" larger than 3, the distribution of x_t is said to exhibit excess kurtosis, or to have fat tails. It is also well-known that reliable estimates of the kurtosis require a fairly large number of observations, larger than the typical sample size of macroeconomic quarterly data. For this reason and in order to limit the discussion, we focus on tests of the joint hypothesis $S = 0$ and $K = 3$. Because of its simplicity the probably most popular test statistic is that of Jarque and Bera (1980),

$$JB = T \Big[\frac{\widehat{S}^2}{6} + \frac{(\widehat{K} - 3)^2}{24} \Big] = T \Big[\frac{\widehat{\mu}_3^2}{6\,\hat{\mu}_2^3} + \frac{(\hat{\mu}_4 - 3\,\hat{\mu}_2^2)^2}{24\,\hat{\mu}_2^4} \Big] \tag{2}$$

(the second expression is added for a better comparison with the generalised statistic below). If the random variable x_t is iid and normally distributed, JB is χ^2-distributed with two degrees of freedom. Thus, normality will be rejected at a 5 % significance level if JB exceeds 5.99. (Throughout the paper, "significance" statements will be based on this level.)

There are two problems with this straightforward rule. First, the small-sample properties of the test are different from the asymptotic result. Second and more seriously, most macroeconomic time series data violate the prerequisite of iid. In the presence of serial correlation, however, the true asymptotic variances of S and $(K - 3)$ are no longer consistently estimated by the denominators of JB, which implies that even asymptotically the rejection probabilities deviate from the desired nominal levels.

We consider two approaches to remedy these distortions, both of which require no deep assumptions on the true data generation process.[7] The first approach, which is borrowed from Lobato and Velasco (2004), modifies the variances 6 and 24 in (2) directly by taking the autocovariances of the series into account. The main correction

[6] There are nevertheless other probability distributions with the same two moments; see distribution S4 in Bai and Ng (2005, p. 60), the skewness and kurtosis of which are reported in the first four tables of this paper.

[7] Vavra and Psaradakis (2011) provide a third and more ambitious generalisation of JB, which is based on smoothed quantiles and also uses more robust measures of the skewness and kurtosis. There is nevertheless no simple rule of thumb for the optimal specification of these quantiles (p. 14), which is the reason why we better wait for additional experience with this method.

terms are here $F^{(3)}$ and $F^{(4)}$ for the skewness and kurtosis statistics, respectively, defined as $F^{(k)} = \sum_{j=-\infty}^{\infty} \gamma(j)^k$ with respect to the population autocovariances $\gamma(j)$ of order j and $k = 3, 4$. For finite samples these sums can be estimated as

$$\begin{aligned} \widehat{F}^{(k)} &= \sum_{j=1-T}^{T-1} \hat{\gamma}(j)\,[\,\hat{\gamma}(j) + \hat{\gamma}(T-|j|)\,]^{k-1} \\ \hat{\gamma}(j) &= (1/T) \sum_{t=1}^{T-|j|} (x_t - \bar{x})\,(x_{t+|j|} - \bar{x}) \end{aligned} \tag{3}$$

($\hat{\gamma}(T)$ is set equal to zero). Lobato and Velasco (2004, p. 676) establish that asymptotically, for weakly dependent processes and under the null hypothesis of normality, their generalized Jarque-Bera statistic (*GJB*) is again $\chi^2(2)$ distributed,[8]

$$GJB = T\left[\frac{\hat{\mu}_3^2}{6\,\widehat{F}^{(3)}} + \frac{(\hat{\mu}_4 - 3\,\hat{\mu}_2^2)^2}{24\,\widehat{F}^{(4)}}\right] \xrightarrow{d} \chi^2(2) \tag{4}$$

This specification is indeed meaningful since $\widehat{F}^{(3)}$ and $\widehat{F}^{(4)}$ are ensured to be positive (p. 678). From a comparison of (2) and (4) it is furthermore easily seen that asymptotically *GJB* reduces to *JB* if the stochastic process is iid, since in this case $\hat{\gamma}(j) \to 0$ for all $j \neq 0$ in (3) and $\hat{\gamma}(0) = \hat{\sigma}^2 = \hat{\mu}_2$. With positive serial correlation in the first few lags of a time series, however, the denominator in *GJB* will be larger than in *JB*, so that *GJB* will fall short of *JB* and the chances of rejecting normality would decrease.

Also a converse conclusion holds true: under certain regularity conditions (technically requiring finite moments up to the sixteenth order) *GJB* diverges to infinity if the null is violated, that is, if for the population moments $\mu_3 \neq 0$ or $\mu_4 \neq 3\,\mu_2^2$. Hence, normality will be rejected with a probability tending to one as $T \to \infty$.

The second approach to correct the asymptotic variance of *JB* was proposed by Bai and Ng (2005). This generalization is less closely related to *JB* than *GJB*. Rather, the testing procedure is more similar to a GMM test of overidentifying restrictions, though some subtle differences still remain. In detail, define

$$z_t = \begin{bmatrix} x_t - \bar{x} \\ (x_t - \bar{x})^2 - \hat{\sigma}^2 \\ (x_t - \bar{x})^3 \\ (x_t - \bar{x})^4 - 3\hat{\sigma}^4 \end{bmatrix}, \qquad \bar{z} = (1/T) \sum_{t=1}^{T} z_t$$

$$y_T = \begin{bmatrix} (1/\sqrt{T}) \sum_t (x_t - \bar{x})^3 \\ (1/\sqrt{T}) \sum_t [\,(x_t - \bar{x})^4 - 3\hat{\sigma}^4\,] \end{bmatrix}$$

$$\widehat{\alpha} = \begin{bmatrix} -3\hat{\sigma}^2 & 0 & 1 & 0 \\ 0 & -6\hat{\sigma}^2 & 0 & 1 \end{bmatrix}$$

[8] The square for $\hat{\mu}_2$ in the kurtosis expression is missing in the definition of their statistics *SK* and *G* (Lobato and Velasco, 2004, pp. 674f), which is a typo.

$$\widehat{\Phi} = \Gamma_0 + \sum_{j=1}^{p} (1 - \frac{j}{p+1}) (\Gamma_j + \Gamma_j') , \quad p = [T^{1/4}]$$

$$\Gamma_j = (1/T) \sum_{t=1}^{T-j} (z_t - \bar{z})(z_{t+j} - \bar{z})' , \quad j = 0, 1, \ldots , p$$

The (4×4) matrix $\widehat{\Phi}$ is a Newey-West estimator of the long-run covariance matrix of z_t that uses the linearly declining weights of the Bartlett kernel, where the maximal lag length $[T^{1/4}]$ (the smallest integer greater than or equal to $T^{1/4}$) is determined by a common rule of thumb (see Greene 2002, p. 267, fn 10).[9] On this basis, Bai and Ng (2005, p. 52) specify a statistic, which we denote *BN*, and (for any consistent estimator of the covariance matrix of z_t) demonstrate that its distribution converges to $\chi^2(2)$:

$$BN \quad = \quad y_T' \, (\widehat{\alpha}\, \widehat{\Phi}\, \widehat{\alpha}')^{-1} \, y_T \quad \xrightarrow{d} \quad \chi^2(2) \tag{5}$$

The tests so far were concerned with the values of the third and fourth moments as they would be implied by the normal distribution. We are now turning to two tests that seek to take account of the shape of the entire distribution. The first one is the Anderson-Darling test. It is based on an evaluation of the squared differences between the hypothesized and the empirical distribution, which however places more weight on the observations in the tails of the distribution. In this respect it follows a similar idea to the kurtosis. Generally, if $F = F(x)$ is the hypothesised distribution and $F_T = F_T(x)$ the empirical cumulative distribution function, their distance is measured as

$$T \int_{-\infty}^{\infty} \frac{[F_T(x) - F(x)]^2}{F(x)\,[1-F(x)]} \, dF(x)$$

To apply this concept to the normal distribution and finite samples, the data must first be transformed into the standardised values,

$$y_t \quad = \quad (x_t - \bar{x}) \,/\, \hat{\sigma} , \qquad t = 1, \ldots , T \tag{6}$$

and arranged in ascending order, $y_1 \le y_2 \le \cdots \le y_T$. These y_i enter the standard normal cumulative distribution function, $\Phi = \Phi(y)$, to set up the Anderson-Darling test statistic *AD*.[10] Together with a simple rule to reject normality at the 5 % significance level, the prescription reads,

[9] Lobato and Velasco (2004, p. 675) emphasise that their statistic (4) does not require the introduction of a kernel function or such a user-chosen number.

[10] It is well-known that $\Phi(\cdot)$ has no closed-form analytical expression. The details of its numerical approximation are given in the appendix.

$$\begin{aligned} A^2 &= -T - \tfrac{1}{T} \sum_{i=1}^{T} \{ (2i-1) \ln \Phi(y_i) + [2(T-i)+1] \ln[1-\Phi(y_i)] \} \\ AD &= A^2 \left[1 + \tfrac{4}{T} - \tfrac{25}{T^2} \right]; \qquad \text{normality rejected if } AD > 0.787 \end{aligned} \tag{7}$$

While the Anderson-Darling test statistic tells us when to reject normality, it does not indicate along which dimension the empirical distribution may differ most from normality. In this respect a parametric approach is more useful that includes the normal distribution as a benchmark case; in particular, when it also provides information about whether or to what extent the empirical distribution may exhibit fat-tail behaviour. A flexible statistical tool for this is the family of the exponential power (EP) densities, or the Subbotin density functions,

$$f(x; b, a, m) = \frac{1}{2a\, b^{1/2}\, \Gamma(1+1/b)} \exp\left\{ \frac{-1}{b} \left| \frac{x-m}{a} \right|^b \right\} \tag{8}$$

where $\Gamma(\cdot)$ is the Gamma function.[11] The three parameters identifying the EP distributions are the location parameter m, the scale parameter a, and the shape parameter b. The latter determines the fatness of the tails and is therefore of prime importance to us. The bell shape of the normal density arises with $b = 2$, while for $b \to \infty$ the distribution tends towards the uniform distribution with support $[-a, a]$. On the other hand, as b decreases from the Gaussian benchmark, the shoulders of the distribution become smaller and the tails become fatter. The benchmark of practical concern in this direction is $b = 1$, which, in a semi-log diagram, yields the tent shape of the Laplacian density (we will encounter it in Sect. 6, Fig. 1, further below).

Since b is a parameter characterising the global shape of the distribution, it can be expected that it will also be a more robust measure of the fatness of tails than the kurtosis. Theoretically, the kurtosis implied by an EP distribution is given by $K = K(b) = \Gamma(1/b)\, \Gamma(5/b) \,/\, [\Gamma(3/b)]^2$ (Chiodi 1995, Sect. 2). Even if one prefers to refer to the kurtosis as a more familiar measure of fatness, this relationship may be used as a check of the direct empirical calculation of K. To get a first impression of its order of magnitude, *vis-à-vis* $K = 3$ for the normal distribution, the kurtosis of the Laplace distribution rises to $K = K(1) = 6$.

There are several likelihood methods to estimate the parameters of an EP distribution.[12] More convenient for us is a moment matching procedure proposed by Mineo (1994, 2003). Based on a generalised index of kurtosis as it can be derived for EP distributions, it permits an isolated estimation of b. For sample sizes of 100 or 200 observations it also appears to give the most accurate results (Mineo 2003, p. 118).

[11] More general versions of (8) can also account for asymmetries; see, e.g., Bottazzi and Secchi (2008), or Zhu and Zinde-Walsh (2009). We neglect this extension since skewness will not be much of a problem for us, the convenient moment-based estimation approach of Eq. (9) below would no longer be applicable to two shape parameters b_1 and b_2, and we feel that our samples are typically too small for reliable results.

[12] These estimations are not always without problems: for sample sizes less than 100 it may happen that the likelihood function has no minimum within a reasonable range (Agró 1995, p. 527).

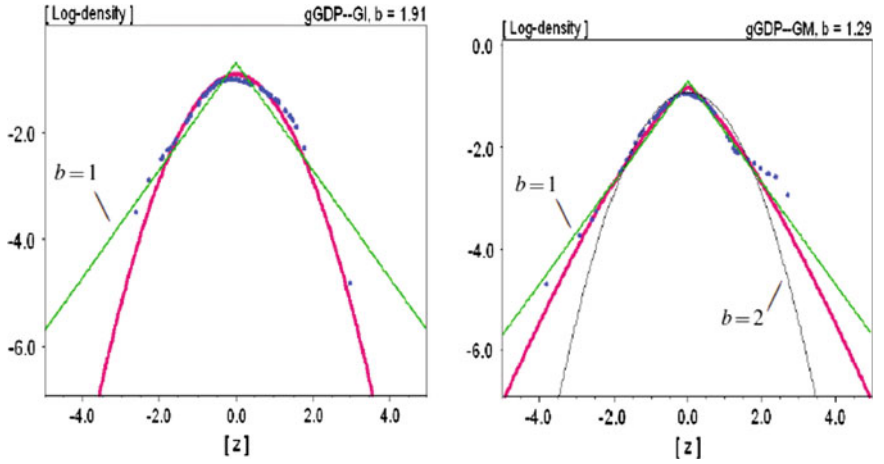

Fig. 1 Estimated densities (EP/nonparametric densities: solid lines/dots)

The approach does not require the optimisation of an objective function, but only the solution of an implicit equation in b, which is indicated by the exclamation mark in the first equation:

$$\frac{\hat{\sigma}}{\hat{d}} \stackrel{!}{=} \frac{\sqrt{\Gamma(3/b)\Gamma(1/b)}}{\Gamma(2/b)}, \qquad \hat{d} = (1/T)\sum_{t=1}^{T}|x_t - \bar{x}| \tag{9}$$

$$\text{iid normality rejected at 5 \% level if } \hat{b} < \begin{cases} 1.578 & \text{for } T = 190 \\ 1.456 & \text{for } T = 95 \end{cases}$$

The expression on the right-hand side of the first equation is the aforementioned alternative index of kurtosis. It is strictly decreasing in b over a sufficiently wide range, hence the equation has a unique root $\hat{b}$.[13] On the other hand, the left-hand side of the equation shows that in the determination of $\hat{b}$ only the second, and no longer the fourth moment is involved, which confirms the expectation articulated above of more robustness.

The simple rule for rejection in (9) is a one-sided test (we are not interested in discriminating $b = 2$ against high values of b). The critical values are readily obtained from 10,000 samples of T random draws from the standard normal. Computing the estimate $\hat{b}$ for each sample, it only remains to determine the 5 % quantile of this collection. The specific values of T referred to will be the typical sample sizes in our empirical analysis.

[13] Estimation of b *via* the implicit equation in (9) is also mentioned in Bottazzi (2004, p. 4), the manual on the software SUBBOTOOLS, which is online freely available at http://cafim.sssup.it/~giulio/software/subbotools. It has, however, to be noted that in older versions—but no longer in the current one—there are two misprints in his formula (the software source code in file `subbofit.c`, on the other hand, is correct).

It should finally be mentioned that Fagiolo et al. (2008) also experimented with the Cauchy, the Student-t and the Lévy-Stable distribution as alternatives to fit fat- and medium-tailed distributions of output growth data. They conclude, however, that the EP density seems to outperform the other three density families.

3 Asymptotic Results for the Empirical Data

Our empirical study is concerned with quarterly output data of the US economy, both in levels and first-differences. Regarding the former, we work with the CBO output gap, while for the growth rates we do not only refer to real GDP (in chained 2005 dollars), but also to the output of the firm sector, which for many models appears to be the more appropriate output concept.[14] The growth rates are annualised and denoted by "gGDP" and "gYF", respectively, the gap series is referenced as "Gap".

The data covers a period of not quite 50 years. It disregards the 1950s and begins in 1960:Q1, and we let it end in 2007:Q2 before the first signs of the financial crisis in the real sector.[15] This amounts to a total of $T = 190$ quarters. Already on the basis of general observations, a part of macroeconomic research divides such an interval into two subsamples, which are commonly called the periods of the GI and the Great Moderation (GM). Estimations are then interested in possible parameter shifts in structural models, representing, for example, a different conduct of monetary policy in the two subperiods. With respect to output there is also rigorous econometric work showing a significant decline in its volatility. In fact, with quarterly growth rates from 1953:Q2 up to 1992:Q2, McConnell and Perez-Quiros (2000) reveal strong evidence for one—and only one—structural break, where the most suitable point estimate for the break date is 1984:Q1. We follow the upshot of their analysis and, maintaining the expressions GI and GM, distinguish two subperiods of almost equal length, GI: 1960:Q1 – 1983:Q4 ($T = 96$ observations) and GM: 1984:Q1 – 2007:Q2 ($T = 94$). The acronym for the full sample period is GIGM.[16]

The main part of Table 1 computes the test statistics from the previous section for the three sample periods and the three original time series. They are recorded in normal font size. If we first consider the full sample period then, as emphasised by the bold face figures, it leaps to the eye that the five statistics lead to very different conclusions concerning non-normality (also "NN" henceforth). One statistic ($\hat{b}$) recognises NN for all three series, and one (BN) rules it out for them. JB and AD conclude NN in two but not identical cases, GJB concludes it in only one case. So not

[14] In essence, the "firm sector" is non-financial corporate business. The data sources are given in the appendix.

[15] While a "black swan" event like the recent crisis in the financial sector is presumably conducive to fat tails here and in the rest of the economy, we are more interested in the question whether non-normality could have also prevailed under less extreme circumstances.

[16] In the context of the so-called New Macroeconomic Consensus there is some discussion about a shift in monetary policy as a major cause for the structural change. Here, however, we need not speculate about this issue.

Table 1 Test statistics (empirical series and residuals from AR(p) estimations)

	GIGM			GI			GM		
	Gap	gGDP	gYF	Gap	gGDP	gYF	Gap	gGDP	gYF
JB:	**7.69**	**11.86**	3.42	1.44	0.17	0.62	0.77	1.84	0.58
	24.73	**27.91**	**6.87**						
GJB:	1.66	**11.53**	3.28	0.38	0.17	0.61	0.17	1.73	0.55
	24.51	**27.64**	**6.80**						
BN:	1.44	4.09	1.39	0.80	0.26	0.56	0.59	2.72	0.46
	5.61	3.43	1.49						
AD:	0.68	**1.47**	**0.82**	0.29	0.36	0.37	0.22	0.62	0.60
	1.80	**1.19**	0.39						
$\hat{b}$:	**1.45**	**1.18**	**1.43**	2.07	1.91	1.84	2.54	**1.29**	**1.40**
	1.09	**1.25**	**1.57**	1.56	1.90	1.93	1.56	1.95	1.68

Note Bold face figures indicate candidates for a rejection of normality. First row of a test statistic refers to the empirical series, second row to the residuals from an AR(3) estimation (in case of the CBO gap) or an AR(2) estimation (in case of the growth rates gGDP and gYF). Regarding the first four statistics, the latter entries are omitted in GI and GM since they offer no clue to non-normality

only across the different types of test statistics, but also even within the class based on skewness and kurtosis (JB, GJB and BN), the evidence is rather mixed. The different outcomes illustrate that one should be cautious with claims of "non-normality" that are referring to no more than one test statistic.

The results can be checked by purging the time series of their autocorrelation structure. Here, the data are conceived of as being generated by a convenient stochastic process. General econometric studies often content themselves with AR(1) processes for this purpose, while Christiano (2007) was more ambitious and employed a four-lag VAR with seven variables (for monthly data, though). We believe that in the present context the uncorrelated residuals from AR(p) estimations are good enough, where for all three sample periods a lag length $p = 3$ is sufficient for the CBO output gap, and $p = 2$ for the two growth rates (parsimony in the number of parameters is not important in this respect, but higher lags yield no further noteworthy improvement in the fit).

The test statistics for these residuals are the entries in smaller font size in Table 1. Regarding GIGM both can happen, that normality is accepted for the depurated series and not for the original data (AD for gYF), and the other way around (in several cases). Most remarkable are GJB and AD for the output gap, according to which the innovations in the AR(3) process are strongly non-normal, while despite the linear structure this property does not carry over to the data produced by these shocks.

Almost any evidence of non-normality disappears when the shorter subperiods of GI and GM are considered, which holds for the original and the AR-filtered series alike. For all but the $\hat{b}$-test, the statistics are far from their critical values. Only for the two growth rates in GM (but not GI), values of $\hat{b}$ are obtained that fall short of the critical values given in Eq. (9). The low value of $\hat{b} = 1.29$ for gGDP is especially

striking since the residuals from the AR(2) estimation are almost perfectly normal according to this criterion.

4 Small-sample Results for the Empirical Data

The critical value for $\hat{b}$ in (9), which in Table 1 lead us to a rejection of normality for gGDP and gYF in GM, was established under the null hypothesis of independent draws from a normal distribution. Since also for the uncorrelated AR(2)-residuals of the series one fails to reject normality of the shape parameter, doubts may arise as to whether the NN conclusion could be maintained once the serial correlation in the original series is properly taken into account (even if it is not overly strong). Regarding the conclusions from the four other test statistics, they are based on asymptotic theory and one may ask for their small-sample properties. These questions bring us to the second stage of the analysis, which is a battery of straightforward MC experiments.

To this end we take the AR(p) estimates of a time series and simulate this process over the empirical sample size (after discarding a longer period at the beginning to rule out any transient effects). The innovations are independently drawn from the normal distribution with a variance equal to the estimation's squared standard error. This is repeated 10,000 times and for each of these MC samples the test statistics are computed.

By construction, the statistics should diagnose normality. According to the asymptotic theory for JB, GJB, BN, AD and the iid assumption for $\hat{b}$ in (9), the rules for rejecting normality should be false in just 5 % of the 10,000 MC samples. Specifically, we have here stylised but empirically relevant conditions on time series data for which we can check how reliably this is done. In econometric terms, we can determine the so-called *size* of the five tests, that is, the probability of committing a type I error by rejecting normality when in fact this null hypothesis is true. This is the first kind of results presented in Table 2, which are independent of the values for the empirical test statistics in Table 1.

Related to this information are the quantiles of the collection of the simulated statistics; the 95 % quantiles for JB, GJB, BN, AD and the 5 % quantile for $\hat{b}$. These critical values for rejection in small samples under the present circumstances are reported in column *crit* in Table 2. The statistics *emp* of the empirical samples (reproduced from Table 1) can now be directly compared to them for a definite conclusion. In addition, we can compute what quantile q a value of *emp* constitutes in the MC distribution and obtain the p-value of the corresponding test statistic from it, which has the following interpretation: if instead of *crit*, the value *emp* of the empirical statistic were employed as a benchmark for rejection, then for JB, GJB, BN, AD the percentage $p = 1 - q$ would be the error rate of falsely rejecting normality, and for $\hat{b}$ it would be $p = q$. Certainly, at this paper's significance level non-normality would only be concluded with a p-value of less than 5 %, and p-values above 5 % would give us an idea of how safe we can feel when accepting normality.

Table 2 Statistics from AR(p) simulations with normal innovations

	size	*crit*	*emp*	*p*	*size*	*crit*	*emp*	*p*
GIGM	Gap							
JB :	33.6	17.71	7.69	22.7				
GJB :	2.7	4.30	1.66	25.2				
BN :	5.5	6.09	1.44	72.6				
AD :	53.3	2.44	0.68	61.5				
$\hat{b}$:	12.1	1.39	1.45	6.7				
GIGM	gGDP				gYF			
JB :	4.6	5.78	11.86	**1.0**	4.8	5.84	3.42	14.4
GJB :	4.3	5.61	11.53	**0.9**	4.4	5.60	3.28	14.2
BN :	6.3	6.37	4.09	17.9	6.1	6.28	1.39	59.4
AD :	4.6	0.78	1.47	**0.1**	4.7	0.78	0.82	**3.9**
$\hat{b}$:	4.9	1.58	1.18	**0.0**	5.0	1.58	1.43	**1.1**
GM	gGDP				gYF			
JB :	4.3	5.58	1.84	32.8	4.4	5.57	0.58	72.7
GJB :	3.8	5.20	1.73	32.2	4.2	5.46	0.56	72.8
BN :	2.6	5.29	2.72	31.9	2.6	5.34	0.46	86.7
AD :	5.4	0.80	0.62	13.0	5.2	0.80	0.60	13.8
$\hat{b}$:	4.9	1.46	1.29	**1.0**	5.1	1.45	1.40	**3.2**

Note Based on 10,000 MC samples for each series and subperiod. Size and p-values in per cent, column "*crit*" indicates the critical quantiles of the MC distributions (95 and 5 %, respectively), and "*emp*" the empirical statistics from Table 1

Table 2 is limited to the series and sample periods for which non-normality was not outright denied by the tests in Table 1. To begin with the evaluation of the size in Table 2, its values for the single statistics (except perhaps BN) are all satisfactorily close to the nominal level of 5 % if we look at the two growth rates. By contrast, there are dramatic deviations for JB and AD when these tests are applied to the CBO output gap. Regarding JB, this is due to an increase of the asymptotic rejection level from $\chi^2(2) = 5.99$ to a 95 % quantile of 17.71, for AD the previous standard level of 0.787 increases here to 2.435. The decline of the 5 % level of $\hat{b}$ from 1.578 (for $T = 190$) to 1.394, which raises the size of this test to 12.1 %, is a weaker but still unpalatable phenomenon. These deteriorations are mainly caused by the strong serial correlation of 0.93 in the output gap, which of course are accounted for by the AR(p) coefficients in the simulations. In comparison, the size effects from the correlation between 0.23 and 0.30 in the growth rates appear rather unimportant.

The approach of capturing the autocorrelation structure of the output gap by the Monte Carlo simulations has also consequences for the NN conclusions. Now, there is no statistic left that would recommend a rejection of normality; only $\hat{b}$ has a p-value that is above but not too far away from the 5 % level. Therefore, to sum up

our evidence concerning the CBO output gap, we do not have sufficient support of any non-normal features in this series.[17]

Prospects of non-normality are better for the growth rates of output. Regarding the full sample period, four of the five p-values reject normality for gGDP (even strongly so), and normality of the firm sector growth rates is rejected by AD and $\hat{b}$. Regarding the GM subsample, recall that only the $\hat{b}$ test provided evidence against normality in Table 1. This tentative result is now fully confirmed by the Monte Carlo experiment; cf. the bottom part of Table 2. The fact that here the three statistics based on skewness and kurtosis do not offer the least clue against normality suggests that the test based on the EP distributions measures something more general. Given the p-values of 13.0 % and 13 % for AD, this test, which likewise considers the shape of an entire distribution, seems to be somewhere in the middle between the two principles. Hence, summarising claims of non-normal growth rates can only be properly assessed with additional information about the particular specification of their non-normality.

Finally, we can also try to make sense of the fact that for gGDP in GM the estimated $\hat{b} = 1.29$ reveals a non-negligible non-normality, whereas the estimated innovations are nearly normally distributed ($\hat{b} = 1.95$ in Table 1). With the linear AR(2) filter, as we have just seen for $b = 2$, this should be an extremely rare event. A tentative alternative conclusion could thus be that generally an AR(p) process, or even a multivariate VAR, may not be a good hypothesis, and that the strong differences in the two statistics are rather indicative of a strong nonlinear mechanism in this sample period.

5 A Two-regime Monte Carlo Experiment

As already pointed out by McConnell and Perez-Quiros (2000), the unique structural break they identify consists of a significant decline in output volatility. This phenomenon is also clearly visible in our three empirical series, even in time series diagrams. Table 3 documents a decrease of about one-half from GI to GM not only in the standard deviations of the empirical data, but also in the standard deviations of the residuals from the AR(p) estimations. One needs no formal statistical tools to classify this as a significant change. On the other hand, the changes in the mean growth rates or the AR(p) slope coefficients are much more moderate.

Intuitively, if we have a time series with strong noise in the first half and weak noise in the second half, one may suppose that this gives rise to a higher kurtosis, since the higher values (in modulus) in the first half become a rarer event when considering the full sample. The argument should be valid even if the random forces

[17] This finding, in particular, means that De Grauwe's (2012) reference to a non-normal output gap, which he points out his model is successfully able to reproduce, is not warranted; not even if one goes beyond his argumentation with the kurtosis or the Jarque-Bera statistic. So far, however, this observation only indicates that the issue of "non-normality" requires further discussion of what his model should more precisely achieve.

Table 3 Standard deviations of the empirical series and their AR(p) residuals (upper and lower row, respectively)

Gap		gGDP		gYF	
GI	GM	GI	GM	GI	GM
2.93	1.43	4.45	2.12	4.12	2.62
0.98	0.44	4.28	1.90	3.89	2.47

Table 4 Two-regime Monte Carlo simulations, pooling AR(p) from GI and GM

	Gap			gGDP			gYF		
	crit	emp	p	crit	emp	p	crit	emp	p
JB:	65.48	7.69	49.9	50.49	11.86	45.5	22.25	3.42	46.9
GJB:	14.92	1.66	51.9	48.45	11.53	44.9	20.94	3.28	46.5
BN:	7.01	1.44	75.7	8.12	4.09	47.1	5.90	1.39	68.9
AD:	4.66	0.68	81.6	2.27	1.47	25.4	1.21	0.82	17.9
$\hat{b}$:	1.02	1.45	38.9	1.09	1.18	14.8	1.32	1.43	14.8

are normally distributed in each of the two subsamples, and a certain amount of non-normality may also be indicated by other test statistics. The conjecture that a structural change in the volatility may contribute to the non-normality results above can, however, be readily checked by another Monte Carlo experiment.

Consider the GDP growth rates over GIGM and take the NN results based on JB, GJB and AD in Table 2 as an example. The simulated data generation process was an AR(2) over the full length of $T = 190$ quarters, where the normal innovations yielded values of the test statistics that are in their vast majority below the empirical estimates (conversely for the $\hat{b}$ statistic). We now maintain the null hypothesis of normal innovations but introduce the regime shift from GI to GM into the simulations. That is, the first $T_{GI} = 96$ periods of the simulation (again after a suitable transitory period) adopt the AR(2) coefficients estimated over GI, and the second $T_{GM} = 94$ periods use the coefficients as they have been estimated over GM. This, in particular, includes the different variances for the normal random draws (which according to Table 3 are equal to $(4.28)^2$ and $(1.90)^2$, respectively).

The question we then ask is whether or to what extent the structural break tends to increase the simulated JB, GJB, AD statistics (and to decrease the values of $\hat{b}$). The answer in the middle of Table 4 is unambiguous. Compared to the experiment in Table 2, we observe a drastic increase in the critical 95 % quantiles of these statistics (and a sizeable decrease of the 5 % quantile of $\hat{b}$). The quantiles even exceed the corresponding empirical value (or fall short of it in case of $\hat{b}$), with the consequence that the p-values distinctly rise above the 5 % levels.

The same effects are obtained for the growth rate gYF and, for completeness, the output gap. Overall, there is no single p-value lower than even 10 %. It can therefore be said that the previous findings of non-normality over the full sample period may be spurious; all evidence of non-normality could be explained by normal innovations in a simple linear stochastic process once one takes account of the significant decline in their volatility. While admittedly, the experimental design with the sudden change at the break data is somewhat crude, the high p-values in Table 4 do not give us much

reason to expect that a smoother transition from one regime to the other would lead to an essential weakening of the conclusion.

6 The Laplacian as an Alternative Hypothesis

Because of the unconvincing evidence of non-normality over GIGM and since in the subsamples GI and GM the only indicator of a possible non-normality is the shape parameter b of the EP distribution, we will in the remainder of the paper be exclusively concerned with this statistic. Given that for the growth rates in GM normality was strongly rejected by the tests in Table 2, we may put forward an alternative hypothesis. As mentioned above, already for systematic reasons the Laplace distribution with $b = 1$ is usually considered to be a natural antithesis to the normal distribution with $b = 2$. As furthermore the growth rate estimates $\hat{b} = 1.29$ and $\hat{b} = 1.40$ are closer to 1 than to 2, $b = 1$ does not appear to constitute an unreasonable alternative benchmark distribution.

Let us begin with a geometric account of the goodness-of-fit of the shape parameter estimations for the growth rates in GI and GM. More specifically, we can also gain an intuitive impression of how far their distributions are from the two benchmark distributions. To this end, we draw the densities functions of the standardised GDP growth rates z in the semi-log diagrams in Fig. 1, where with respect to the estimated scale and location parameters $\hat{a}$ and $\hat{m}$ and the original growth rates $x = \text{gGDP}$, the standardised values are given by $z = (x - \hat{m})/\hat{a}$.[18] The diagrams for the firm sector growth rates do not look very different, so Fig. 1 is sufficiently representative.

The left-hand panel in Fig. 1 deals with the GI subsample. The exponential power density constituted by the estimated shape parameter $\hat{b} = 1.91$ is drawn as the bold (red) line. The dots distributed around it are the density values of the $T = 96$ observations of this period. That is, we use a standard nonparametric approach to compute a kernel density estimator $\hat{f} = \hat{f}(z)$ of the empirical values of z and for each observed z_t plot the point $(z_t, \hat{f}(z_t))$.[19] Over a wide range and especially in the middle part, these points really nestle into the smooth curve of the theoretical density function. The estimation thus inspires confidence, its outcome being more than just the result of a somewhat abstract and technical concept.

The log-density of the normal distribution looks very similar to the EP density with $b = 1.91$, which is the reason why it has been omitted in this diagram. The other polar case of the Laplacian density, where $b = 1$, is the tent-shaped thin solid line. Over most of the empirical range it yields a clearly worse fit. One might nevertheless argue that it could perhaps provide better results for the moderate and more extreme negative values of z_t, which overall may even suggest an asymmetric estimation approach (as it was hinted at in fn 11) with different values of b for

[18] A prescription of how $\hat{a}$ and $\hat{m}$ are obtained after $\hat{b}$ has been estimated before can be found in the Appendix.

[19] We employ the Epanechnikov kernel for this purpose; see Davidson and MacKinnon (2004, pp. 678–683) for the computational details.

positive and negative values of z. Here, we mainly abstain from this idea since given the relatively small sample size, this would also require an econometric discussion of the risk of overfitting.

The right-hand panel in Fig. 1 presents the results for the GM period. The estimated density function with $\hat{b} = 1.29$ spreads wider away from that of the normal distribution (the curve with $b = 2$ in the diagram), thus giving greater weight to the outer regions. To some part this seems to be an implication of the sharper turnaround in the centre; to another part the shallower slope of the points $(z_t, \hat{f}(z_t))$ with z_t between 1 and 3 may contribute to it, although there are again no observations with $|z_t| \geq 4$. Both the outer positive and negative points in the diagram indicate that now even the Laplacian density would not accomplish too bad a fit.

The latter observation prompts the idea of investigating the Laplacian as a specific alternative hypothesis to normality. More precisely, we can put forward $b = 1$ as another null hypothesis and test it with similar Monte Carlo experiments to Table 2, now with re-estimations of the $\hat{b}$-statistic only. However, we wish to weaken the underlying structural assumptions, that is, we no longer estimate an AR(p) process (now with Laplacian innovations). Instead, we randomly draw directly from the standardised Laplace distribution, though in such a way that also the empirical first-order autocorrelation ρ is taken into account.[20] Of course, the normal distribution ($b = 2$) can be dealt with in the same way.

We run this experiment with the sample sizes $T = 96$ and $T = 94$ for GI and GM, respectively, again 10,000 simulation runs for each case.[21] Re-estimating the shape parameter $\hat{b} = \hat{b}^c$ for each such run c ($c = 1, \ldots, 10,000$), a distribution $\{\hat{b}^c\}$ is obtained to which we can relate the empirical estimate $\hat{b}$ from above. The tests that we thus carry out are one-sided. We reject normality if this $\hat{b}$ is below the 5 % quantile $Q_{0.05}$ of the MC distribution under the null of $b = 2$, and we reject the Laplacian if this $\hat{b}$ is above the 95 % quantile $Q_{0.95}$ of the distribution under the null of $b = 1$. The p-values are determined accordingly.

The results presented in Table 5 leave us a clear and pronounced message. For both growth rate series alike, as emphasised by the bold face figures, it can be concluded that over the GI period the Laplacian is rejected and normality is accepted. By contrast, over GM it is just the other way around: here normality is rejected and the Laplacian is accepted, in the sense that it cannot be ruled out.

These results are based on the particular values of the autocorrelation in the empirical data. Within the typical range of serial correlation in the quarterly growth rates of aggregate output, however, the critical quantiles of the MC distributions $\{\hat{b}^c\}$ remain quite insensitive. Regarding the choice between the normal and the Laplace distribution we can therefore conclude this subsection with putting forward a simple

---

[20] This sampling could be viewed as an AR(1) process with Laplacian innovations, the variance of which is linked to ρ (and, however, ρ only); see the Appendix for this and the generation of iid. pseudo random numbers from an EP distribution. Considering the standardised Laplace distribution suffices since the estimation of b is independent of the other parameters m and a.

[21] Apart from the strong tendency towards normality documented in Table 1, subjecting the CBO output to the same exercise is not very informative because its high autocorrelation of $\rho = 0.93$ leads to an extremely wide dispersion in the re-estimated values of b.

Table 5 Testing the normal ($b = 2$) and the Laplace ($b = 1$) distribution

	empirical		null: $b = 2$		null: $b = 1$	
	ρ	$\hat{b}$	$Q_{0.05}$	p	$Q_{0.95}$	p
gGDP						
GI:	0.26	1.91	1.44	**36.8**	1.68	1.5
GM:	0.23	1.29	1.44	1.6	1.65	**29.1**
gYF						
GI:	0.30	1.84	1.44	**31.0**	1.73	2.9
GM:	0.26	1.40	1.43	3.9	1.69	**20.0**

Note Each case based on 10,000 Monte Carlo samples of T random draws from the EP distribution with shape $b = 1$ or $b = 2$, respectively, correlated with coefficient ρ (T being the empirical sample size). $Q_{0.05}$ and $Q_{0.95}$ are the 5 and 95 % quantiles of the MC distribution of the re-estimated $\hat{b}$. p-values in per cent and for one-sided tests, that is, for $b = 2$ ($b = 1$) p is the percentage of values in the MC distribution that are less than the empirical $\hat{b}$ (larger than this $\hat{b}$)

rule of thumb. Referring to a correlation coefficient $\rho = 0.25$ and a sample size $T = 95$, it reads,

$$\begin{aligned} \hat{b} &> 1.671 \qquad \text{reject } b = 1 \\ \hat{b} &< 1.438 \qquad \text{reject } b = 2 \\ \hat{b} &\in [1.438, 1.671] \text{ compatible with both } b = 1 \text{ and } b = 2 \end{aligned} \tag{10}$$

Perhaps easier to recall, we can also say that the inconclusive range is given by 1.55 ± 0.12, while above that interval normality may be accepted and below it a Laplace distribution.

7 On the Precision of the Estimates of the Shape Parameter

Regarding the stylised facts of the macro economy, a model may be judged by, *inter alia*, how well its output growth rates are able to reproduce the empirical shape of the EP distribution. In order to put the model's degree of matching into perspective, we need to know something about the precision of the estimation of $\hat{b}$. Conventionally, we thus ask for the standard error of $\hat{b}$. Readily available for this is the asymptotic variance, which can be explicitly computed as

$$\mathrm{Var}(\hat{b}) = \frac{\hat{b}^3}{(1+1/\hat{b})\,\Psi'(1+1/\hat{b}) \;-\; 1} \tag{11}$$

where $\Psi'(\cdot)$ is the trigamma function, that is the second derivative of the logarithm of the Gamma function (Agró 1995, pp. 524f; Bottazzi and Secchi 2008, p. 5). As it should be, $\mathrm{Var}(\hat{b})$ is independent of the location and scale of the distribution. On the other hand, it changes with the level of the estimate. While the denominator in (11) is rising with $\hat{b}$, the increase in the numerator is stronger. Hence, the more normal the distribution, so to speak, the higher the variance. These variations are sizeable. For example, $\hat{b} = 1$ and $\hat{b} = 2$ give rise to a variance of 3.45 and 19.89, respectively, meaning that the standard error more than doubles.

Now, one may be sceptical about employing (11), not only because of the relatively small size of our samples, but also since it derives from the maximum likelihood estimation of independent random draws from the EP distribution. In addition to (11), we therefore make use of two bootstrap procedures to determine the confidence intervals around $\hat{b}$.

A first and obvious approach takes up the Monte Carlo experiments in the previous section where 10,000 samples of autocorrelated data were generated under the null hypothesis of $b = 1$ and $b = 2$. Here, we only have to replace these polar values with the empirical estimates $\hat{b}$. This procedure can be viewed as a parametric bootstrap. The standard deviation of the collection of the re-estimated values $\{\hat{b}^c\}$ gives us the bootstrapped standard error, and suitable quantiles of it the lower-and upper-bounds of a confidence interval.

The second approach is a nonparametric bootstrap that directly samples from the empirical data set with its T observations. Because of the serial correlation, three block bootstraps BB1, BB4 and BB10 are considered, which sample (with replacement, of course) from the overlapping blocks of length 1 (the degenerate case), 4 and 10, respectively. These frequency distributions are likewise referred to as $\{\hat{b}^c\}_{c=1}^{10,000}$.

Table 6 reports some basic indicators of the dispersion of the bootstrap distributions, namely, the 5 and 95 % quantiles and the standard deviation, that is the bootstrapped standard error of the estimation. Once again, these statistics are computed for GI and GM, and for the output growth rates of GDP and the firm sector. What holds for all four cases is that the confidence intervals are not symmetric around the estimated value of b: the positive deviations of $\hat{b}^c$ from the estimate $\hat{b}$ are larger than the negative deviations. These distortions can be even so serious that the standard errors would be misleadingly high. In the calculation of the latter we have therefore truncated the $\hat{b}^c$ at 5.

Concerning the question of whether the data could be compatible with the Laplace distribution, the answer from the confidence intervals is strongly negative for the GI period and weakly negative for GM. The normal distribution can be dismissed for the GDP growth rates in GM if we consult the block bootstraps, but not if the parametric bootstrap is employed. For gYF, none of the bootstrap distributions stays away from $b = 2$. Generally, owing to the short sample periods, the precision of the estimations is so limited that we have to be satisfied if at least one of the polar cases $b = 1$ or $b = 2$ is ruled out. However, this may contradict the results from Table 5, which for GM could not reject the hypothesis $b = 1$, whereas here the 5 % quantiles of $\{\hat{b}^c\}$ are bounded away from unity.

Table 6 Dispersion statistics of the bootstrapped distributions $\{\hat{b}^c\}$

	gGDP				gYF			
	emp	$Q_{0.05}$	$Q_{0.95}$	*ser*	*emp*	$Q_{0.05}$	$Q_{0.95}$	*ser*
GI	1.91				1.84			
PB:		1.40	3.29	0.633		1.37	3.19	0.608
BB1:		1.36	3.17	0.600		1.34	3.12	0.590
BB4:		1.35	3.18	0.609		1.38	2.96	0.525
BB10:		1.39	3.27	0.627		1.38	2.96	0.524
$\sigma_b^\star$:				**0.442**				**0.395**
GM	1.29				1.40			
PB:		1.03	2.12	0.349		1.12	2.38	0.409
BB1:		1.05	1.88	0.282		1.08	2.10	0.341
BB4:		1.08	1.84	0.240		1.10	2.11	0.332
BB10:		1.10	1.78	0.215		1.09	2.28	0.383
$\sigma_b^\star$:				**0.189**				**0.293**

Note Sample size of the $\{\hat{b}^c\}$: 10,000. $Q_{0.05}$ and $Q_{0.55}$ are their 5 and 95 % quantiles, *ser* the standard error, and *emp* the empirical estimate. PB is the parametric bootstrap with the empirical autocorrelation, BBn the block bootstrap with block length n ($n = 1, 4, 10$), while $\sigma_b^\star$ is the standard deviation that equates the normal density at its mode to the (estimated) density of BB10 at its mode

Because of the asymmetry already noted, the standard error has to be interpreted with caution. On the other hand, such a statistic is not only a succinct information, but also of practical use. In assessing the deviations of a model-generated $\hat{b}$ from its empirical counterpart and making them comparable to the deviations of other summary statistics (also called moments), these magnitudes have to be suitably weighted; in particular, if, following Ruge-Murcia (2012) mentioned in the Introduction, a model is estimated by the method of simulated moments. The common treatment here is a multiplication of the model deviation with the inverse of the variance of the moment. This controls for the scale and the weight is proportionately higher the higher the precision of the estimate, or the lower its variance.

A problem arises with asymmetric distributions like our bootstraps since then the variance may imply a certain overstatement, such that the weight of this moment is unduly low. To discuss this issue, consider the bold line of the distribution BB10 for gGDP over GI in Fig. 2 (BB10 is quite as good as any of the other bootstrap distributions). The kind of asymmetry is clarified by the vertical dotted line, which is the median of the distribution. The idea of the conventional weighting factor rests on the supposition that the corresponding moment is nearly normally distributed around $\hat{b}$, at least in some neighbourhood of this estimate. If we try to approximate such a distribution by a normal distribution around $\hat{b}$ with the variance σ^2_{BB10} of the bootstrap, we see that it assigns probabilities that are far too low for values near $\hat{b}$, and too high for estimates less than 1.30 (roughly).

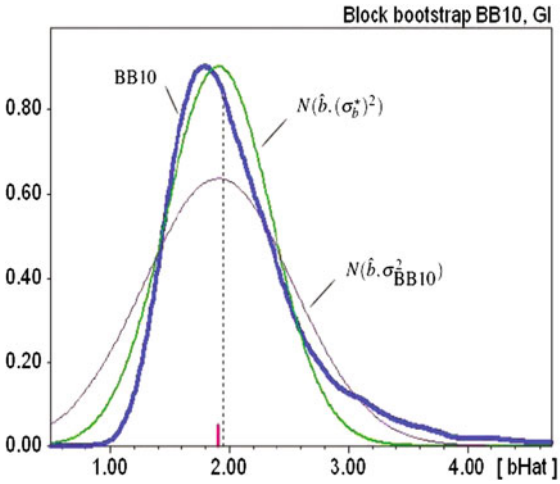

Fig. 2 Bootstrapped density BB10 of the shape parameter estimate $\hat{b}$ ('bHat'). *Note* Bootstrap of the GDP growth rates in GI. The dotted line indicates the median of BB10, the bold (*red*) bar at the bottom the empirical estimate $\hat{b}$

Table 7 Mode-consistent and asymptotic standard errors for $\hat{b}$

$\hat{b}$:	1.00	1.29	1.40	1.84	1.91	2.00
$\sigma_b^\star$:	–	0.189	0.293	0.395	0.442	–
$\sqrt{\text{Var}(\hat{b})\,/\,T}$:	0.191	0.262	0.291	0.408	0.429	0.458

Note Underlying are $T = 96$ for $\hat{b} = 1.84, 1.91$ (GI), $T = 94$ for $\hat{b} = 1.29, 1.40$ (GM), and $T = 95$ for the benchmark cases $\hat{b} = 1, 2$.

Clearly, responsible for these distortions is too low a value of this normal density function at $\hat{b}$. A better approximation would be a normal distribution that is equally high in the centre as BB10 at its mode. Accordingly, we specify $\sigma_b^\star$ as the value that renders the density of $N(\hat{b}, (\sigma_b^\star)^2)$ at $\hat{b}$ equal to the value of the estimated density function of BB10 at its mode. Figure 2 illustrates that over the main range of BB10, this normal distribution is indeed a suitable approximation. Hence, if reference is made to a single standard error, this mode-consistent standard error $\sigma_b^\star$ appears to be the most appropriate concept. Its values are shown as the boldface figures in Table 6, certainly all of them being smaller than the standard deviations of the asymmetric bootstrap distributions.

Even if the specification of $\sigma_b^\star$ is not backed up by rigorous econometric theory, we believe that the intuitive argument of Fig. 2 makes good sense. To put the results obtained for the mode-consistent standard errors into perspective, we should nevertheless return to the asymptotic variance $\text{Var}(\hat{b})$ in (11) and compare the corresponding errors $[\text{Var}(\hat{b})/\,T]^{1/2}$ to them. This is done in Table 7. It shows that three of the four $\sigma_b^\star$ are amazingly close to the asymptotic standard errors. The difference between the two statistics is larger for gGDP in GI, where $\hat{b} = 1.29$, but the ranking

across the $\hat{b}$ is still preserved. On the whole, the table suggests that despite the small samples and the neglect of serial correlation, the asymptotic standard error is not so unreliable after all. At least as far as the information content of a single statistical number is concerned, one may thus save the effort of a bootstrapping procedure and invoke (11) directly.

8 Conclusion

The primary motivation for this paper were recent claims in the literature that the US output growth rates fail to be normally distributed. However, although all of these studies have long sample periods of more than 40 years in common, none of them mentions the possibility of a structural break that might or might not invalidate the results. A straightforward Monte Carlo experiment could demonstrate that this neglect is indeed unwarranted. Simulating two AR(p) processes for the two periods of the GI and Great Moderation (GM) with their different variances of the—normally distributed—innovations, the pasted growth rate series is typically found to exhibit non-normal behaviour of the type measured in the empirical data. Therefore, as long as no new and more sophisticated evidence is provided, the previous results of non-normality appear to be spurious.

Correspondingly, we were then looking for non-normality within the two shorter subsamples. Of the five test statistics that we considered, the only indication of non-normality is the estimated shape parameter $\hat{b}$ of the exponential power distribution over GM, which distinctly falls short of the benchmark for normality, $b = 2$. Here, it has to be noticed that this finding did not take the serial correlation in the data into account, and that the estimated $\hat{b}$ for the residuals of a suitable AR(p) process does not essentially deviate from $b = 2$, or only moderately so. This suggests that if it is normally distributed innovations that ultimately drive the economy, the transmission mechanisms may be of a distinctly nonlinear nature. Alternatively, of course, there may also be non-normalities in the shock processes themselves.

As a first and largely atheoretical step in this direction, we put forward the hypothesis that the growth rates were obtained from a normal ($b = 2$) *vis-à-vis* a Laplace ($b = 1$) distribution with its fatter tails, where both of them exhibit the empirical autocorrelation. The p-values from this Monte Carlo experiment allowed us the interpretation that normality prevailed in GI, whereas the Laplace distribution took over in GM.

This is a nice and pronounced statement that has not been put up to discussion before. On the other hand, it has to be admitted that the message cannot be fully maintained from the perspective of the bootstrapped confidence intervals around the estimated $\hat{b}$. This qualification may be taken as a final example that any claim of non-normality requires an additional discussion of the specific measurement approach, in order to clarify the kind of statements that could be legitimately made. The main

problem that, however, remains is the limited number of empirical observations over a structurally stable period of time that we have.[22]

Appendix

Data sources
The data of potential output are from the report "The Budget and Economic Outlook: Fiscal Years 2012–2022" (January 2012), downloadable at http://www.cbo.gov/publication/42912. Real GDP was obtained from the Bureau of Economic Analysis, http://www.bea.gov/national/index.htm#gdp.

The firm sector output series was extracted from the database `fmdata.dat` in the zip file `fmfp.zip`, provided by Ray Fair for working with his macroeconometric model. It is a plain textfile downloadable from http://fairmodel.econ.yale.edu/fp/fp.htm. The acronym to identify the series is 'Y', as explained in Appendix A.4, Table A.2., of the script *Estimating How The Macroeconomy Works* by R.C. Fair, January 2004, which can be downloaded from http://fairmodel.econ.yale.edu/rayfair/pdf/2003a.pdf .

Approximation of the standard normal cumulative distribution function
Let $\phi = \phi(x)$ be the probability density function of the standard normal, $\phi(x) = \exp(-x^2/2) \,/\, \sqrt{2\pi}$, and $\Phi = \Phi(x)$ the standard normal cumulative distribution function. Then according to Abramowitz and Stegun (1964, p. 932, algorithm 26.2.17), up to an absolute error $|\varepsilon(x)| < 7.5 \times 10^{-8}$, the latter is approximated as follows,

$$\Phi(x) = \begin{cases} 1 - c & \text{if } x \geq 0 \\ c & \text{if } x < 0 \end{cases} \qquad \text{where}$$

$$c = \phi(x)\,[\,b_1 z + b_2 z^2 + b_3 z^3 + b_4 z^4 + b_5 z^5\,] \;+\; \varepsilon(x)$$
$$z = z(x) = 1 \,/\, (\,1 + b_0\,|x|\,)$$
$$b_0 = \;\; 0.2316419 \qquad b_1 = 0.319381530$$
$$b_2 = -0.356563782 \qquad b_3 = 1.781477937$$
$$b_4 = -1.821255978 \qquad b_5 = 1.330274429$$

Estimation of a and m for an exponential power distribution
Suppose that the shape parameter b has already been estimated before as described by eq. (9) in the main text. Setting, in a ML estimation, the partial derivative of the log-likelihood function with respect to m equal to zero, $\hat{m}$ can then be obtained as the solution of the implicit equation in m,

[22] If one believes in non-normality as a stylised fact that a theoretical (stochastic) model should be able to reproduce, one may require this feature for its long-run behaviour, but may also check its small-sample properties—which preferably could be similarly weakly conclusive, or even inconclusive, as the empirical data.

$$\sum_{t=1}^{T} |x_t - m|^{\hat{b}-1} \ \mathrm{sgn}(x_t - m) \ = \ 0$$

where sgn is the sign function, $\mathrm{sgn}(y) = 1$ $(0, -1)$ if $y > 0$ ($y = 0$ or $y < 0$, respectively); cf. Mineo (2003, p. 112). On this basis, Chiodi (1988) has proposed the following expression as an unbiased estimate of a,

$$\hat{a} \ = \ \left[\frac{\sum_{t=1}^{T} |x_t - \hat{m}|^{\hat{b}}}{T - \hat{b}/2} \right]^{1/\hat{b}}$$

(quoted from Mineo and Ruggieri 2005, p. 4, Eq. (9)).

Random variates from EP distributions

In general, the generation of pseudo random numbers drawn from an EP distribution involves draws from a Gamma distribution, which in turn requires some computational effort (see, e.g., Zhu and Zinde-Walsh 2009, p. 91, or Li (2011), Sect. 2, which both allow for an asymmetric shape also). For the class of standardized distributions with shape $b > 1$ (besides $m = 0$, $a = 1$), Chiodi (1995, Sect. 4) set up a faster and easy-to-implement algorithms which has the advantage that it only needs the generation of uniformly distributed random numbers. A random number z is here generated in the following two stages[23]:

1. Repeat
   draw U and V from the uniform distribution over $[-1, +1]$
   and put $W = |U|^b + |V|^{b/(b-1)}$
   until $W \leq 1$.
2. Put $z = U \cdot [-b \ \ln(W)/W]^{1/b}$.

While the procedure fails to be applicable to $b = 1$, we checked that it is robust and works well for values of b arbitrarily close to unity. In our experiments with $b = 1$ it is thus perfectly sufficient to have recourse to the approximation $b = 1.00001$.

Of course, the draws thus obtained are iid. To take account of an autocorrelation ρ put, in round t, $z_t = \rho \, z_{t-1} + \sqrt{1 - \rho^2} \, \tilde{z}$, where besides $|\rho| < 1$ it is supposed that z_{t-1} is a draw from the previous round and $\tilde{z}$ a draw from the EP distribution, both of them with the same variance σ^2. It is easily seen that then $\mathrm{Var}(z_t) = \sigma^2$ and $\mathrm{Corr}(z_t, z_{t-1}) = \rho$. It is well-known that for normal distributions, $b = 2$, z_t is normally distributed, too. We know of no mathematical proof that establishes the analogous statement for general values of b. The property can, however, be confirmed by simulation studies, even for b close to one, although (very) large samples are required for a satisfactory convergence of the sample density function towards the theoretical density (the smaller b or the higher ρ, the larger the samples).

[23] The procedure can still be accelerated by suitable squeeze methods, at the price of a more complicated computer code. Since the original version is already fast enough, this does not seem worth the effort.

References

Abramowitz, M., & Stegun, I. (1964). Handbook of mathematical functions. Electronic copy of p. 932 at http://people.math.sfu.ca/cbm/aands/page_932.htm

Agró, G. (1995). Maximum likelihood estimation for the exponential power function parameters. *Communications in Statistics-Simulation and Computation*, *24*(2), 523–536.

Alfarano, S., Milaković, M., Irle, A., & Kauschke, J. (2012). A statistical equilibrium of competitive firms. *Journal of Economic Dynamics and Control*, *36*(1), 136–149.

Ascari, G., Fagiolo, G., & Roventini, A. (2012).*Fat-tail distributions and business-cycle models*. Working paper series, Department of Economics, University of Verona.

Bai, J., & Ng, S. (2005). Tests for skewness, kurtosis, and normality for time series data. *Journal of Business and Economic Statistics*, *23*(1), 49–60.

Blatt, J. M. (1983). *Dynamic economic systems: A post-Keynesian approach*. Sharpe: Armonk, N.Y.

Bottazzi, G. (2004). *Subbotools user's manual*. LEM Working Paper Series 2004/14, Sant'Anna School of Advanced Studies, Pisa.

Bottazzi, G., & Secchi, A. (2008). *Maximum likelihood estimation of the symmetric and asymmetric exponential power distribution*. LEM Working Paper Series 2008/19, Sant'Anna School of Advanced Studies, Pisa.

Chiodi, M. (1988). *Sulle distribuzioni di campionamento delle stime di massima verosimiglianza dei parametri delle curve normali di ordine p*. Istituto di Statistica, Facoltà di Economia e Commercio di Palermo : Technical report.

Chiodi, M. (1995). Generation of pseudo random variates from a normal distribution of order *p*. *Statistica Applicata (Italian Journal of Applied Statistics)*, *7*(4), 401–416.

Christiano, L. J. (2007). Comment. *Journal of Business and Economic Statistics*, *25*(2), 143–151.

Davidson, R., & MacKinnon, J. G. (2004). *Econometric Theory and Methods*. Oxford: Oxford University Press.

De Grauwe, P. (2012). Booms and busts in economic activity. *Journal of Economic Behavior and Organization*, *83*(3), 484–501.

DeLong, J. B., & Summers, L. H. (1986). Are business cycles symmetrical? In R. J. Gordon (Ed.), *The American Business Cycle: Continuity and Change* (pp. 166–179). Chicago: University of Chicago Press.

Fagiolo, G., Napoletano, M., Piazza, M., & Roventini, A. (2009). Detrending and the distributional properties of U.S. output time series. *Economics Bulletin*, *29*(4), 3155–3161.

Fagiolo, G., Napoletano, M., & Roventini, A. (2008). Are output growth-rate distributions fat-tailed? Some evidence from OECD countries. *Journal of Applied Econometrics*, *23*, 639–669.

Fagiolo, G., Napoletano, M., & Roventini, A. (2007). How do output growth-rate distributions look like? Some cross-country, time-series evidence. *The European Physical Journal B*, *57*, 205–211.

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

Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, *6*(3), 255–259.

Li, S. (2011). *Regression models with the error term following the asymmetric exponential power distribution*. Mimeo: Department of Economics, Rutgers University.

Lobato, I. N., & Velasco, C. (2004). A simple test of normality for time series. *Econometric Theory*, *20*(4), 671–689.

McConnell, M. M., & Perez-Quiros, G. (2000). Output fluctuations in the United States: What has changed since the early 1980's? *American Economic Review*, *90*(5), 1464–1476.

Mineo, A. M. (1994).*Un nuovo metodo di stima di p per una corretta valutazione dei parametri di intensità e di scala di una curva normale di ordine p* (pp. 147–154). CISU, Roma: Atti della XXXVII Riunione Scientifica della Società Italiana di Statistica.

Mineo, A. M. (2003). On the estimation of the structure parameter of a normal distribution of order *p*. *Statistica*, *63*(1), 109–122.

Mineo, A. M., & Ruggieri, M. (2005). A software tool for the exponential power distribution: The normalp package. *Journal of Statistical Software*, *12*(4), 1–24.

Ruge-Murcia, F. (2012). Estimating nonlinear DSGE models by the simulated method of moments: With an application to business cycles. *Journal of Economic Dynamics and Control*, *36*(6), 912–938.

Vavra, M., & Psaradakis, H. (2011). *Testing normality in time series*. Mimeo: Birkbeck College, University of London.

Zhu, D., & Zinde-Walsh, V. (2009). Properties and estimation of asymmetric exponential power distribution. *Journal of Econometrics*, *148*(1), 86–99.

Part III
Financial Market Modelling

Heterogeneous Beliefs and Quote Transparency in an Order-Driven Market

Polina Kovaleva and Giulia Iori

1 Introduction

It has been repeatedly shown in the market microstructure literature that trend-following strategies are common among traders. As Chiarella et al. (2009) have vividly demonstrated, this behaviour of heterogeneous traders contributes substantially to fat-tailed distribution of asset returns and long-term memory in volatility—the phenomena that frequently emerges from empirical market data. In addition, persistent patterns in order flow have been documented by Biais et al. (1995); Bouchaud et al. (2002) and Ranaldo (2004). These market features, however, undermine market efficiency hypothesis and thus require thorough investigation of their origins. The aim of this chapter is to explore the role of order book transparency in the persistence of these stylised properties using an artificial double auction set-up.

This chapter contributes to the research on determinants of abnormal returns and long-range dependencies in market data on the one hand, and the repercussions of market transparency on the other (Lo 1991; Bouchaud et al. 2002; Lillo and Farmer 2004). In an extension of Chiarella et al. (2009) agent-based framework we associate the chartist principle with the intertemporal demand and supply recorded in the limit order book. We distinguish between three quote transparency regimes and evaluate the consequent traders' interactions. A related study has been realised by Yamamoto (2011) who includes both chartist component based on returns and a probabilistic mechanism that reproduces the reaction of traders to order book imbalances. In contrast with the latter, our experiments suggest that order aggressiveness switching mechanism implied by our model does not arise endogenously in a dark market.

P. Kovaleva (✉) · G. Iori
Department of Economics, City University, Northampton Square, London EC1V 0HB, UK
e-mail: polina.kovaleva.1@city.ac.uk

G. Iori
e-mail: g.iori@city.ac.uk

R. Dieci et al. (eds.), *Nonlinear Economic Dynamics and Financial Modelling*,
DOI: 10.1007/978-3-319-07470-2_10,

In this chapter, we examine the role of transparency in generating stylised facts from financial markets. The repercussions of quote transparency on market liquidity and resilience, bid-ask spread and the price discovery process are addressed in a companion paper (Kovaleva and Iori 2014).

The remainder of this chapter proceeds as follows. Section 2 outlines the expectations mechanism that traders use to formulate their orders. Section 3 defines market specifications and provides the basic overview of market dynamics. Section 4 evaluates the capability of our artificial market to reproduce empirical distribution of order flows and asset returns for different degrees of transparency. Section 5 concludes and points out the direction for future research.

2 The Model

Our model builds on the framework of Chiarella et al. (2009). We assume a double auction market, for a single non-dividend paying stock, where N_A heterogeneous agents trade during a repeated number of rounds. In each time period t a random agent i is called to trade. This agent formulates his order placement strategy by (i) relating the current price of the asset to its fundamental value, (ii) analysing information about the visible market depth and (iii) assuming some random component in the asset returns. Traders are fundamentalists and expect in the long run (i.e. over a period τ^f) the market to revert to trading at the fundamental price of the asset p_t^f. In the short run traders take into account the visible order imbalance to evaluate in which direction the market is likely to move. The visible order imbalance D_{lt} here is defined as the signed log difference between the total amount of stocks demanded Q_{lt}^b and offered Q_{lt}^a at l best quotes at time t

$$D_{lt} = \operatorname{sgn}\left(Q_{lt}^a - Q_{lt}^b\right) \cdot \ln\left(1 - \frac{\left|Q_{lt}^a - Q_{lt}^b\right|}{Q_{lt}^a + Q_{lt}^b}\right). \tag{1}$$

If the combined volume of orders to buy is higher than the combined volume of orders to sell, that is $D_{lt} > 0$, the trader expects a price increase, and vice versa. We impose that the number of visible orders l, that traders use to assess the gap between the demand and supply, describes the degree of market transparency.[1] In our analysis, we will distinguish between different levels of quote transparency $l = \{\infty, 5, 0\}$.

Finally, the trader faces an aggregate uncertainty factor ξ_t that we assume to be normally distributed with zero mean and variance σ_ξ^2, $\xi \sim N(0, \sigma_\xi^2)$. Overall, trader

[1] Although in reality traders may extract additional value from the information on the visible quotes, such as the ID of potential counterparties and guess any hidden volumes, we show below that even tracking the buy-sell imbalance alone generates some interesting patterns.

i builds his expectation of the return on the asset that can be achieved within his time horizon τ^i according to the following rule:

$$\hat{r}^i_{t,t+\tau^i} = \frac{1}{\nu^i_1 + \nu^i_2 + \eta^i} \left[\nu^i_1 \frac{\ln(p^f_t / p_t)}{\tau^f} + \nu^i_2 \bar{D}_{lt} + \eta^i \xi_t \right], \tag{2}$$

where $\bar{D}$ is the average order imbalance over the past τ^i periods, and the coefficients ν^i_1 , ν^i_2 and η^i guide fundamentalist, imbalance and noise components of trader i's strategy, respectively. The projected rate of return $\hat{r}^i_{t,t+\tau^i}$ yields the maximum expected price over τ^i periods

$$\hat{p}^i_{t+\tau^i} = p_t \exp(\hat{r}^i_{t,t+\tau^i} \tau^i). \tag{3}$$

We assume that every trader in the market is risk-averse and holds a portfolio of stocks S^i_t and cash C^i_t, so that the wealth of trader i at time t is $W^i_t = S^i_t p_t + C^i_t$.[2] The optimal composition of the portfolio is determined via the maximisation of a negative exponential utility function with a constant absolute risk-aversion $\varphi^i = \varphi(1 + \nu^i_1)/(1 + \nu^i_2)$, so that trader i's demand for the stock at time t is given by:

$$\pi^i_t(p) = \frac{\ln(\hat{p}^i_{t+\tau^i}/p)}{\varphi^i V^i_t p}, \tag{4}$$

where V^i_t is the variance of the spot returns $V^i_t = \frac{1}{\tau^i} \sum_{j=1}^{\tau^i} (r_{t-j} - \bar{r}^i_t)^2$ with the mean spot return $\bar{r}^i_t = \frac{1}{\tau^i} \sum_{j=1}^{\tau^i} r_{t-j} = \frac{1}{\tau^i} \sum_{j=1}^{\tau^i} \ln\left(\frac{p_{t-j}}{p_{t-j-1}}\right)$. There exists price p^*, a satisfaction level, that makes the current portfolio composition optimal, i.e. $\pi^i_t(p^*) = S^i_t$. If the current market price of the stock is above this level, then the trader is willing to sell, otherwise, the trader intends to buy more stocks.

At each trading round the agent who is called to trade computes his range of admissible prices. We assume that short selling and borrowing are not allowed. The maximum, p_M, and minimum, p_m, prices at which a trader can place his order are determined by his budget constraint. The trader is boundedly rational: he picks randomly a price p from the current range of admissible prices. Then, depending on the location of this price p relative to the satisfaction price p^* and the best bid b_t and ask a_t quoted in the limit order book, the trader formulates his order. The possible scenarios of the order placement are sketched in Fig. 1 with the precise order classification in Table 1.

[2] This model bears close resemblance to the framework described in Chiarella et al. (2009) with the imbalance factor in place of chartist. For this reason we only delineate here the key aspects of the traders' decision-making routine, highlighting the role of order book imbalance. Interested readers are referred to Chiarella et al. (2009) for the demand function derivations and further details on order placement mechanism.

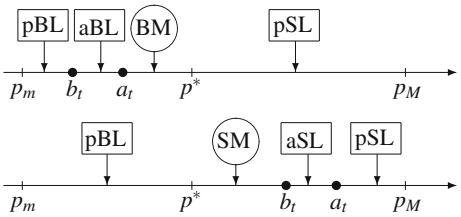

Fig. 1 The topology of order submissions: $[p_m, p_M]$ is the range of admissible prices, p^* is the satisfaction price, and b_t and a_t are the current best bid and ask prices in the market, respectively

Table 1 The resulting order submission for a randomly picked price $p \in [p_m, p_M]$ for tick size Δ

Buy				Sell			
Type	Interval	Price	Size	Type	Interval	Price	Size
BM	$[\max(a_t, p_m + \Delta); p^*]$	a_t	$\pi_t^i(a_t) - S_t^i$	SM	$[p^*; \min(b_t, p_M)]$	b_t	$S_t^i - \pi_t^i(b_t)$
aBL	$(\max(b_t, p_m);$ $\min(a_t, p^*))$	p	$\pi_t^i(p) - S_t^i$	aSL	$(\min(b_t, p^*);$ $\min(a_t, p_M))$	p	$S_t^i - \pi_t^i(p)$
pBL	$[p_m; \min(b_t, p^* - \Delta)]$	p	$\pi_t^i(p) - S_t^i$	pSL	$[\max(a_t, p^* + \Delta); p_M]$	p	$S_t^i - \pi_t^i(p)$

In our simulations, we distinguish between six classes of orders: a buy market order (BM), or an intention to buy the asset at a price higher or equal to the current best ask; a sell market order (SM), or an intention to sell the asset at a price lower or equal to the current best bid; aggressive buy (aBL) and aggressive sell (aSL) limit orders placed inside the current bid-ask spread; passive buy limit orders placed below the current best bid (pBL), and passive sell limit orders placed above the best ask (pSL).

3 Order Book Simulations

In this section, we outline the main features of the simulations and provide a summary of numerical parameters. The core objective of this chapter is to study the emergent properties in stock returns and order flow arising from the restricted access to market depth information imposed on heterogeneous traders. In the simulation analysis below we contrast three market specifications with full, intermediate and no quote visibility.

We define two extreme cases: *a transparent market* ($l = \infty$), where traders are aware of full order sizes and use entire order book to evaluate imbalance D_{lt}, and *a dark market* ($l = 0$), where market depth is not publicly revealed and traders have no resources to estimate the order imbalance.[3] The latter specification serves as a benchmark in our analysis. Naturally, in a transparent market agents observe the depth of the entire limit order book before trading. While in certain markets such an assumption is perfectly viable (e.g. NYSE OpenBook), many other exchanges

[3] All agents have a zero weight of imbalance component $\nu_2^i = 0, \forall i$ in Eq. (2).

(e.g. Tokyo Stock Exchange, Euronext Paris) confine the available information to the five best quotes on each side of the limit order book at any time. As the third environment, we consider *a quasi-transparent market* ($l = 5$), which accommodates this restriction, but otherwise traders formulate their strategies according to the same principle as in a transparent market.

We adopt the simulation parameters that match the baseline values in Chiarella et al. (2009). There are $N_A = 5000$ agents, who randomly arrive to the market and execute trades according to time and price priority protocol. Their initial cash and stock endowments are uniformly distributed on $[0, 50]$ and $[0, 50p_0]$, respectively. When called to trade, a trader formulates his strategy and overwrites his previous unfilled order recorded in the book if any. The common risk aversion factor among traders that participate in this market is 0.1. The trading horizons of agents are driven by $\tau = 200$, so that the average horizon is approximately two trading days τ_d of 100 trading sessions each. The stock initially trades at $p_0 = £400$ above its fundamental value $p_0^f = £300$ and the price grid is determined by the minimum tick size $\Delta = £0.005$. We assume that the fundamental asset value follows a Wiener process with zero drift and volatility σ_ξ, where $\sigma_\xi = 0.001$ is the dispersion of the Gaussian noise term ξ_t. The expected mean reverting time to the fundamental value in $\tau^f = 20$ days. The fundamental price, order imbalance and noise coefficients are drawn from three independent exponential distributions for the entire population of traders: $\nu_1^i \sim \exp(1/\sigma_{\nu_1})$, $\nu_2^i \sim \exp(1/\sigma_{\nu_2})$, $\eta^i \sim \exp(1/\sigma_\eta)$, $\forall i = \overline{1, N_A}$, and remain constant over the course of trading. We select $\sigma_{\nu_1} = 0.1$, $\sigma_{\nu_2} = 0.0037$ and $\sigma_\eta = 0.01$ in the baseline.

First hand graphical analysis of the transaction prices generated in three trading environments (Fig. 2) reveals that transparency can introduce higher heterogeneity of beliefs and thereby enhance the deviation from the fundamental asset value. Indeed, there is a stark difference between price volatility in a dark market (Fig. 2c) compared to both transparent and quasi-transparent market. This outcome is in line with Verardo (2009) and Yamamoto (2011). The first two specifications, however, yield very similar results.

In contrast with the observations of Chiarella et al. (2009), we do not detect prolonged excursions of the transaction prices from the fundamental level, which implies that chartist indicator based on imbalance rather than returns trend stabilises market performance and assists price discovery. In addition, the returns series plotted in Fig. 3 suggest that both transparent and quasi-transparent markets frequently move to extremes.

4 Stylised Facts

Our point of interest is the ability of the market to reproduce phenomena frequently detected in financial markets. In particular, we examine how the degree of market transparency alters the likelihood of abnormal price fluctuations, path dependence in order submissions and other memory properties.

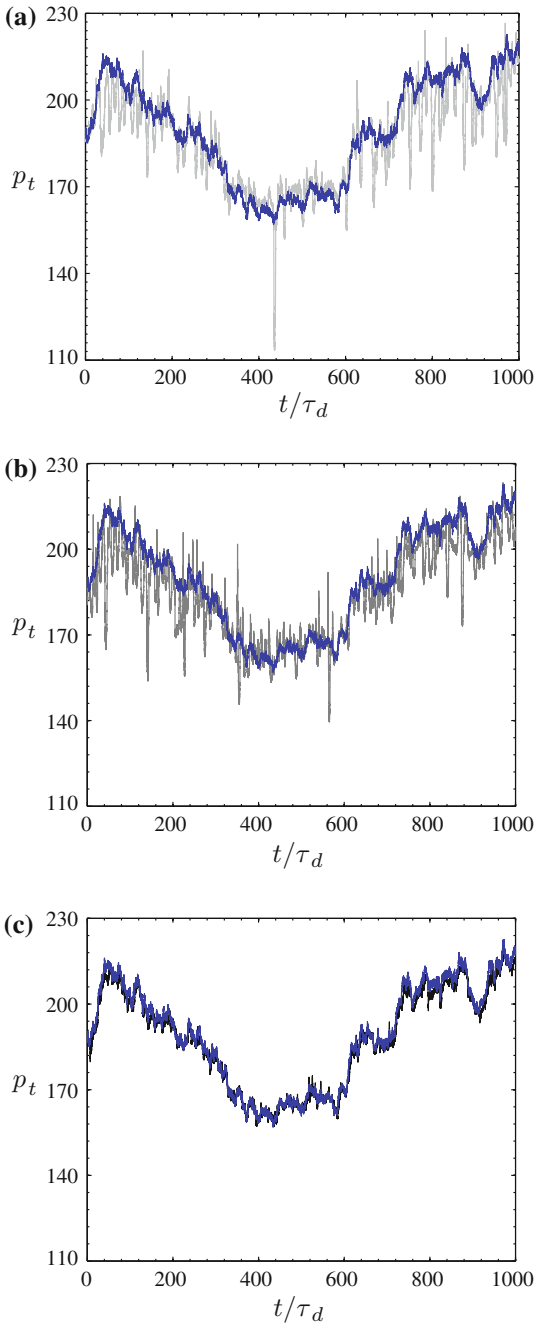

Fig. 2 Fundamental (*blue line*) and transaction price series sampled at daily frequency under three transparency regimes: **a** transparent, **b** quasi-transparent, **c** dark

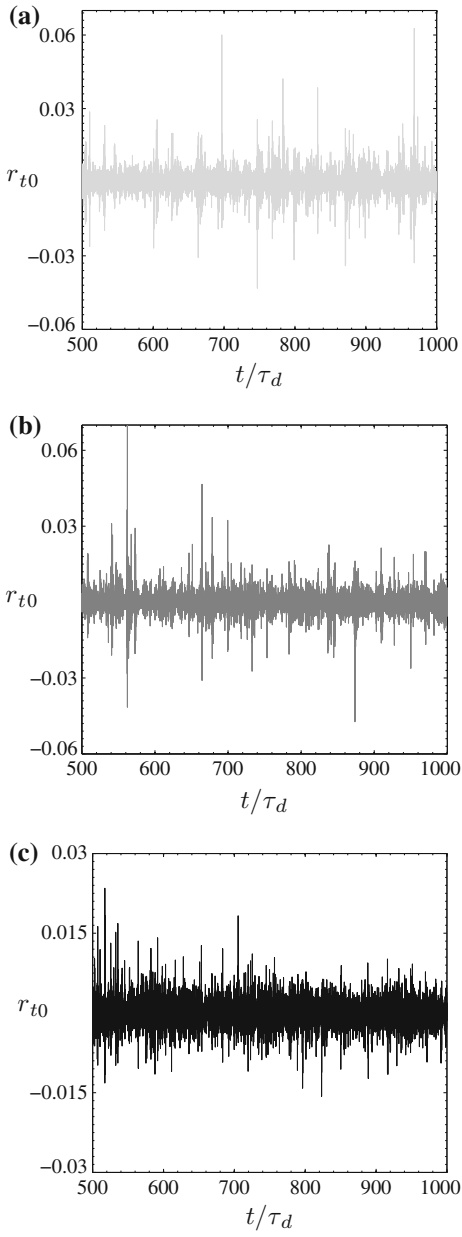

Fig. 3 Returns series sampled at daily frequency under three transparency regimes: **a** transparent, **b** quasi-transparent, **c** dark

4.1 Abnormal Returns

The mean return is effectively zero under all specifications, whereas the diapason of returns expands significantly once traders have the opportunity to account for a limit order book depth to form their expectations of the future stock price (Table 2).

Table 2 Summary statistics of returns and order flow under four transparency regimes

Variable		Transparency regime		
		Transparent	Quasi-transparent	Dark
Returns	Min	−0.0541	−0.0528	−0.0215
	Max	0.0500	0.0559	0.0246
	Mean	0.0000	0.0000	0.0000
	S.D.	0.0024	0.0024	0.0014
	Skewness	−0.1527	0.1850	0.2320
	Kurtosis	33.7944	34.9786	18.8458
Order Flow	No. of orders	96,982.5	97,005.3	96,966.1
	Average order size	24.04	24.11	24.30
	Average BM	28.10	28.10	29.20
	Average BL	23.18	22.95	22.18
	Average SM	30.41	30.67	32.05
	Average SL	19.66	19.61	19.38

The results represent averages over 100 simulation runs

Analogously to the model of Chiarella et al. (2009), the reason for large price deviation from the fundamental level in both the partially and fully transparent cases is the antagonistic effect of fundamentalist and imbalance components that override traders' expectations once a big price change occurs. Both transparent and quasi-transparent regimes lead to much greater amplitude of abnormal returns than a dark regime. Of the former two regimes, the transparent microstructure produces marginally smaller price swings, as depicted in Fig. 2 as well as from the differences between the maximum and minimum return in Table 2.

Though on average it is impossible to profit from trading in any of the considered markets, the quasi-transparent market offers the highest, albeit still very close to fully transparent case, probability of receiving abnormal returns as indicated by kurtosis. Notably, only the transparent market embraces two empirical facts in the financial markets: asymmetric chances of receiving positive versus negative returns and non-trivial probability of abnormal returns. This is a convenient result since the model of Chiarella et al. (2009), with chartist impact in place of order imbalance impact, could not replicate the empirical negative skewness of returns.

4.2 Order Placement Patterns

In order to assess the coherence of our artificial stock market we investigate order flow properties documented in the abundant research on this subject, including, among others, the seminal paper of Biais et al. (1995) and empirical studies of Bouchaud et al. (2002); Ranaldo (2004) and Lo and Sapp (2005) on the determinants of order aggressiveness.

The general observation about the order placement is that the availability of information about book depth has no discernible effect on order flow in a transparent or a

Table 3 Unconditional order type probabilities from the artificial market, Australian Stock Exchange and Paris Bourse

Order type	Simulation results			Australian Stock Exchange			
	Transparent	Quasi-transparent	Dark	BHP	NCP	TLS	Paris Bourse
BM	15.89	16.06	16.61	27.72	15.91	23.99	19.03
aBL	9.17	9.40	11.22	5.64	3.58	1.79	10.10
pBL	25.74	25.80	23.50	24.70	29.17	25.72	14.52
SM	15.51	16.09	17.09	18.28	13.81	22.40	34.23
aSL	9.18	9.46	11.06	4.28	4.52	1.29	9.37
pSL	24.51	23.19	20.53	19.38	33.01	24.81	12.96

The simulation results represent averages over 100 simulation runs

quasi-transparent market in that neither the cumulative number of submitted orders, nor the average order size changes substantially. As indicated in Table 2, under these two transparency regimes the average size of aggressive market orders diminishes, albeit insignificantly, and is accompanied by a marginal increase in the average size of passive limit orders relative to the order sizes in the dark market.

In terms of unconditional probabilities, as shown in the right column in Table 3, most of the traders are liquidity providers in the environments with and without market depth information. Notice, however, that the order flow is more symmetric in the transparent case. Concerning the unconditional probabilities of six order types in a quasi-transparent market, we compare the distribution to the empirical data from Paris Bourse in Biais et al. (1995) and the Australian Stock Exchange in Hall and Hautsch (2006), both of which allow reduced quote transparency. In the data from Paris Bourse, the sellers, and most impatient ones, outnumber the buyers. On the contrary, out of five stocks from the Australian Stock Exchange analysed by Hall and Hautsch (2006) buyers dominate in two markets compared to fairly symmetric order flows in the other three, which we reproduce in Table 3 for our order classification. In terms of order flow pattern and overall order flow aggressiveness the simulated markets bear the closest resemblance to the market for News Corporation shares (NCP): the proportions of buy and sell orders are symmetric with a large fraction of passive limit orders. In their recent study of 100 stocks listed on INET in 2004, Hasbrouck and Saar (2009) find that not only limit orders compose the bulkiest part of the order flow, but also tend to have larger size than market orders. However, in many instances limit orders are cancelled within seconds after submission, and these fleeting, or flash orders usually exceed in size the limit orders that remain patiently in the book awaiting execution. This empirical evidence overturns the traditional treatment of limit orders as passive liquidity provision strategies. Instead, such application of big limit orders embodies aggressive liquidity-seeking behaviour that depletes the order book and harms market liquidity. Therefore, given the bias of fleeting orders, it is hard to retrieve the strength of liquidity streams from empirical data.

Further, the estimates in Table 3 indicate that the probability of market orders in a quasi-transparent market increases accompanied by a small increase in the probability of orders inside the spread, that is, aggressive limit buys (aBL) and sells (aSL). In the study of Yamamoto (2011), where traders react to the order imbalance by means

of a probabilistic switching mechanism, once the information about order volumes recorded in the book becomes limited, the use of market orders slightly escalates in comparison to a perfect transparency case. Similarly, in our framework, once traders are constrained to observe only the best five quotes, market orders accrue a slightly higher share in the order flow relative to perfect transparency. This partially explains a little revival in trading activity: a higher number of marketable orders in a quasi-transparent environment leads to a faster turnaround of assets.

We examine the order submission patterns to verify whether traders respond to the observable state of the limit order book in line with the empirical inference. Using order flow and transaction data from the Swiss Stock Exchange, a pure order-driven electronic stock market, Ranaldo (2004) consolidated findings on the determinants of order aggressiveness. In relation to the main variables in our model, we focus on the following properties identified in Ranaldo (2004):

P1: The wider the spread is, the weaker will be the order aggressiveness.
P2: The higher the volatility is, the stronger will be the order aggressiveness.
P3a: The thicker the book on the buy (sell) side is, the stronger will be the order aggressiveness of the incoming buyer (seller).
P3b: The thicker the book on the sell (buy) side is, the weaker will be the order aggressiveness of the incoming buyer (seller).

In order to validate the capability of the artificial market to reproduce these properties on aggregate, we compute the frequencies of all six order types submissions given the value of the corresponding market parameter. For instance, for the bid-ask spread variable we first count the frequencies of all order types submitted when the current spread is tight, then order frequencies when the spread at the time of submission is wide. Small and large sizes of the spread are defined against the average spread. In each simulation run we calculate the probability of submitting a concrete order when the parameter, e.g. the spread, is high minus the probability of the same order submission when the value of this parameter is low. Therefore, a positive number in Table 4 implies a higher probability of a given order type for a larger value of the corresponding market parameter, i.e. spread, volatility and depth; a negative number—a lower probability.

As reported in Table 4, we recover P1 and detect a negative correlation between order aggressiveness and the spread size across all markets independently of the transparency regime: as the spread widens, traders rely more on limit orders (the probability of orders inside the spread increases, and outside the spread decreases) and less on market orders. Condition P2 holds for sellers in all four cases, but not for buyers. The asymmetry between the behaviour of patient buyers and sellers in respect to volatility is caused by the CARA preferences.[4] Conditions P3a and P3b,

[4] Higher volatility does not affect the expectations of the trader directly, but alters the region of admissible prices defined by p_m and p^*. The subinterval of buy prices becomes shorter than the subinterval of sell prices proportional to the total interval length $p_M - p_m$ when the spot volatility is high. Consequently, it is more likely that the current market price of the security lies in the subinterval of selling prices (above p^*), and if the trader intends to purchase the asset, he submits a passive buy limit order.

Table 4 The changes in the probabilities of order types (in %) conditional on the market characteristics given an increase in the corresponding market parameter

Factor	Order type					
Transparent	BM	aBL	pBL	SM	aSL	pSL
Spread	−4.96	9.72	−4.65	−4.31	9.84	−5.65
Volatility	−3.80	0.43	3.51	1.07	1.35	−2.57
Buy depth	4.77	−0.65	−3.77	−4.57	−1.74	5.96
Sell depth	−4.30	−1.76	5.97	4.65	−0.75	−3.82
Quasi-transparent						
Spread	−5.54	9.92	−4.40	−4.15	10.37	−6.22
Volatility	−4.13	0.31	3.81	1.28	1.58	−2.85
Buy depth	4.55	−0.40	−3.81	−4.07	−1.43	5.16
Sell depth	−4.36	−2.04	6.36	4.53	−1.24	−3.25
Dark						
Spread	−6.29	12.45	−6.21	−5.08	12.38	−7.24
Volatility	−3.80	0.63	3.06	1.47	1.47	−2.83
Buy depth	1.12	−1.27	0.23	0.32	−1.47	1.08
Sell depth	2.28	−1.70	−0.54	−0.69	−2.03	2.68

The results represent averages over 100 simulation runs

evidently, hold only when at least some market depth information is disclosed to traders. For instance, both in transparent and quasi-transparent markets when the total supply in the book, i.e. sell depth increases, the probability of an incoming buy market orders and sell limit orders declines, whereas the buy limit order and sell market order submissions become more likely. In this sense, accounting for market depth provokes a feedback effect. In other words, in the transparent and quasi-transparent cases, an incoming buyer has a higher probability to use a market order when competition is tight, while an incoming seller has an increased chance of trading via a passive limit order, as the demand for the asset increases.

Another aspect that characterises markets is path dependences in order submissions. We analyse the order flow pattern using an approach similar to Biais et al. (1995). In their empirical study of Paris Bourse data, Biais et al. (1995) uncovered a number of important properties including the so-called *diagonal effect*, whereabout orders of the same type tend to follow each other, so that the diagonal elements of the conditional distribution matrix have the highest weights. Typical conditional distributions of orders in the four markets, including Biais et al. (1995) estimates, are given in Table 5. Each estimate in this table stands for the number of times that the order type in row i was submitted after the order type in column j throughout 1,000 trading days. Dividing frequencies in each row by the total number of orders of the corresponding type, we convert to probabilities and then calculate a percentage deviation of a conditional probability of a particular order type from the unconditional probability of this order type. The last manipulation ensures comparability of the diagonal effect strength across specifications. The same transformation is applied to the Biais et al. (1995) data once their order types are regrouped to match our

Table 5 Percentage deviations of order type probabilities conditional on the previous order type from their unconditional probabilities

Order type	Previous order type					
	BM	aBL	pBL	SM	aSL	pSL
Transparent						
BM	**35.0**	**3.5**	−12.3	−39.0	−2.3	**14.5**
aBL	33.6	−6.2	−11.9	2.1	−9.0	−4.9
pBL	−37.6	1.6	**14.1**	24.6	6.6	−9.0
SM	−42.6	2.5	13.3	**35.1**	**9.3**	−13.1
aSL	6.3	−5.2	−3.6	26.7	−6.7	−12.8
pSL	23.0	2.2	−7.4	−32.3	−0.6	12.7
Quasi-transparent						
BM	**35.9**	**5.2**	−13.6	−40.7	−0.8	**16.7**
aBL	30.8	−5.5	−11.9	3.0	−7.3	−4.9
pBL	−36.4	0.3	**14.2**	24.4	5.0	−9.7
SM	−41.4	0.7	13.9	**33.8**	**6.7**	−13.3
aSL	5.6	−5.5	−3.3	26.4	−5.4	−14.1
pSL	26.0	4.1	−9.2	−35.8	0.3	15.3
Dark						
BM	**28.0**	2.5	−8.5	−35.4	1.6	14.3
aBL	24.4	−1.9	−10.4	3.7	−4.8	−7.3
pBL	−34.6	2.4	**11.9**	22.2	**4.2**	−7.7
SM	−34.8	0.7	10.5	**26.3**	3.9	−8.3
aSL	6.1	−0.0	−4.8	19.4	−4.5	−13.1
pSL	25.2	**3.2**	−8.1	−36.8	3.2	**16.1**

(continued)

Table 5 (continued)

Order type	Previous order type					
	BM	aBL	pBL	SM	aSL	pSL
Biais et al. (1995)						
BM	**57.5**	3.6	−3.7	−20.2	−31.0	−8.8
aBL	9.8	**52.7**	25.4	−22.1	0.1	−27.2
pBL	−3.6	23.9	**62.4**	−17.5	−22.2	−22.6
SM	−19.0	−24.7	−20.0	**33.3**	−4.1	−17.2
aSL	−18.7	5.1	−21.9	−2.3	**51.5**	15.4
pSL	−11.6	−24.1	−22.5	−12.2	9.1	**85.1**

The results represent averages over 100 simulation runs

six types. Along each column, the most likely order sequence is printed in bold. Although the diagonal effect may not be a universal property of real markets, it is insightful to verify its relation to the degree of quote transparency. The described evaluation method proves convenient for this purpose.

Overall, the diagonal effect is dissipated in our artificial marketplaces, unlike the clear pattern in Biais et al. (1995) shown in the last panel in Table 5. This said, we observe significant path dependence in market orders equally for buyers and sellers, which is further facilitated when market transparency is higher. Series of passive limit orders are slightly more rare, though such orders are still more likely to arrive in batches rather than independently. In fact, empirical (Biais et al. 1995) and experimental (Majois 2010) evidence reveals that most frequent sequences can appear off the diagonal with a finer order classification. Chordia et al. (2002) document that positive returns tend to be continued, whereas negative returns most frequently provoke market reversals. This evidence corresponds to a weakening diagonal effect in the domain of sellers: while buys—especially aggressive market orders—tend to arrive one after the other and push the transaction price up, sell orders appear to be less likely to cluster in this manner under transparent and quasi-transparent regimes. The differences in estimated conditional order distributions in Table 5 suggest that quote transparency generates significantly stronger regularities in market order submissions, resulting in larger price impacts.

The justification of the diagonal effect, according to Biais et al. (1995), arises from imitation among traders, strategic order splitting, or similar expectations. Although disentangling these origins is a challenge, in a series of classroom experiments Majois (2010) concludes that the diagonal effect stems from order splitting. Any conscious imitation or strategic behaviour is redundant in our setting, hence the sole reason for traders to copy each other is some kind of herding mechanism, enabled by matching expectations. All traders possess the same market depth information and have identical fundamental asset valuation. On the other hand, an order placed by the previous trader always affects the expectations of the one arriving after him by changing the depth and possibly the distance to the fundamental price. In the case with visible order imbalance all traders see the book and become more correlated in their actions, resulting in convergent future price expectations and order placements across agents. Therefore, in the absence of intentional imitation between market participants this framework captures the similarities in traders' order placements that translate into more pronounced diagonal effect for higher transparency. Furthermore, this artificial market reproduces fairly symmetric conditional distribution of order types alike Paris Bourse data of Biais et al. (1995).

4.3 Memory Properties

Memory in the financial time series stipulates predictability of future market dynamics, thereby creating a relevant argument for the discussion of market efficiency. We investigate the memory effects in key variables, which have an immediate impact on agents' decision-making outcomes: returns, absolute returns and order imbalance.

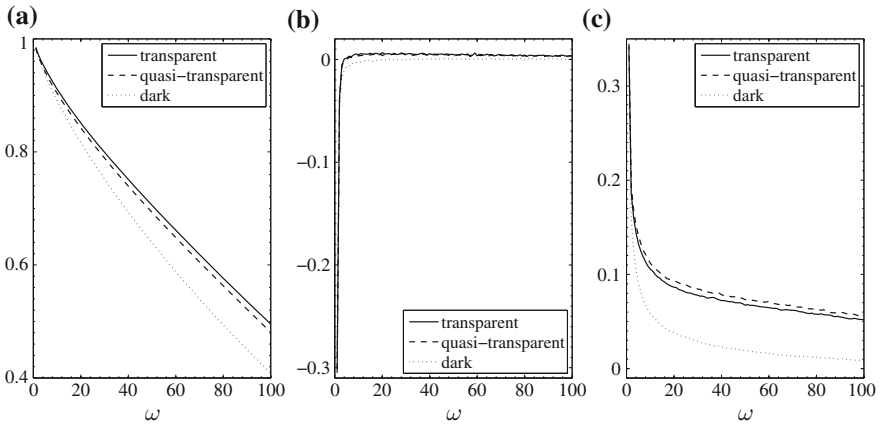

Fig. 4 Autocorrelation functions of **a** order imbalance, **b** returns, and **c** absolute returns series estimated as averages over 100 simulation runs for each of the four specifications

Figure 4 depicts the autocorrelation functions in returns and imbalance for three model specifications. The total order imbalance variable pertains a strong memory, as Fig. 4a demonstrates. This observation is not surprising. It confirms that the order book depth is sustainable and a single trade does not reverse or change significantly the imbalance between buy and sell sides. The memory is most persistent for the transparent market, but decays rather quickly in the dark market. Reducing market transparency helps to diminish the long-term path dependence in the order imbalance. Based on Ljung-Box test we find that serial autocorrelations in absolute returns and in order imbalance remain statistically different from zero for all 100 included lags and further. The simulation outcomes of Yamamoto (2011) also confirm substantial persistence in order imbalance both for transparent and quasi-transparent market set-ups. Moreover, Lillo and Farmer (2004) detect long memory in the order flow, particularly in the signs of submitted orders, for a range of stocks listed on the London Stock Exchange and sampling years. This result is connected to our artificial markets in two ways. First, it justifies the autocorrelations in Fig. 4a. Consider, for example, successive arrival of sell orders: limit orders add depth to the sell side and increase the order book imbalance, as defined in Eq. (1), market orders absorb the volume on the buy side which again has a positive impact on D_{lt}. Second, the order signs memoryness echoes the discussion of the diagonal effect in submitted order sequences. In their review of the statistical evidence on long memory in order flow Mike and Farmer (2008) argue that long memory property in order signs has a profound impact on price dynamics and induces a power law distribution with high probabilities of extreme returns. They explain it by order splitting and other dynamically optimised strategies, whereas in our set-up it is rather the unconscious herding caused by matching price expectations that is responsible for the correlation in buy and sell volumes.

On the contrary, Fig 4b shows that there is no persistent correlation in returns apart from the first few lags, which implies informational efficiency regardless of the degree of market transparency. The negative autocorrelation in returns for small lags is caused by the bid-ask bounce, generated by the impatient traders who demand liquidity. This effect is the smallest for the dark market, which is explained by the thicker limit order book in this set-up. The correlations beyond lag 4 are not economically significant, as the autocorrelations are smaller than the size of the spread, and therefore cannot be exploited profitably. Naturally, the memory in returns dissipates most quickly in the dark market since no trend-following mechanism is implicit in traders' behaviour in this case. Regarding the autocorrelation in absolute returns that is usually interpreted as a proxy of volatility, Fig. 4c shows a positive dependence in the absolute returns even for distant lags. Overall, the higher the transparency is, the stronger will be the memory in absolute returns. However, the impact of restricting the info only to the five best quotes is somewhat ambiguous since it intensifies the long memory in absolute returns compared to the full depth visibility case. The combination of uncorrelated returns and significant positive dependence in absolute returns quantifies the volatility clustering.[5]

Further, we apply the test that discards conveniently short period memory and concentrate mainly on long-term correlations. We calculate the modified rescaled range statistic, which is more appropriate in the context of stark departures from the normal distribution in the data:

$$\tilde{Q}_{x,n}(\omega) = \frac{1}{\sqrt{V_n(\omega)}} \left[\max_{1 \le k \le n} \sum_{i_k=1}^{k} \left(x_{i_k} - \bar{x}_n \right) - \min_{1 \le k \le n} \sum_{i_k=1}^{k} \left(x_{i_k} - \bar{x}_n \right) \right], \quad (5)$$

where $n < N$ is the subsample size drawn from N-observations sample, $\bar{x}_n$ is the subsample mean, $V_n(\omega)$ is the subsample autocovariance for lag ω. This statistic is adjusted to the subsample size to obtain the rescaled range coefficient $\beta_{x,n}(\omega) = \ln \tilde{Q}_{x,n}(\omega) / \ln n$. In case $\beta_n(\omega) < 0.5$, there is a mean reversion in the data series. If $\beta_n(\omega) > 0.5$ there is a positive correlation in the data and a long memory property.

Figure 5 depicts rescaled range coefficients for asset returns and absolute returns. First, we monitor the evolution of memoryness in absolute returns for various time horizons. The strong dependence in absolute returns plotted in Fig. 5a supports the volatility clustering phenomena from real financial markets: there is a slow decay in the serial correlation of the volatility of asset returns. As more lags ω are omitted, the coefficient $\beta_n(\omega)$ decreases and slowly approaches the benchmark level 0.5, which indicates that memory of distant past absolute returns taken in isolation fades away, hence the spot volatility of returns is independent in the long term. In other words, periods of high volatility in returns are followed by periods of low volatility.

[5] In terms of graphical analysis, Fig. 3a and b depict clustering in returns series under the transparent and quasi-transparent regimes, compared to the more monotonous variability of returns in the dark market in Fig. 3c.

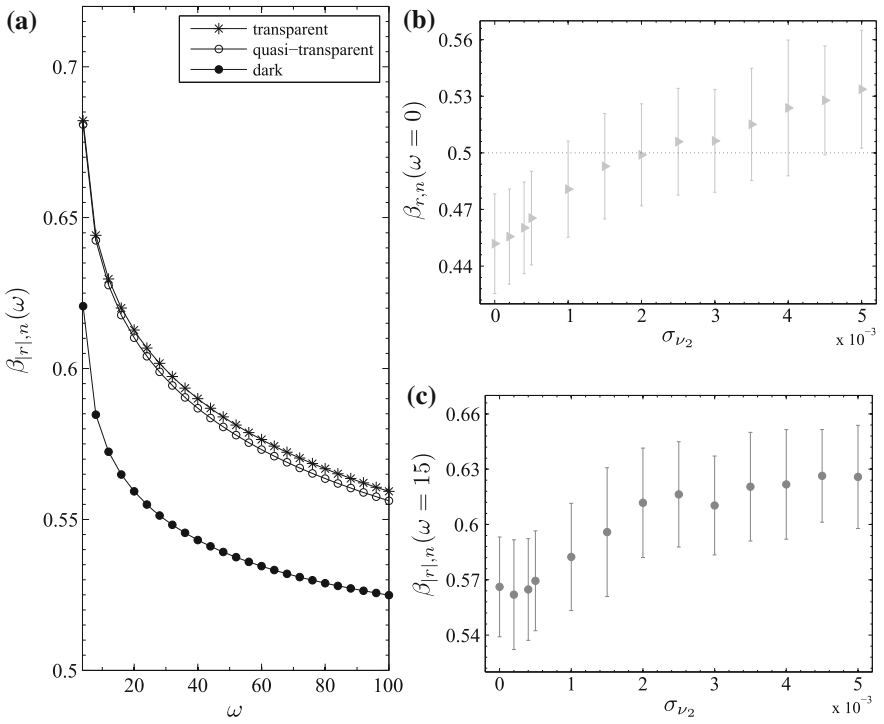

Fig. 5 Modified rescaled range test coefficient of **a** absolute returns for various lags (error bars computed over 100 simulations), **b** absolute returns in a transparent market with the optimal number of discarded lags, and **c** returns in a transparent market with all lags included. The subsample size is 5,000 time steps

The clustering is very poor in the dark market, which is indicated by low values of $\beta_{|r|,n}(\omega)$ even for small lags, in contrast with transparency regimes where all or part of the book volume is revealed to traders.

Next, we investigate how the strength σ_{ν_2} of order imbalance component in traders' price expectations affects the memoryness of returns and absolute returns.[6] Since the preliminary autocorrelation analysis revealed that the path dependence in returns is short-lived, we calculate the rescaled range coefficients without omitting any lags. Fig. 5b depicts the standard rescaled range coefficient for asset returns with all lags included as a function of imbalance impact σ_{ν_2}. We observe that when order imbalance impact is meagre, the returns clearly exhibit a mean reversion pattern. This includes the case of a dark market which corresponds to the first point on this plot with $\sigma_{\nu_2} = 0$. However, some positive correlation emerges as the standard deviation of imbalance weight of imbalance grows above 0.005. Regarding latent

[6] The results across the two market specifications with some quote visibility are reasonably similar; thus, we present just the estimates for the transparent case.

volatility measured by the absolute returns, the short-term dependencies converge to zero at the truncation lag $\omega = 15$, according to Lo (1991) criterion. The graph of modified rescaled range coefficient in Fig. 5c shows that even for low values of imbalance component there is a long memory effect, albeit moderate. Higher weight of imbalance in traders' price predictions, at the same time, translates into more persistent long range dependencies in the absolute returns, which match empirical findings Lo (1991). The inclusion of a sizable imbalance weight is therefore an important aspect of this model, given that it allows to replicate more accurately real market dynamics. In addition, this analysis highlights that the combination of baseline parameters for our artificial market implies memoryless returns (the confidence interval of $\beta_{r,5000}(\omega = 0)$ overlaps with 0.5 in this instance), and strong long-term correlations in absolute returns ($\beta_{|r|,5000}(\omega = 15) \approx 0.62$). The model is also robust with respect to reasonable changes in σ_{ν_2}, as follows from Fig. 5b and c.

5 Conclusion

This chapter conducted an analysis of the impact that varying degrees of order book depth transparency pertain for market performance. The evidence on benefits and adverse effects caused by limited market transparency is gained in the artificial double auction market simulations. Three distinct transparency regimes were included in our analysis: a fully observable limit order book—a transparent market, a market with visible orders restricted to five best quotes on each side of the book—a quasi-transparent market, and a market with market depth unknown to traders—a dark market.

The core implication of our agent-based model is that disclosing the information about order book depth deteriorates efficiency of order-driven markets, as indicated by the ascending memory properties of absolute returns. The transparency level imposed by certain exchanges whereupon only several best orders on each side of the limit order book are publicly displayed is controversial, since it further increases volatility clustering. At the same time, an ability to evaluate the cleavage between demand and supply available in the market gives the opportunity to earn higher return for traders with diverse preferences and risk perceptions.

It follows from the experiments reported in this chapter that market depth matters for shaping its future dynamics. Inclusion of order book imbalance variable into expectations allows to replicate the empirical conditional distribution of order flow. In contrast with Chiarella et al. (2009) results, we observe that following the trend in order book imbalance instead of past returns stabilises the price in the short term and reproduces the negative skewness feature of returns. In the future analysis, we intend to investigate how transparency regimes affect market quality indicators, such as transaction costs, liquidity and speed of trading. In addition, the range of transparency regimes can be expanded to incorporate more complex rules adopted in real exchanges.

References

Biais, B., Hillion, P., & Spatt, C. (1995). An empirical analysis of the limit order book and the order flow in the Paris Bourse. *The Journal of Finance*, *50*(5), 1655–1689.

Bouchaud, J.-P., Mezard, M., & Potters, M. (2002). Statistical properties of stock order books: empirical results and models. *Quantitative Finance*, *2*(4), 251–256.

Chiarella, C., Iori, G., & Perelló, J. (2009). The impact of heterogeneous trading rules on the limit order book and order flows. *Journal of Economic Dynamics & Control*, *33*(3), 525–537.

Chordia, T., Roll, R., & Subrahmanyam, A. (2002). Order imbalance, liquidity, and market returns. *Journal of Financial Economics*, *65*(1), 111–130.

Hall, A. D., & Hautsch, N. (2006). Order aggressiveness and order book dynamics. *Empirical Economics*, *30*(4), 973–1005.

Hasbrouck, J., & Saar, G. (2009). Technology and liquidity provision: The blurring of traditional definitions. *Journal of Financial Markets*, *12*(2), 143172.

Kovaleva, P., & Iori, G. (2014). The impact of reduced pre-trade transparency regimes on market quality.

Lillo, F., & Farmer, J. D. (2004). The long memory of the efficient market. *Studies in Nonlinear Dynamics & Econometrics*, *8*(3), 1–32.

Lo, A. J. (1991). Long-term memory in stock market prices. *Econometrica*, *59*(5), 1279–1313.

Lo, I., & Sapp, S. G. (2005). Order submission: The choice between limit and market orders. Working paper 2005–42, Bank of Canada.

Majois, C. (2010). Order aggressiveness and the diagonal effect in experimental double auction markets. *Economics Letters*, *107*(2), 304–309.

Mike, S., & Farmer, J. D. (2008). An empirical behavioral model of liquidity and volatility. *Journal of Economic Dynamics and Control*, *32*(1), 200–234.

Ranaldo, A. (2004). Order aggressiveness in limit order book markets. *Journal of Financial Markets*, *7*(1), 53–74.

Verardo, M. (2009). Heterogeneous beliefs and momentum profits. *Journal of Financial and Quantitative Analysis*, *44*(4), 795–822.

Yamamoto, R. (2011). Order aggressiveness, pre-trade transparency, and long memory in an order-driven market. *Journal of Economic Dynamics and Control*, *35*(11), 1938–1963.

The Simplicity of Optimal Trading in Order Book Markets

Daniel Ladley and Paolo Pellizzari

1 Introduction

How should a trader optimally execute a trade? As academic understanding of financial markets and the effect of their structure has grown this question has become more nuanced and sophisticated. In early models, markets were assumed to have a single price and react smoothly to changes in demand. In this context, the question of optimal trading was often one of timing—when should a trader trade. As these models became more sophisticated and market makers started to play a role, issues such as order splitting and information hiding came to the fore. More recently with the inclusion of architectures such as order books the question has acquired new facets—not just when should a trader trade, but also at what price and with what tool. Should a trader trade now with a market order? This guarantees trade at a specified, but potentially inferior, price. Or should they post a limit order in the belief that prices will improve and that greater returns will be made? The ability of traders to select the best order may potentially have a large effect on their profits. The size of this effect is increasing as algorithmic trading aimed at picking off inefficient submission becomes more common. The trader's choice will be contingent on their own information, but importantly also the state of their environment—the order book. How this information should be used and just which pieces are important, however, is an open question.

In this paper, we investigate the importance and effect of information on trading strategies and market dynamics. We draw conclusion from two models. The first permits continuous prices, that is, there is no minimum tick size, and trading strategies

D. Ladley (✉)
Department of Economics, University of Leicester, Leicester LE17RH, UK
e-mail: dl110@leicester.ac.uk

P. Pellizzari
Department of Economics, Ca' Foscari University, 30121 Venice, Italy
e-mail: paolop@unive.it

R. Dieci et al. (eds.), *Nonlinear Economic Dynamics and Financial Modelling*,
DOI: 10.1007/978-3-319-07470-2_11,

are conditioned on the prices of the best bid and ask quotes. In this model, strategies are optimised through the use of Evolution Strategies, an optimisation technique based on evolution and adaptation of the most profitable strategies. In the second model, traders submit orders on a discrete grid of ticks. Strategies are identified via the algorithm of Goettler, Parlour and Rajan (2005) ensuring that they are optimal for the specified game. We find that the amount of information traders use in their strategies has little effect either on the dynamics of the market or on the behaviour of the traders either under the optimal strategies or the linear approximations. We conclude that optimal trading strategies in a microstructure context may be simpler than believed and importantly may be characterised by a linear combination of the information available at the best quotes. Further, we conclude from this that models of financial markets do not need to concern themselves with interpreting the full information set available to traders strategies. Indeed, restricting consideration to the best quotes has little effect on results.

The dynamics of order book markets constitute complex situations through which traders interact. Traders and academics, when analysing or modelling these markets, are both faced with the task of combining large amounts of information to find an optimal strategy. One reason for this is the complexity of the environment—the amount of information available to traders in the book. Even under a Markov assumption—that the entire payoff-relevant history may be captured by the current state of the book—the information available is very substantial. Order books typically constitute price grids. At each discrete price there may be any quantity available to buy or sell (under the constraint that the highest buy price must be less than the lowest sell price). As a result the size of the information space is potentially infinite. Some of this information is undoubtedly more important than other pieces. Orders far from the best prices are unlikely to result in trades, and therefore are potentially less important. Their presence in the book, however, would have an effect on extreme price movements, and therefore may not be ignored. As such, different pieces of information will be more or less important and may have a smaller or larger effect on trading behaviour. Constructing the strategy—the optimal mapping between states and actions—in these markets is therefore a daunting task.

Since the relatively early stages of the academic microstructure literature models have been constructed in an attempt to do this. Frequently, however, this requires strong assumptions in order to maintain analytical tractability. For example Parlour (1998) considers a book of only four ticks in which two have infinite liquidity, while Rosu (2009) assumes continuous prices and time, permitting instantaneous revision of quotes. There have also been attempts to model these markets and trading strategies numerically. The simplest case is the literature on zero intelligence models, for example, Ladley and Schenk-Hoppé (2009), in which traders ignore information about the book and remove strategic considerations all together. While these models allow the full market architecture and realistic order submissions, they completely abstract from the central problem we are concerned with here. Other models such as Chiarella, He and Pellizzari (2012) and Chiarella, Iori and Perello (2009) use exogenously specified rules for determining the choice between market and limit order submission and the appropriate price and quantity. These decisions are dependent on

the traders demand and the best quotes in the market. They are restricted, however, by the pre-specified functional structure—there is no guarantee (or claim) that these strategies are optimal in this setting. A third avenue of research of particular interest is in the papers of Goettler et al. (2005) and Goettler, Parlour and Rajan (2009), which use the numerical technique of Pakes and McGuire (2001) to solve an order book market game for a Markov perfect equilibrium in which the trading strategies are optimal. While this may appear to solve the problem these techniques are still numerical demanding. The algorithm attempts to identify the optimal response in all relevant states in the state space. As the size of the order book (the number of orders present) increase, however, this state space grows exponentially. As a result this algorithm is only able to find optimal strategies under a constrained space—either information must be discarded or this algorithm is restricted to books with a relatively small number of ticks and with few orders present.

An important insight to this question is made by Bouchaud, Farmer and Lillo (2009). In this paper, the authors discuss how there may frequently be gaps in the order book—prices at which no orders are present. Even with these *static* gaps the book may be considered to be *dynamically* complete, that is, orders will appear and accumulate as they are needed—they are issued on the fly to provide liquidity. As such, knowledge of many levels of the order book may not be fully revealing of the state of the world if there are traders present within the market that will provide liquidity when it is needed. Information beyond the best quotes may be unreliable. Manahov, Soufian and Hudson (2013) consider a related problem in which traders with different levels of cognitive abilities trade within financial markets. In this case, cognitive ability is reflected by greater capacity for complex strategies and reasoning through larger genetic programs. They find that more intelligent traders enhance price discovery, but damage price stability and liquidity. It is, however, important to emphasise that this study is concerned with the cognitive ability of traders and not the information they have at hand or the size of the strategy space, as we focus on here.

As is the case in many other works, we assume traders are risk neutral profit-maximisers despite the fact that, as pointed out in Parlour and Seppi (2008), agents' decisions should in the end be coherent with their portfolio and consumption choices, which typically display risk-aversion. However, to keep the models numerically manageable, we use reduced-form trading preferences and assume that trading benefits, modelled through private values, proxy for the utility stemming from trading. See the first section of the extensive survey by Parlour and Seppi for a thorough analysis of this issue.

The paper is organised as follows. The next section gives details on the set-up of the market, defines the strategies used by traders and the equilibrium concepts used in this paper. Section 3 describes the two models of optimal trading in a continuous double auction, based on the use of linear and Markov perfect equilibrium strategies. Simulation results are presented in Sect. 4, which reports aggregate data on the order book dynamic equilibria together with an illustration of the optimal strategy used by traders. Some discussion and conclusive remarks end the paper.

2 Set-up

We model a standard order book-based Continuous Double Auction (CDA) where at each time step a single trader enters the market. The trader is randomly allocated a type, buyer or seller with equal probability, and a positive reservation value $v \in V$ or positive cost $c \in C$ for a single unit of the traded asset. We assume that $V = C$ and $|V| = k$, that is, that agents' values and costs belong to the same set of k discrete positive values. Additionally, values and costs are uniformly drawn from V and C and are constant over time: $v_i \in V$ is the i-th buyer's private valuation of the asset, and can be thought as the maximum price he will rationally pay for the asset. Symmetrically, $c_j \in C$ is the j-th seller's private cost for the asset and can be regarded as the minimum price at which he is rationally willing to sell the asset. We will assume, as done frequently in other works, that every agent buys or sell a *single* unit of the asset and, likewise, deal with cancellation in a simplified and standard way: at the end of every time step each order stored in the book is cancelled[1] with (a small) exogenous probability $P_c > 0$ that is independent of time, state of the book and of the specific agent acting in that period.

At any time t the book is a double sequence of outstanding unit orders

$$S_t = \{0 \leq ... \leq b_{3t} \leq b_{2t} \leq b_{1t} < a_{1t} \leq a_{2t} \leq a_{3t} \leq ...\},$$

where $b_{1t}, b_{2t}, ...$ and $a_{1t}, a_{2t}, ...$ are the lists of buy and sell orders in the books. We often omit the time index for simplicity. The highest bid b_1 and lowest ask a_1 are referred as best bid and best ask, respectively. The distance $a_1 - b_1$ is referred to as the spread.

Traders submit a single order when they enter the market. The quantity is fixed at one unit, but the trader must decide the price computed using a function of the state of the book and their valuation: without loss of generality we describe the model for the i-th buyer (the situation for the sellers can be easily recovered, given the symmetry of the environment). The bidding function (or strategy)

$$B_{it} = f_i(a_{1t}, b_{1t}, I_{it})$$

provides the limit price B_{it} (a *bid*, in this case), given the values of the best quotes a_{1t}, b_{1t}. The set I_{it} contains all of the information available to the agent both public and private. This set may include the state of the book and their private valuation/cost. The submission of B_{it} changes the book and results in an immediate trade, a *marketable order*, if the bid is greater than or equal to the best ask, that is, $B_{it} \geq a_{1t}$. In this case, the two agents involved in the transaction get the associated profits: the buyer earns $v_i - a_{1t}$ and the j-th seller, who issued a_{1t} previously, is paid $a_{1t} - c_j$ where c_j is his cost. The book is then updated so that the best ask $a_{1,t+1}$ in the next tick will be given by a_{2t}. If instead $B_{it} < a_{1t}$, the new order is inserted[2] in the book,

[1] We never cancel the order in the time step in which it is submitted.

[2] We always use the standard price-time priority to break ties.

maintaining its ordering, to be possibly used in future trades. Any profit occurring after t is accrued in the same way to the parties with no time-discount. In particular, if $b_{1t} < B_{1t} < a_{1t}$ the order is called *price improving* as it raises the current best bid. Bids for which $B_{1t} \leq b_{1t}$ are less aggressive as the relative limit price is queued after the best bid and, as a consequence, at least one trade is needed before execution is possible.

Notice that in this set-up an immediate transaction may result from many different orders. Indeed, any bid for which $B_{1t} \geq a_{1t}$ produces a transaction and gives the very same profit, regardless of B_{1t}. In other words, there are non-trivial subsets of bidding functions that are formally different and provide different limit prices, but are *profit-equivalent*. This is especially true for strategies that often generate marketable orders, and has implications for the interpretation of the numerical results of the following sections.

An equivalent description holds for the generic j-th seller whose limit ask is given by $C_{jt} = g_j(a_{1t}, b_{1t}, J_{jt})$, where J_{jt} is the information set available (to the seller) at time t. We skip the details for brevity.

Agents are risk neutral and maximise the expected payoff (immediate or delayed), selecting a strategy to issue orders (bids or asks). Once the rules for the auction regarding cancellation and quantities, and the description of the agents are given, different models are obtained specifying the features of the strategies and the information that is processed. The i-th buyer will attempt to solve the problem

$$\max_{f_i \in \mathscr{F}} E[\text{pay}_{it} | O_{-i}, v_i], \tag{1}$$

where pay_{it} is the random profit resulting from bidding what is prescribed by $f_i(a_{1t}, b_{1t}, I_{it})$ at time t, $\mathscr{F}$ is the set of admissible bidding function and O_{-i} denotes the (fixed) strategies used by the other traders. To simplify notation, we omit O_{-i} and v_i whenever this is not harmful. More formally:

$$\text{pay}_{it} = \begin{cases} v_i - a_{1t} & \text{if the order is immediately executed: } B_{it} \geq a_{1t}; \\ v_i - B_{it} & \text{if the order is executed at some time } t' > t : B_{it} < a_{1t}; \\ 0 & \text{if the order is (randomly) cancelled before execution.} \end{cases}$$

The expectation in (1) is taken over all the states of the book that can be faced at t and over all the trajectories of states that can materialise for $t' > t$, starting from the initial condition S_t at time t, under the use of strategies O_{-i}. Unless unrealistically strong assumptions are made, the previous optimisation problem is analytically intractable due to the path-dependency of the book and the intricacies of the auction mechanism. Finding a numerical solution of (1) is still a non-trivial task. This, however, may be tackled in several ways, which will be detailed in what follows.

We will assume, hereafter, that all agents of the same type use the same strategy, and are interested in the equilibrium situation in which no agent has the incentive to change strategy given what other agents do. In detail, we aim at approximately computing a set of bidding (asking) functions

$$O^* = \{f_1^*, ..., f_k^*, g_1^*, ..., g_k^*\}$$

such that for any buyer $i = 1, \ldots, k$, say, we have

$$E[\text{pay}_{it}|f_i^*, O_{-i}^*] \geq E[\text{pay}_{it}|f_i, O_{-i}^*], \forall f_i \in \mathscr{F}, f_i \neq f_i^*, \tag{2}$$

where O_{-i}^* are the strategies optimally played by all the other agents/types. The intuition behind (2) is well known and requires an equilibrium to be a set of policies in which no agent has the incentive to deviate if the other traders stick to their optimal strategy.

3 The Models

In this section, we describe two models of traders' behaviour in a CDA. Several features of the auction (most notably, due to the "double" path dependency, uncertain execution and random cancellation) and the strategic interplay of different types make optimal decisions hard to select or even approximate.

The first model is arguably mimicking a minimal and memory less level of strategic reasoning. Traders submit their orders only based on the best quotes at the time of entering the market. Limit prices are simple weighted averages of a_1 and b_1 (plus a constant). On the top of the best quotes, the information set available to any trader is the empty set. A similar model was used in Pellizzari (2011).

The second model, see Pakes and McGuire (2001), Goettler et al. (2005) and (2009), allows traders to make use of further information—the first l quotes on either side of the book. The expected payoffs of all possible orders in each state of the book are explicitly computed by estimating the execution probability of each order submission (clearly, for marketable orders the execution probability is taken to be 1). As such the profit maximising order may be selected and, effectively, the price setting function may therefore be of arbitrary shape and complexity.

A more detailed description is given in the next subsections.

3.1 Linear Strategies

We assume that the bid/ask to be submitted by traders at time t is given by

$$B_{it} = f_i(a_{1t}, b_{1t}, I_i) = \min(\overline{B}, \alpha_i a_{1t} + \beta_i b_{1t} + \gamma_i), \tag{3}$$

for buyers and

$$A_{jt} = g_j(a_{1t}, b_{1t}, J_j) = \max(\overline{A}, \delta_i a_{1t} + \phi_i b_{1t} + \eta_i), \tag{4}$$

for sellers, where $\alpha_i, \beta_i, \gamma_i, \delta_j, \phi_j, \eta_j$ are real constant to be determined and $I_i = J_j = \emptyset, \forall i, j$. Essentially, all traders compute the limit price to submit by offsetting a linear combination of the best ask and the best bid. Slightly abusing terminology, we refer to these bidding functions as *linear strategies* in the following and notice that f can be thought of as a function of the coefficients α, β, γ as well as a function of the best bid and ask. We enforce a minimal level of rationality and assume that no buyer bids more than some (large) constant price $\overline{B}$ and no seller's ask is satisfied with less than some (small) constant amount $\overline{A}$, but we do not otherwise constraint agents and they are free to pick any linear strategy even though, say, the resulting bid may exceed the private valuation of the asset, and hence, successful execution would cause a net loss. It is also clear from (3, 4) that bids and asks are continuous real values: this is to be contrasted with values and costs that are discrete.

Using the previous linear formulation, we can describe the strategies of all traders as vectors in $\mathbf{R}^3$ so that the bidding function (3) for the i-th type is determined by $\mathbf{x}_i = (\alpha_i, \beta_i, \gamma_i)$. Analogously, the asking function for the j-th seller can be thought of as $\mathbf{y}_j = (\delta_i, \phi_i, \eta_i)$. Given a set of strategies for traders other than the i-th one:

$$O_{-i} = \{\mathbf{x}_1, ..., \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, ..., \mathbf{x}_k, \mathbf{y}_1, ..., \mathbf{y}_k\},$$

he will attempt to maximise the profits solving the problem

$$\max_{\mathbf{x}_i \in \mathbf{R}^3} E[\text{pay}_i | \mathbf{x}_i, O_{-i}].$$

A trading equilibrium is a set of triplets (strategies)

$$O^* = \{\mathbf{x}_i^*, \mathbf{y}_j^*, i, j = 1, ..., k\}$$

such that

$$\mathbf{x}_i^* = \arg\max_{\mathbf{x} \in \mathbf{R}^3} E[\text{pay}_i | \mathbf{x}_i, O_{-i}^*],$$

for all buyers indexed by $i = 1, \ldots, k$ and with an analogous property holding for all sellers, $j = 1, \ldots, k$.

Numerically, the set of equilibrium strategies can be approximated by repeatedly solving the optimisation problem for each type, assuming all the other agents stick to their strategies, and running the algorithm over all types until "convergence is reached". The details of the method are outlined in Pellizzari (2011) and are based on Evolution Strategies. This optimisation technique, thoroughly surveyed in Beyer and Schwefel (2002), evolves the parameters of the population through a number of generations in which the tentative bidding functions are mutated, evaluated, deterministically ranked and discarded based on a fitness measure, before giving birth to the next generation. It is of particular interest here that a meta-parameter related

to the strength of innovation is endogenously evolved together with the unknown parameters and can be used to gauge whether convergence has been successfully reached.

3.2 Markov Perfect Equilibrium Strategies

The second model embodies a different approach in which beliefs of the probabilities of order execution are explicitly calculated. An equilibrium in this framework is a set of probabilities of execution for any limit order in any state of the book. Moreover, we require such an assignment P of probabilities to be consistent, meaning that if agents trade based on the beliefs P, the realised probability of execution is indeed P, so that there is no discrepancy between beliefs and reality.

We assume that the bidding function f_i takes values in V and that the $l \geq 1$ best quotes are known on each side of the market.[3] We refer to l as to the *information level* of the trader, with $l = 1$ being the situation in which no quotes other than the best bid and ask are known. More formally, the i-th buyer's bidding function is

$$f_i : V^{2l} \longrightarrow V, \quad (b_1, ..., b_l, a_1, ..., a_l) \mapsto B_{it},$$

where the bid B_{it} maximises

$$P(B_{it}|S_t)pay_{it},$$

and $P(B_{it}|S_{it})$ is the (perceived) probability that the order will be executed in state S_t either immediately or after some time. In equilibrium, traders decide their bid based on the belief P: $V^{2l+1} \rightarrow [0, 1]$ representing the probability that an order $B_{it} \in V$ issued in state $S_t \in V^{2l}$ at time t will be executed (before exogenous cancellation).

The probabilities are iteratively found as outlined in Pakes and McGuire (2001), aiming at producing $P_n \rightarrow P$ for large n: for any bid $b \in V$ and a state S, at the start of the simulation we set $\forall b, S,\ P_0(b, S) = 1$ and $m_0^{b,S} = 1$.

It is important that the initial probability P is optimistic to facilitate the exploration of the parameter space. The counter m records the number of times a state has been visited—here initialised to 1. The trader who arrives at the market in each period selects the optimal order based on the current estimates of probabilities. Each probability is updated each time step as follows. For a state in which an order executes:

$$P_{t+1}(b, S) = \frac{m_t^{b,S}}{m_t^{b,S}+1} P_t(b, S) + \frac{1}{m_t^{b,S}+1}, \quad m_{t+1}^{b,S} = m_t^{b,S} + 1.$$

[3] We also consider a special case where $l = 0$. In this case prices are selected at random uniformly from the distribution $(0, v_i)$ for buyers and $(c_j, \overline{A})$ for sellers. This constitutes a Zero Intelligence (ZI) strategy as defined by Gode and Sunder (1993).

For a state in which the order is cancelled: $P_{t+1}(b, S) = \frac{m_t^{b,S}}{m_t^{b,S}+1} P_t(b, S)$, $m_{t+1}^{b,S} = m_t^{b,S}$.

For states in which an order is neither cancelled nor executed: $P_{t+1}(b, S) = P_t(b, S)$, $m_{t+1}^{b,S} = m_t^{b,S}$.

A number of algorithmic devices are used to improve speed and avoid premature convergence.[4]

After running the model for T time steps we test for convergence in probabilities. The model is run for a further X time steps during which the updating procedure described above is not applied and probabilities are held constant. Through out this period the number of times orders are submitted in each state and the number of times those orders are executed are both recorded. At the end of the period for any state in which more than 100 orders are submitted the realised probability of order execution is compared with $P(b, S)$, namely the probability of execution estimated by the numerical algorithm. The average mean squared error over all such states is calculated. If this value is less than 0.001 the model is said to be converged, that is, the equilibrium has been identified. If this is not the case the model is run for a further T time steps with probability updating and the model retested. This is repeated until the model is converged. Once this is achieved statistics are collected from the model.

3.3 Further Comments

The two models reviewed in the previous section have some similarities, but are also different in important aspects. Agents in both frameworks share a common set of discrete values/costs and attempt to maximise the gain from trade in a risk-neutral fashion. In the Markov Perfect Equilibrium model, traders must pick a bid/ask among k possible prices (ticks), explicitly computing the expected profit of each option available. The bidding function takes discrete values, but is not restricted in any other way and, in particular, has the potential to reveal that optimal trading may be characterised by some form of nonlinearity.

In contrast, agents using linear strategies can submit orders at any price and this model is not endowed, as was the case for the Markov Perfect Equilibrium market, with a natural tick-size. Hence, in the linear strategy equilibrium, the best quotes can be arbitrarily close at times and this can possibly increase the liquidity and efficiency of the trading process. The strategy of each type of buyer/seller is relatively simple and depends only on three coefficients, whereas a full set of probabilities must be known to take any trading decision in the other model. Importantly, the form of

[4] Every 100,000 time steps we set $m_t^{b,S} = 1, \forall b, S$. Moreover, with probability p_R rather than submitting the utility maximising order a trader instead submits a randomly chosen order in the current configuration. The effect of this is to help prevent local equilibrium. In particular, due to poor early performance, certain actions may no longer be chosen, however, as strategies are refined over time these orders may be once again acceptable. The random selection of these orders allows them to be reintroduced to the strategy.

Table 1 Values and description of the parameters used in the numerical simulations

Variable	Description	Value
V	Buyer valuations	{0.05, 0.10, ..., 0.90, 0.95}
C	Seller valuations	{0.05, 0.10, ..., 0.90, 0.95}
P_c	Probability of cancellation	0.01
$\overline{B}$	Maximum bid	1.0
$\overline{A}$	Minimum ask	0.0
P_R	Probability to issue a random order	0.01
X	Convergence assessment period	1,000,000
T	Optimisation period	1,000,000,000

the bidding functions in the linear strategies market is rather restrictive, and the possibility to devise or approximate any nonlinear trading scheme is ruled out. The following section presents the results of a set of numerical simulations, and discusses the extent to which the differences between the two models have an effect on the book dynamics and traders' actions or profits.

In both models traders are risk neutral. If the traders were risk averse they would trade to minimise the risk of non-execution by placing fewer limit orders and more market orders. This may not necessarily result in a wider spread as, being risk averse, traders would place their orders less far back in the book. Hence, while the proportion of equilibrium orders may be different, the effects of information levels demonstrated in this paper are not likely to change.

4 Results

This section compares the book dynamics prevailing in equilibrium in the two strategic models. For comparison, we also provide results obtained in a market populated by non-strategic Zero Intelligence (ZI) traders.

The simulations are based on the parameter values listed in Table 1. Numerical results for the linear strategy model are based on 20 independent simulations and averages or other statistics are computed using an ensemble of 106,400 states of the book.[5] For the Markov perfect equilibrium model results are calculated over 20 repetitions for each information level and are averaged over 1,000,000 states of the order book.

Table 2 shows the average state of the book under different models: together with ZI traders ($l = 0$), we have considered three different information levels l and the use of linear strategies.

[5] States are obtained from 20 independent simulations of 7 days of trading. We approximate a continuous flow of traders using a large population of 760 agents, 380 buyers and 380 sellers: hence, statistics are based on $106,400 = 20 \times 7 \times 760$ states.

Table 2 Summary statistics of the book under the four different information levels

Model	ZI	$l = 1$	$l = 2$	$l = 3$	Linear
Best bid	0.424	0.466	0.465	0.464	0.472
Best ask	0.576	0.534	0.534	0.536	0.524
Spread	0.152	0.068	0.069	0.071	0.051
Quantity at best bid	1.76	2.39	2.40	2.32	–
Quantity at best ask	1.76	2.42	2.44	2.40	–

The market populated by ZI traders is substantially different from any strategic market, with much wider best quotes on average and an inflated spread. Clearly, the lack of strategic considerations in this case results in too many orders being randomly placed behind the best quotes and with a low probability of ending in a trade. Conversely, any market populated by strategic traders shows a much narrower spread, close to the gap between adjacent traders' values or costs. There is virtually no difference for different levels l of information and little practical discrepancy between the set of the Markov equilibria and the linear strategies equilibrium. The average equilibrium spread using linear strategies 0.051 compared to about 0.070 for the other models (regardless of l), but it must be noticed that in the latter cases the spread cannot be less than 0.05, as offers on opposite sides are discrete and cannot overlap.[6] As such the presence of a minimum price increment (tick) in the Markov model has only a small effect on the equilibrium market behaviour.[7]

To understand why the information level has little effect on behaviour it is beneficial to consider the problem faced by traders. In the model, in equilibrium, the traders' estimates of the probabilities of orders executing are always correct. For a given state X in information level l this probability is the average, weighted by frequency of occurrence, of all states that in information level $l + 1$ would map to state X. For instance, consider the state X for $l = 1$ of $\{B_1 = 0.4, A_1 = 0.6\}$ (i.e. the best bid is 0.4 and the best ask 0.6). There are a large number of states in $l = 2$ which map to this, including $\{B_1 = 0.4, A_1 = 0.6, B_2 = 0.3, A_2 = 0.7\}$, $\{B_1 = 0.4, A_1 = 0.6, B_2 = 0.3, A_2 = 0.8\}$, $\{B_1 = 0.4, A_1 = 0.6, B_2 = 0.3, A_2 = 0.9\}$ etc. All of these states in $l = 2$ would be represented by X in $l = 1$. The greater number of states allows traders to specify their strategy more finely, but they do not measure the probability of execution over the set any more accurately. As such, there may be some states where traders are more aggressive at $l = 2$ than they would be in X at $l = 1$ and, similarly, some where they are less aggressive. The chosen action

[6] The quantities at the best quotes for the linear model are not given as with continuous pricing there is never more tha one order at this price.

[7] The effect of the width of the price grid—the number of ticks in the market—was also considered. Doubling the number of ticks in the price grid led to an increase in the spread of 50 % while the quantities at the best quotes were found to be 50 % greater under the smaller set of prices. Importantly, however, a larger price grid was found to have no effect on the behaviour of the model across information levels, that is, for all information levels the spread and quantities available were the same.

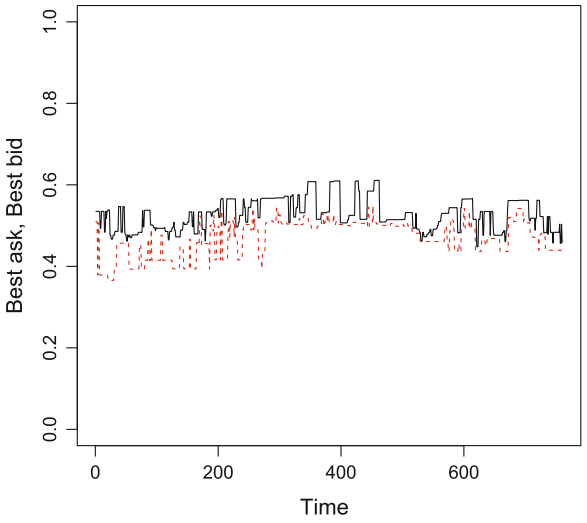

Fig. 1 Example of time evolution of best bid and ask in equilibrium using linear strategies. Best bid is given as a dashed line while the best ask is the *solid line*. Each time step corresponds to a single trader entering the market

at level $l = 1$ may therefore be viewed as the payoff maximising action averaged over all possible states at $l = 2$. This explains why the information level has little influence on the aggregate behaviour being actions averaged across all states.

A snapshot of the best quotes realised with linear strategies is depicted in Fig. 1. The graph shows that there is considerable variability in the trading session as well as frequent periods in which the spread falls to minute levels (periods when the two lines nearly intersect). This demonstrates why the average spread in the presence of linear strategies is smaller than in the Markov perfect equilibria.

Table 3 shows the distribution of spreads for all the markets. Again the statistics for the four markets with strategic traders are very similar. In all cases over half of the time the best bid and ask are within one tick of the equilibrium price. In 90 % or more of the cases the spread is within two ticks and in nearly all cases the spread is within three ticks. In contrast, the ZI market shows much more variability in the spread. In only 14 % of observations is the spread within one tick of the mid price indicating that the market is much more volatile and less efficient. This indicates that for markets populated by strategic traders the price is relatively stable and, importantly, there are only a small number of market situations, which traders are faced with. As such the degree of strategic sophistication traders' require may be low.

Table 4 shows the relative shares of the type of orders submitted in different markets. Again, the ZI results differ markedly from the ones of the strategic models: marketable orders are halved with respect to the other markets, few orders are aggressively improving the extant quotes and, as a consequence, most of orders are placed behind the best quotes. These results broadly match those highlighted by Ladley and Schenk-Hoppé (2009) who found that the ZI model produced too few

Table 3 Distribution of ranges of bid and ask spreads for traders using ZI, Markov ($l = 1, 2, 3$) and linear strategic

		0.45−0.55	0.40−0.60	0.35−0.65
	0.45−0.55	0.14	0.28	0.39
ZI	0.40−0.60	0.28	0.49	0.65
	0.35−0.65	0.39	0.65	0.85
	0.45−0.55	0.61	0.79	0.80
$l = 1$	0.40−0.60	0.78	0.97	0.98
	0.35−0.65	0.80	0.98	1.00
	0.45−0.55	0.62	0.79	0.81
$l = 2$	0.40−0.60	0.78	0.95	0.97
	0.35−0.65	0.81	0.97	0.99
	0.45−0.55	0.62	0.78	0.81
$l = 3$	0.40−0.60	0.77	0.94	0.96
	0.35−0.65	0.80	0.96	0.99
	0.45−0.55	0.50	0.69	0.74
Linear	0.40−0.60	0.68	0.89	0.95
	0.35−0.65	0.72	0.94	0.99

Rows correspond to bid price and columns to ask prices

Table 4 Distribution of types of orders under the four different information levels along with number of cancellations and trades

Model	ZI	$l = 1$	$l = 2$	$l = 3$	Linear
Market orders	0.113	0.233	0.233	0.233	0.257
Price improving limit orders	0.073	0.108	0.104	0.109	0.181
Limit orders at best quote	0.045	0.162	0.167	0.161	–
Limit orders behind best quote	0.769	0.497	0.496	0.498	0.563

orders market orders and limit orders at the best quotes and too many behind the best quotes relative to empirical data. In reality, as well as in this model, strategic behaviour leads to fewer limit orders being wasted—being placed behind the best quotes with little chance of execution. Sophisticated traders choose not to submit these orders and submit price improving orders instead.

The market with linear strategies is slightly more efficient than the Markov markets, as seen in the fractions of market(able) orders, 25.7 %, as compared to 23.3 %. This implies that the traded volume is almost 5 % bigger in the market with linear strategies than in the Markov ones due to the smaller spread available in the first market. As before, orders at the best quote are meaningless in the linear model. We therefore, provide in the table only the share, 56.3 %, of non-improving orders for the model with linear strategies. Despite some differences, all the strategic markets are rather similar as shown by a more accurate comparison, say, between the linear model and the one in which $l = 3$. The share 16.1 % of "at the best quote" orders for the Markov model can be split in equal parts and tallied in the "improving" and "behind the quotes" orders, respectively, assuming that with equal probability an

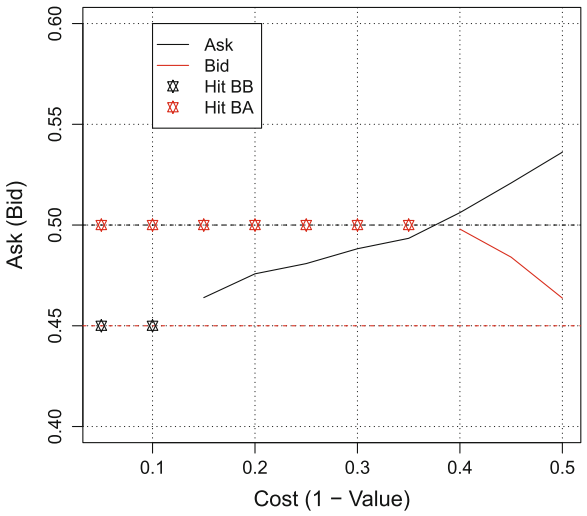

Fig. 2 Trading behaviour of intra-marginal buyers and sellers facing best quotes 0.50 and 0.45. *Black* (*red*) stars denote market orders submitted by sellers (*buyers*) and *black* (*red*) *solid lines* show the median ask (*bid*) when limit orders are posted. The horizontal axis shows the costs for sellers and 1 less the values for buyers

order at the best quote falls in either of the neighbouring category. In such a way the fifth column of Table 4 would have 18.9 % of improving and 57.8 % of "behind the best quote" orders, which should be compared to 18.1 and 56.3 % of the sixth column, relative to the linear strategy equilibrium.

It is of interest to also look at the behaviour of the traders in equilibrium, particularly when they use linear strategies that are relatively simple. Recall that the models contemplate heterogeneous agents with different values and costs: while some may be strongly intra-marginal, feeling an intense pressure to finalise a trade to get profits, others—the extra-marginal ones—will basically have no chance to trade in equilibrium, being outstanding quotes at levels that do not make possible execution at a profit compared to reservation values. Moreover, as hinted in Sect. 2, even though different strategies are evolved in distinct simulations, they are however almost perfectly profit-equivalent.

A way to represent what agents do is to show what they bid/ask facing some frequently visited states of the book. We take the two symmetrical configurations in which the best quotes are 0.50, 0.45 and 0.55, 0.50, respectively. Figures 2 and 3 depict the median of the limit orders posted by intra-marginal sellers and buyers across all the simulations. When the best quotes are 0.50 and 0.45 (dashed in Fig. 2), there is fierce competition among sellers who pushed the ask downwards to get closer to the outstanding bid. On the one hand, the strongest sellers, with costs equal to 0.05 or 0.10, issue marketable orders hitting the best bid and cashing 0.45 for one unit of the asset (see the black stars in the picture): they get less than the equilibrium price, but trade is immediate and large profits are secured anyway. On the other hand, sellers

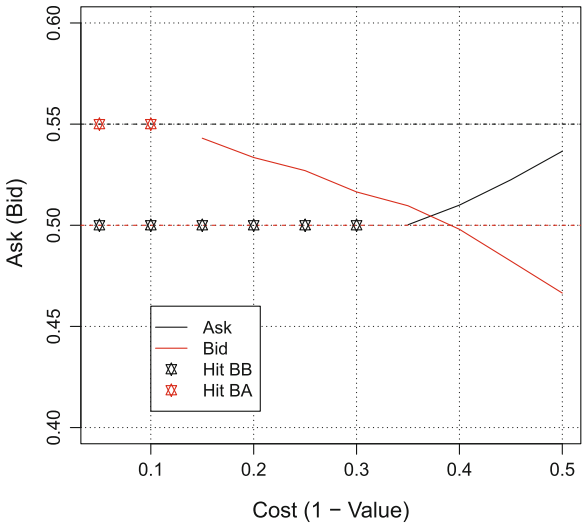

Fig. 3 Trading behaviour of intra-marginal buyers and sellers facing best quotes 0.55 and 0.40. *Black* (*red*) stars denote market orders submitted by sellers (*buyers*) and *black* (*red*) *solid lines* show the median ask (*bid*) when limit orders are posted. The horizontal axis shows the costs for sellers and 1 less the values for buyers

whose cost exceeds 0.10 prefer to post limit orders that are not immediately executed, see the black solid line: in particular, we observe that the median order is improving when $c = 0.15, ..., 0.35$ and behind the best quote when $c = 0.40, 0.45, 0.50$.

Buyers in Fig. 2 find an attractive (low) ask and the ones whose value is larger or equal to 0.65 content themselves with a marketable order, see the red stars representing bids hitting the quote 0.50 and notice that the horizontal axis shows $1 - v$ for buyers. Agents with values $v = 0.60, 0.55, 0.50$ prefer to improve the outstanding best bid in order to gain priority, see the red solid line.

Figure 3 almost perfectly matches Fig. 2, after swapping the roles of buyers and sellers. Even when the depicted behaviour is distinct, this results in minute differences in profits and even more so if one takes into account that the figures represent *median* behaviours. Take, for instance, the seller whose cost is 0.35 in Fig. 3: he will decrease the ask to 0.502, virtually zeroing the spread and securing for himself an expected profit that is very similar to the one immediately cashed by the symmetric buyer whose value is 0.65 in Fig. 2.

Overall, the pictures represents a rather sensible and, *ex post*, intuitive behaviour on the part of traders: strongly intra-marginal agents typically trade immediately using marketable orders, either because there is fierce competition on their side or because the quote on the opposite side is (already) captivating. The weakly intra-marginal traders improve the best quote to gain priority or patiently queue their orders in the hope that future, less unbalanced, states of the book will make their offers competitive.

5 Conclusion

In this paper, we have used two models of order book markets to investigate the importance of information and strategic sophistication. The results provide insights into the effect and importance of information on optimal trading in order book-based continuous double auctions. The statistical measures of market and trader behaviour differed little across levels of information. These statistics, however, were very different from those obtained under the zero-intelligence model where lack of knowledge and the resulting random behaviour results in sub-optimal trading. We may therefore conclude that the crucial piece of information for traders in constructing their optimal strategy is knowledge of the best quotes. Further knowledge about the book conveys no value in this context: intuitively, this may be related to the dynamic nature of the book, where orders are likely to be added close to the best quotes as and if they are needed. As Bouchaud et al. (2009) argue the book may be dynamically complete.

Key to the effect above is the finding that in equilibrium only a relatively small number of order book states occur as shown by the large percentage of observation in which the spread occupies a relatively narrow band around the equilibrium price. As such the possible situations that traders must develop optimal responses for are small in number. Traders strategies may therefore be relatively simple and easily learnt. Moreover, the similarity between the optimal Markov strategies and the linear approximation indicates that optimal trading may be approximated by a simple functional form further easing the cognitive burden placed on traders.

The work presented in this paper could be extended to consider more complex market settings. In this paper, we have considered a relatively simple market—a fixed equilibrium price, unit quantities and exogenous cancellation. All three of these aspects could be made more sophisticated. A moving equilibrium price would exacerbate the risk for limit order submitters—increasing the chance of non-execution or picking off if the price moved away from or towards the order. Non-unit orders could increase the impact of a trader on the book as they would potentially be able to remove liquidity at multiple price ticks. Endogenous cancellation and resubmission of orders would allow traders to adapt their order placement to the changing state of the market. All three of these changes would possibly increase the value of information beyond the first tick. It was surprising in the current setting that only the first level of information was valuable. Identifying the requirements for this to be the case more generally, however, would be a potentially valuable advance.

References

Beyer, H.-G., & Schwefel, H.-P. (2002). Evolution strategies: a comprehensive introduction. *Natural Computing*, *1*, 3–52.

Bouchaud, J.-P., Farmer, J. D., & Lillo, F. (2009). How markets slowly digest changes in supply and demand. In T. Hens & K. R. Schenk-Hoppé (Eds.), *Handbook of financial markets: Dynamics and evolution* (pp. 57–160). North-Holland, San Diego: Handbooks in Finance.

Chiarella, C., He, X.-Z., & Pellizzari, P. (2012). A dynamic analysis of the microstructure of the moving average rules in a double auction market. *Macroeconomic Dynamics*, *16*, 556–575.

Chiarella, C., Iori, G., & Perello, J. (2009). The impact of heterogeneous trading rules on the limit order book and order flows. *Journal of Economic Dynamics and Control*, *33*(3), 525–537.

Gode, D. K., & Sunder, S. (1993). Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *Journal of Political Economy*, *101*(1), 119–37.

Goettler, R. L., Parlour, C. A., & Rajan, U. (2005). Equilibrium in a dynamic limit order market. *Journal of Finance*, *60*(5), 2149–2192.

Goettler, R. L., Parlour, C. A., & Rajan, U. (2009). Informed traders and limit order markets. *Journal of Financial Economics*, *93*(1), 67–87.

Ladley, D., & Schenk-Hoppé, K. R. (2009). Do stylised facts of order book markets need strategic behaviour? *Journal of Economic Dynamics and Control*, *33*(4), 817–831.

Manahov, V., Soufian, M., & Hudson, R. (2013). The implications of trader cognitive abilities on stock market properties. Intelligent Systems in Accounting, Finance and Management Forthcoming.

Pakes, A., & McGuire, P. (2001). Stochastic algorithms, symmetric markov perfect equilibrium, and the 'curse' of dimensionality. *Econometrica*, *69*(5), 1261–1281.

Parlour, C. A. (1998). Price dynamics in limit order markets. *Review of Financial Studies*, *11*(4), 789–816.

Parlour, C. A., & Seppi, D. J. (2008). Chapter 3—limit order markets: A survey. In A. V. Thakor & A. W. Boot (Eds.), *Handbook of financial intermediation and banking* (pp. 63–96). Elsevier, San Diego: Handbooks in Finance.

Pellizzari, P. (2011). Optimal trading in a limit order book using linear strategies, Working Papers 16, Department of Economics, University of Venice "Ca' Foscari".

Rosu, I. (2009). A dynamic model of the limit order book. *Review of Financial Studies*, *22*(11), 4601–4641.

Regime Switching Models in the Foreign Exchange Market

Wai-Mun Chia, Mengling Li and Huanhuan Zheng

1 Introduction

A large number of studies have confirmed the application of heterogeneous bounded-rational strategies based on technical and fundamental analysis in the trading of financial assets by investment professionals. For instance, Boswijk et al. (2007), Chiarella et al. (2012) and Yamamoto and Hirata (2013) have documented the evidence of behavioural heterogeneity and the simultaneous presence of both fundamentalists and chartists in the stock markets. Similar observation is also found in the options market (Frijns et al. 2010) and the foreign exchange markets (Gilli and Winker 2003; Westerhoff and Reitz 2005; Menkhoff et al. 2009; de Jong et al. 2010). Such heterogeneity in investment behaviour is also supported by the experimental evidence of Hommes (2011) and the questionnaire survey of Allen and Mark (1990), Cheung and Chinn (2001) and Gehrig and Menkhoff (2004), among others. A more comprehensive summary of these surveys can be found in the study of Menkhoff and Taylor (2007).

W.-M. Chia
Division of Economics, Nanyang Technological University, 14 Nanyang Drive, Singapore 637332, Singapore
e-mail: aswmchia@ntu.edu.sg

M. Li
Division of Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, Singapore 637371, Singapore
e-mail: mli6@e.ntu.edu.sg

H. Zheng (✉)
Institute of Global Economics and Finance, The Chinese University of Hong Kong, 12 Chak Cheung street, Shatin N. T, Hong Kong
e-mail: arwenzh@gmail.com

H. Zheng
Department of Economics and Related Studies, The University of York, Heslington, York YO10 5DD, UK

R. Dieci et al. (eds.), *Nonlinear Economic Dynamics and Financial Modelling*,
DOI: 10.1007/978-3-319-07470-2_12,

Heterogeneity can either be cross-sectional where different types of investors adopt different strategies or time-variant where the same investor can switch among heterogeneous strategies over time (Boswijk et al. 2007; Chiarella et al. 2012; Yamamoto and Hirata 2013). Accounting for such a switch in trading behaviour is important in both theoretical and empirical work. Theoretically, Huang and Zheng (2012) and Huang et al. (2010, 2012) show that, besides the cross-sectional heterogeneity, accounting for the regime-switching beliefs itself can improve the models capability to generate financial market stylized facts that match well with actual financial market data. Empirically, Manzan and Westerhoff (2005), De Grauwe and Grimaldi (2006), de Jong et al. (2010) and Chiarella et al. (2012) show that models with the regime-switching behaviour tend to outperform those without in terms of better in-sample explanatory power and out-of-sample forecasting ability. Besides, these models also demonstrate better predictive power than a simple foreign exchange rate random walk model that outperforms many structural exchange rate models (Meese and Rogoff 1983).

Existing regime-switching models that are used to estimate behavioural heterogeneity can be broadly classified into three categories according to their switching mechanisms. The first category is pioneered by Boswijk et al. (2007, BHM hereafter) where investors switch their strategies based on past performance. In the original work of Boswijk et al. (2007), investors cluster evolutionarily to the strategy that generate higher past realised profits according to some discrete choice probability, with root tracing back to Brock and Hommes (1998). Many empirical studies later follow similar switching rule, but refer to various switching criteria based on backward looking indicators. For example, in Manzan and Westerhoff (2007), investment decision is updated according to the deviation between actual and fundamental prices, whereas studies by de Jong et al. (2010), Jongen et al. (2012) and ter Ellen et al. (2013) assume that investors switch to strategies with relatively accurate past forecasting. This strand of regime-switching models is by far the most commonly applied type of models in the current literature. The second strand of empirical models follows closely the work of Lof (2012) where investors update their trading strategies according to business cycles. The switching rule is governed by a smooth-transition function of some macroeconomic variables such as GDP growth rate, industrial production and consumer price index. In this class of models, investors follow a chartist strategy during economic expansion and switch to a fundamentalist strategy during economic contraction. The third strand of models follows the work of Chiarella et al. (2012) where the switching originates from the structural change in expectations based on some unobserved conditions in the financial markets that are governed by a Markov process. In this set up, investors initially play the same trading strategy, but subsequently deviate from their original forecasting formula and switch completely to a different strategy depending on the market conditions. Putting these together, it is noted that behavioural heterogeneity can be cross-sectional and time-variant. It can also arise within the same trading strategy. In the first and second types of models, investors switch evolutionarily between trading strategies that are government by some rule. In the third type of models, even though investors do not switch

between strategies, strategy may switch according to different market conditions that are governed by a Markov switching process.

Although there is clear evidence that the regime-switching models provide a good empirical specification to estimate behavioural heterogeneity, it remains unclear as in which type of switching mechanism performs better. In view of the lack of performance comparison, in this chapter, we aim to compare different switching mechanisms and evaluate their performance in estimating the foreign exchange market.[1] Specifically, we identify the comparative advantage of each switching mechanism in terms of its goodness-of-fit, estimation efficiency and predictive power, using a framework that is as simple as possible. In doing so, we first develop a benchmark model that is sufficiently general to incorporate all the three types of switching mechanisms while keeping other factors unchanged. We then estimate the three models separately.

We choose to compare the three switching mechanisms in the foreign exchange market with the context of AUD/USD for several reasons. First, AUD/USD is the fourth-most-traded currency pair, which accounts for 7 % of the global foreign-exchange market (Economist 2013). The common practices of various types of traders in such a market is relatively general and representative. Second, the Federal Reserve of United States has kept the interest rate close to zero for more than five years, while the interest rate in Australia has been high (its current interest rate is still 2.75 % after several round of rate cuts), which makes this currency pair one of the most important vehicle for carry trade. Third, to the best of our knowledge, no empirical literature on behavioural heterogeneity has studied AUD/USD despite its importance in the foreign exchange market. Our studies complement existing literature with the trading behaviour on AUD/USD.

Using monthly data from January 2000 to June 2013, our empirical results suggest that (1) behavioural heterogeneity can arise not only due to cross-sectional heterogeneity with investors taking different strategies or time-variant heterogeneity with investors switching their strategies over time, but also within-group heterogeneity when investors who stick to the same strategy shift their expectations over time; (2) while strategy switching based on a smooth-transition function of various macroeconomic variables proposed by Lof (2012) provides the best in-sample explanatory power and strategy switching based on past performance as in Boswijk et al. (2007) exhibits better out-of-sample forecasting power, there is no significant evidence to confirm any of these two switching mechanisms outperforms that of Chiarella et al. (2012) in terms of forecasting accuracy.

The remaining of the chapter is organized as follows. Section 2 describes the model. Section 3 presents the data and methodology. Section 4 discusses the estimation results of the three models. Sections 5 and 6 compare the three models in terms of their estimation efficiency and out-of-sample forecasting power, respectively. Section 7 concludes.

[1] Existing papers typically apply BHM to estimate HAMs, using various foreign exchange data. ter Ellen et al. (2013) apply weekly survey data to estimate HAM with the BHM method. They find existence of value traders and momentum traders as well as the switching between them. Perhaps due to the monthly data that we have applied, we find no evidence of switching behaviour in the AUD/USD exchange market with the BHM method.

2 Model Specifications

Like many other asset markets, the foreign exchange market is also believed to consist of investment professionals who use both fundamental and technical analysis in their trading activities (Menkhoff and Taylor 2007). Many existing studies have indeed confirmed the simultaneous presence of fundamentalists and chartists in the foreign exchange market. These include the work of Cheung and Chinn (2011) and Gehrig and Menkhoff (2004). Additionally, many researchers who incorporate fundamentalist and chartist strategies in their asset pricing models show that simulated data in these models match with almost all stylised facts (De Grauwe and Grimaldi 2006; Huang et al. 2012). These models also provide empirical specifications that outperform the commonly used random walk model (Chiarella et al. 2012; de Jong et al. 2010). Motivated by these findings, we consider a foreign exchange market that consists of both fundamentalists and chartists who submit their orders to a market maker. The market maker then adjusts the exchange rate up or down according to the size of aggregate order.

2.1 Benchmark Model

2.1.1 Fundamentalists

In the benchmark model, the fundamentalists perceive the deviation between spot exchange rate S_t and fundamental exchange rate u_t as a trading opportunity. According to them, expected price movement is caused by the mispricing of either an undervaluation or an overvaluation of a currency. While some of them believe in the persistence of the mispricing in the short term, others may expect a reversion back to its fundamental value. Together, their aggregate demand function is given by:

$$D_{f,t} = \alpha_{ft} \left(u_t - S_{t-1} \right), \tag{1}$$

where α_{ft} measures the extent to which the fundamentalists act on their belief. With $\alpha_{ft} > 0$, the majority of the fundamentalists believe in the mean-reverting of price, that is, they expect the currency to appreciate (depreciate) in the future if S_{t-1} is below (above) u_t. In this case, the aggregate actions of the fundamentalists stabilise the currency. With $\alpha_{ft} < 0$, the majority of the fundamentalists believe that the spot rate will continue to deviate from its fundamental, at least for a while. In this case, the aggregate actions of the fundamentalists destabilise the currency and worsen the deviation. Our specification of the fundamentalists is slightly different from the traditional definition, where they always trade to drive prices towards its fundamentals. The traditional definition is a special case in our set-up.

There is a lack of consensus in estimating the fundamental value u_t in the foreign exchange market (Taylor 1995). While many theoretical works propose to estimate

the fundamental exchange rate using monetary models based on money stock and real income (Mark 1995), recent works suggest an extended framework that includes money supply, output, interest rate, expected inflation and trade balance (Neely and Sarno 2002).[2] There is, however, no convention on estimating the corresponding coefficients, such as the money-demand income elasticity. Therefore, these methods that work well in the existing theoretical literature are not sufficient for our purposes. Most of the foreign exchange rate determination models draw upon purchasing power parity (PPP) and uncovered interest rate parity (UIP). However, the calculation of true PPP value of an exchange rate is ambiguous and not straightforward (Ellen et al. 2013). Therefore, in our studies, we choose to proxy the fundamental exchange rate based on UIP, which requires only directly observable data and is relatively easy to estimate. The application of UIP is also justified by the observation that foreign exchange traders watch closely the interest rates movement underlying the currency pairs (in our case, the federal funds rate in the United States and the cash rate in Australia), the two most important indicators underlying UIP. Specifically, we measure the fundamental foreign exchange rate for AUD/USD based on the following equation:

$$u_{t+1} = S_t(1 + r_t^{AU}/12)/(1 + r_t^{US}/12), \tag{2}$$

where r_t^{US} is the annualised US money market rate and r_t^{AU} is the annualised Australia money market rate.

Fundamentalists who expect mean reversion are those who believe in UIP. They will buy the currency of low-interest country and sell the currency of high-interest country and anticipate the exchange rate of low-interest country to appreciate. These are known as the UIP traders. On the other hand, for fundamentalists who expect the price deviation to persist will sell the currency of low-interest country and buy the currency of high-interest country. They will profit from the interest rate differential as the exchange rate remains constant or depreciates. They are known as the carry traders. Therefore, when the majority of the fundamentalists are UIP traders, $\alpha_{ft} > 0$. When carry traders form the majority, $\alpha_{ft} < 0$. We distinguish the fundamentalists as carry traders and UIP traders, beyond the archetypical specifications of fundamentalists in HAMs. This set up is in accordance with the importance of interest differential as a determinant of the foreign exchange expectations, in addition to the conventional fundamental and momentum considerations in other financial markets (Jongen et al. 2012).

2.1.2 Chartists

The chartists in the model are those who conduct technical analysis to form their expectation of future exchange rate. While there are many technical rules, the most commonly applied rule is the momentum rule. Empirical studies have indeed con-

[2] See Taylor (1995) for a survey on various models of foreign exchange rate determination.

firmed the presence of such a trading rule. Besides, the momentum rule is found to be relatively more profitable as compared to the other technical rules, such as moving average (Jongen et al. 2012). Motivated by these findings, we assume that the chartists form their expectation based on the most basic form of momentum rule AR(1):

$$E_{c,t}(S_t) = S_{t-1} + \beta_t \left(S_{t-1} - S_{t-2}\right), \tag{3}$$

where β_t is the extrapolation rate of the chartists and it measures the degree of expected autocorrelation. When $\beta_t > 0$, the chartists expect the price trend to persist (bandwagon expectation). On the other hand, when $\beta_t < 0$, the chartists expect the past price trend to reverse (contrarian expectation). The aggregate demand of the chartists is given by:

$$D_{c,t} = \eta[E_{c,t-1}(S_t) - S_{t-1}] = \lambda_t \left(S_{t-1} - S_{t-2}\right), \tag{4}$$

where $\eta > 0$ is a parameter that measures the extent to which the chartists act on their beliefs and the second line is obtained by letting $\lambda_t = \eta\beta_t$. Note that $\lambda_t > 0$ ($\lambda_t < 0$) if and only if $\beta_t > 0$ ($\beta_t < 0$).

2.1.3 The Market Maker

Within a market maker framework, the market maker collects orders from all traders and subsequently quotes the spot exchange rate according to the aggregate demand with a speed of $\gamma > 0$. We use $\omega_{f,t}$ and $\omega_{c,t}$ to denote the market weights (or fractions) of fundamentalists and chartists, respectively. The exchange rate is updated according to the aggregate demand and a noise term ε_t, which can be written as:

$$\begin{aligned}\Delta S_t &= S_t - S_{t-1} \\ &= \gamma\left(\omega_{f,t} D_{f,t} + \omega_{c,t} D_{c,t}\right) + \varepsilon_t \\ &= \gamma\omega_{f,t}\alpha_{ft}\left(u_t - S_{t-1}\right) + \gamma\omega_{c,t}\lambda_t\left(S_{t-1} - S_{t-2}\right) + \varepsilon_t .\end{aligned} \tag{5}$$

where γ is the speed of exchange rate adjustment and the third line is obtained by substituting $D_{f,t}$ and $D_{c,t}$ from Eqs.(1) and (4).

In the following, we specify the three types of switching mechanisms separately. In particular, the proportion of fundamentalists, $\omega_{f,t}$, and chartists, $\omega_{c,t}$, are modelled to evolve according to different rules—switching based on past performance (Boswijk et al. 2007) and switching based on macroeconomic fundamentals (Lof 2012), while the intensity of actions of fundamentalists and chartists, α_t and λ_t, are modelled to be contingent on the foreign exchange market state for the Markov-switching beliefs (Chiarella et al. 2012).

2.2 Strategy Switching Based on Past Performance

Following BHM, we assume that traders update their market weight every period according to the most recent realised profits. The proportion of fundamentalists, $\omega_{f,t}$, and chartists, $\omega_{c,t}$, evolve according to a discrete choice model with multinomial logit probabilities:

$$\begin{aligned}\omega_{f,t} &= \frac{\exp\left(\rho\pi_{f,t-1}\right)}{\exp\left(\rho\pi_{f,t-1}\right)+\exp\left(\rho\pi_{c,t-1}\right)} \\ \omega_{c,t} &= \frac{\exp\left(\rho\pi_{c,t-1}\right)}{\exp\left(\rho\pi_{f,t-1}\right)+\exp\left(\rho\pi_{c,t-1}\right)} = 1-\omega_{f,t},\end{aligned} \tag{6}$$

where ρ is a scaled intensity of choice that measures the sensitivity to the relative profitability of the trading rules, and $\pi_{f,t-1}$ and $\pi_{c,t-1}$ represent the most recent realised return of fundamentalists and chartists, respectively. Specifically, the profits of each fundamentalist and chartist can be described by:

$$\begin{aligned}\pi_{f,t} &= D_{f,t}\left(S_t - S_{t-1}\right) \\ \pi_{c,t} &= D_{c,t}\left(S_t - S_{t-1}\right),\end{aligned} \tag{7}$$

To comply with BHM, we further assume that $\alpha_{ft} \equiv \alpha_f$ and $\lambda_t \equiv \lambda$. This implies that both fundamentalists and chartists maintain the same degree of acting on their beliefs. In the BHM style of switching, Eq. (5) can be written as:

$$\begin{aligned}\Delta S_t &= S_t - S_{t-1} \\ &= \frac{\gamma\alpha_f \exp\left(\rho\pi_{f,t-1}\right)\left(u_t - S_{t-1}\right)}{\exp\left(\rho\pi_{f,t-1}\right)+\exp\left(\rho\pi_{c,t-1}\right)} + \frac{\gamma\lambda \exp\left(\rho\pi_{c,t-1}\right)\left(S_{t-1} - S_{t-2}\right)}{\exp\left(\rho\pi_{f,t-1}\right)+\exp\left(\rho\pi_{c,t-1}\right)} + \varepsilon_t \\ &= \frac{\gamma\alpha_f\left(u_t - S_{t-1}\right)}{1+\exp\left[-\frac{\rho}{\gamma}\Delta S_{t-1}\left(\gamma\alpha_f\left(u_{t-1} - S_{t-2}\right) - \gamma\lambda\left(S_{t-2} - S_{t-3}\right)\right)\right]} \\ &\quad + \frac{\gamma\lambda\left(S_{t-1} - S_{t-2}\right)}{1+\exp\left[\frac{\rho}{\gamma}\Delta S_{t-1}\left(\gamma\alpha_f\left(u_{t-1} - S_{t-2}\right) - \gamma\lambda\left(S_{t-2} - S_{t-3}\right)\right)\right]} + \varepsilon_t\end{aligned} \tag{8}$$

While the time series of spot exchange rate S_t is directly observable, the fundamental exchange rate u_t is calculated based on Eq. (2). Other parameters to be estimated are γ, α_f, λ and ρ. To prevent under-specification, we estimate $\gamma\alpha_f$, $\gamma\lambda$ and ρ/γ instead. As $\gamma > 0$ is only a scaling factor, it will not affect the statistical significance of the estimation coefficients.

2.3 Strategy Switching Based on Macro Fundamentals

Following Lof (2012), traders are assumed to update their strategy according to macroeconomic fundamentals. The weight of fundamentalists, $\omega_{f,t}$, follows a logistic smooth-transition regressive function (LSTR) of a lagged matrix of macroeconomic fundamental variables X_{t-1}:

$$\omega_{f,t} = \frac{1}{1 + \exp\left[\tau\left(X_{t-1}A - c\right)\right]}, \tag{9}$$

where A is a column vector denoting the weight of the fundamental variables, the parameter $\tau > 0$ measures the sensitivity to the fundamental matrix, and c is the threshold exceeding which the foreign exchange market will be dominated by chartists. The model implies transitions between two regimes: the regime dominated completely by fundamentalists, $\omega_{f,t} = 1$, which tends to occur when $X_{t-1}A < c$ and the regime dominated completely by chartists, which tends to occur when $X_{t-1}A > c$. When $X_{t-1}A = c$, fundamentalists and chartists share the market equally. The selection of macroeconomic fundamental variables X_{t-1} will be based on a series of linearity tests. More details are discussed in the methodology section.

Following Lof (2012), we let $\alpha_{ft} \equiv \alpha_f$ and $\lambda_t \equiv \lambda$. In Lof's LSTR switching model, Eq. (5) can be written as:

$$\Delta S_t = \frac{\gamma\alpha_f\left(u_t - S_{t-1}\right)}{1 + \exp\left[\tau\left(X_{t-1}A - c\right)\right]} + \frac{\gamma\lambda\exp\left[\tau\left(X_{t-1}A - c\right)\right]\left(S_{t-1} - S_{t-2}\right)}{1 + \exp\left[\tau\left(X_{t-1}A - c\right)\right]} + \varepsilon_t. \tag{10}$$

As the time series of u_t, S_t and X_t are directly observable, we can estimate the LSTR for $\gamma\alpha_f$, $\gamma\lambda$, A, c and τ.

2.4 Markov-Switching Beliefs

As in the financial market, the foreign exchange market is characterised by currency appreciation with low volatility in a boom state, and currency depreciation with high volatility in a bust state. Investors' trading behaviour is likely to vary in different states. Motivated by these observations, the beliefs of fundamentalists and chartists as well as the noise term are modeled to be state-dependent.

Following Chiarella et al. (2012), λ_t (to be exact β_t) is assumed to be contingent on the foreign exchange market state m_t, which takes a discrete value of 0 or 1 so that $m_t \in M = \{0, 1\}$. The dynamics of the state aim at capturing the changes in the market conditions through the observed prices. The state m_t is modeled as a stationary ergodic two-state Markov chain on M with transition probabilities given by:

$$P\left(m_t = j | m_{t-1} = i, m_{t-2} = k, \ldots\right) = P\left(m_t = j | m_{t-1} = i\right) = P_{j,i} \tag{11}$$

for $i, j, k \in M$, where $P_{j,i}$ indicates the probability that state (regime) i transits to state j for $i, j \in \{0, 1\}$. The transition probabilities are constants and satisfy the conditions of $\sum_{j=0}^{1} P_{j,i} = 1$ and $0 \leq P_{j,i} \leq 1$ for $i = 0, 1$. The state m_t is a random variable that is not directly observable. However, a filter estimate can be computed from the time series of the exchange rate. Some filters, such as sequential filter, are capable of performing accurate inferences of m_t. It is therefore reasonable to assume that investment professionals can estimate the state with high precision. The regime-dependent α_{ft} and λ_t are then given by:

$$\alpha_{ft} = \begin{cases} \alpha_{f0}, & m_t = 0, \\ \alpha_{f1}, & m_t = 1. \end{cases} \tag{12}$$

and

$$\lambda_t = \begin{cases} \lambda_0, & m_t = 0, \\ \lambda_1, & m_t = 1. \end{cases} \tag{13}$$

The noise term ε_t is assumed to be drawn from an $N(0, \sigma_t^2)$ distribution and σ_t^2 is regime-dependent, that is:

$$\varepsilon_t \sim \begin{cases} N(0, \sigma_0^2), & m_t = 0, \\ N(0, \sigma_1^2), & m_t = 1. \end{cases} \tag{14}$$

In Chiarella et al. (2012), the expectation formation process is regime-dependent, but the fraction of fundamentalists and chartists are constant such that $\omega_{f,t} \equiv \omega_f$ and $\omega_{c,t} \equiv \omega_c$. Under the Markov regime-switching (MS) process, the price dynamic function is given by:

$$\Delta S_t = \gamma \omega_f \alpha_{ft} (u_t - S_{t-1}) + \gamma \omega_c \lambda_t (S_{t-1} - S_{t-2}) + \varepsilon_t.$$

The parameters estimated are $\gamma\omega_f\alpha_{f0}$, $\gamma\omega_f\alpha_{f1}$, $\gamma\omega_c\lambda_0$, $\gamma\omega_c\lambda_1$, σ_0, σ_1, $P_{0,0}$ and $P_{0,1}$.

3 Methodology and Data Description

3.1 Data

We use period average AUD/USD monthly exchange rate, Australia and US money market rates from 2000:1 to 2013:6. The use of monthly data is consistent with most literature in HAM for the foreign exchange market, for example, de Jong et al. (2010), Jongen et al. (2012) and Spronk et al. (2013). Using data with lower frequency may

smooth out much of the behavioural changes of the traders.[3] Even though foreign exchange rate data are available at high frequency, we use monthly data for two reasons. The first is to avoid directly addressing issues arising from short-term noise such as the day-of-the-week effect in exchange rate volatility (see Hsieh 1988). The second is to link the decision-making process to macroeconomic data that are typically available at a relatively low frequency.

Other data collected are potential transition variables, which are used in the LSTR model. These include real effective exchange rate (REER), real GDP (GDP),[4] unemployment rate (UNE), real industrial production (IND), consumer price index (CPI), money supply (M1), short-term (STY) and long-term (LTY) interest rates measured by government bond yield. Among these variables, REER, GDP, IND, CPI and M1 are measured in month-on-month growth rates. Since nominal exchange rate is a relative measure, we take country-on-country differences (Australian value minus US value) of each potential transition variable.

The statistics of all these variables are summarised in Table 1. Figure 1 shows the nominal exchange rate and its deviation from the fundamental value. Over the sample period, the Australian dollar has in general appreciated strongly against the US dollar and AUD/USD exchange rate has been around or below the fundamental value, with two notable exceptions of continuing depreciation. Between 2000 and 2001, due to falling market confidence and rising concerns over political instability, the Australian dollar fell to its lowest level since the currency was floated in 1983, with large deviations from its fundamental value. The second notable depreciation episode happened between August 2008 and March 2009 during the Global Financial Crisis, where the nominal exchange rate exhibits unprecedent deviations from its fundamental value.

The summary statistics suggest a large fluctuation in the AUD/USD exchange rate with a minimum value of 0.929, and a maximum value of 1.996, and a standard deviation of 0.315. Its maximum deviation from the fundamental exchange rate is 0.218 and on average it has been below its fundamental value. This deviation arises from the interest rate differences between the two economies. The Australia average money market rate of 4.990 % has been more than twice of that of the US 2.225 %. The real effective exchange rate growth rates in Australia and the US exhibit significant differences, with the maximum of 8 % and the minimum of −18 %. Such differences are also observed in other variables, with industrial production and money supply growth rates being more significant. All potential transition variables are then standardised to accommodate the numerical estimation of the nonlinear model.

[3] Some empirical studies with HAM also use annual data of the financial market indices. For example, Boswijk et al. (2007) use annual S&P500 index since the index can be traced back to 1871, and thus, have a large sample even with annual data. However, Australia adopted a fixed exchange rate regime prior to 1984. Therefore, using annual exchange rate data would largely reduce our sample size and negatively affect the asymptotic properties of the estimates.

[4] Since real GDP data is available at quarterly frequency only, we first calculate the quarterly growth rate and then calculate the average monthly growth rate assuming a geometric growth pattern in each quarter.

Table 1 Summary statistics

Variable	Mean	Standard deviation	Minimum	Maximum
S_t	1.350	0.315	0.929	1.996
$S_t - u_t$	−0.006	0.044	−0.104	0.218
r_t^{AU}	4.990	1.086	2.750	7.250
r_t^{US}	2.225	2.119	0.070	6.540
REER	0.297	3.098	−18.753	8.003
GDP	0.098	0.230	−0.372	0.775
UNE	−0.941	2.096	−4.692	2.610
IND	1.405	2.949	−7.346	9.054
CPI	0.046	0.322	−0.805	1.494
M1	−0.003	2.142	−15.885	5.491
STY	2.217	1.185	−0.400	4.555
LTY	1.368	0.602	0.026	2.618

Notes all numbers are expressed in percentage, except S_t and $S_t - u_t$

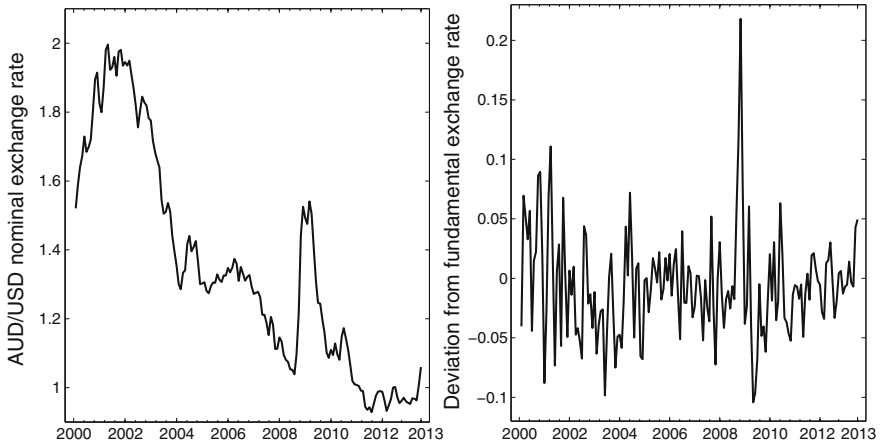

Fig. 1 Nominal exchange rate and its deviation from fundamental exchange rate

3.2 Linearity Tests

To determine which set of variables are valid transition variables in the LSTR model, following Luukkonen et al. (1988), linearity tests based on Taylor approximations of the model are performed. The potential transition variables are divided into four groups: (1) foreign exchange market indicators (REER); (2) business cycle indicators (GDP, UNE, IND); (3) money supply (M1) and inflation rate (CPI); and (4) interest rates (STY, LTY). First, we consider the univariate transition function and select the variable that yields the strongest rejection of linearity. Next, we consider multivariate transition functions with two, three and four transition variables, separately. Linearity tests are performed on each possible set of two to four transition variables. We never

Table 2 Estimation results for the BHM model

Parameter	$\gamma\alpha_f$	$\gamma\lambda$	ρ/γ
Coefficient	−2.225	0.723	17.571
p-value	0.270	0.000	0.738

include more than one variable from each of the four groups. This approach allows us to avoid multicollinearity within the transition function since variables within each group are likely to be highly correlated. Then, for each of the two-, three- and four-transition-variable case, we choose the set of variables that yields the strongest linearity rejection as the optimal set of transition variables. We select the set of transition variables {REER, CPI, STY} from the linearity tests since it exhibits the best goodness-of-fit.

4 Estimation Results

In this section, we present the estimation results from the proposed model with the three types of regime switching, including (1) switching between strategies based on past performance (BHM), (2) switching between strategies based on macroeconomic fundamentals (LSTR) and (3) switching between regime-dependent beliefs (MS).

4.1 BHM Estimation Results

The parameter estimates for the BHM-type switching, defined by Eq. (8), are presented in Table 2. The model is estimated by nonlinear least squares following Boswijk et al. (2007). The logit switching rule in the BHM model essentially represents a special case of the generalised logistic smooth transition where the transition variables are differences in the realised profits of fundamentalists and chartists. The estimate of $\gamma\alpha_f$ is negative (−2.225). This implies that among the fundamentalists, carry traders dominate the foreign exchange market. The estimate of $\gamma\lambda$ is positive (0.723). This suggests that among the chartists bandwagon expectations dominate. The intensity of choice parameter ρ is not identified in the estimation procedure, but is captured by ρ/γ (17.571). This is not statistically significant, suggesting that there is no direct evidence of switching between the fundamentalists and the chartists based on past performance. This implies that the relative profitability of the two trading rules is not a significant strategy switch in the AUD/USD foreign exchange market. However, this result does not exclude the possible switch driven by other mechanisms (for example, LSTR and/or MS).[5]

[5] No switching can be taken as a criteria to exclude a model since the existence of strategy switching in the AUD/USD foreign exchange market is not a stylised fact. The relatively poor performance of BHM in the AUD/USD foreign exchange market, however, cannot and should not be generalised to the other financial markets. There is significant difference between equity market and foreign

The upper panel of Fig. 2 plots the estimated evolution of the fraction of fundamentalists and the AUD/USD exchange rate. The lower panel of Fig. 2 is a scatter plot of the fraction of fundamentalists $\omega_{f,t}$ against the relative profitability of fundamentalism and chartism trading rules captured by $\gamma \Delta S_{t-1}[\lambda_t(S_{t-2} - S_{t-3}) - \alpha_f(u_{t-1} - S_{t-2})]$. As suggested by the estimated coefficients in Table 2, there is no significant fluctuation in the fraction of fundamentalists, staying around 0.5. The scatter plot is a relatively flat line, indicating that agents respond sluggishly to differences in performance. The negative slope of the scatter plot, albeit small in absolute magnitude, somehow indicates that a positive difference in profits between chartism and fundamentalism strategy results in a smaller fraction of fundamentalists. Around the period of September 2008, the fraction of fundamentalists deviated from its mean level, suggesting some switching from fundamentalists to chartists with the relative appreciating US dollar.

4.2 LSTR Estimation Results

Following Lof (2012), we estimate the switching between strategies based on macroeconomic fundamentals, defined by Eq. (10), using nonlinear least squares, for each set of transition variables separately. In the LSTR model with a multivariate transition function, τ, A and c cannot be all identified at the same time without further restrictions. Therefore, we restrict the summation of the elements in A to be one such that $X_{t-1}A$ represents the weighted sum of multiple transition variables. The estimation results with three transition variables {REER, CPI, STY} are presented since it has the best goodness-of-fit. The other linearity test results and estimation results are discussed in the appendices.

Table 3 summarises the estimation results. The results identify two distinct regimes with statistically significant $\gamma\alpha_f$ and $\gamma\lambda$. Since $\gamma\alpha_f < 0$ (-19.553), among the fundamentalists carry traders dominate. On the other hand, $\gamma\lambda > 0$ (0.370) suggests that among the chartists bandwagon expectations dominate. These results are found to be consistent with those estimated using the BHM model. The significance of $\gamma\alpha_f$ and $\gamma\lambda$ also implies the presence of between-group behavioural heterogeneity. The intensity of choice parameter τ (-0.933) supports a smooth transition between regimes, but this is found to be not significant. However, the insignificance of the intensity of choice parameter with a large standard deviation is a common result in switching-type regression models since large changes in τ only cause a small variation of the fraction of fundamentalists $\omega_{f,t}$ (see for example, Boswijk et al. 2007 and de Jong et al. 2010). Teräsvirta (1994) suggests that this effect is not relevant as long as there is significant heterogeneity in the estimated regimes.

(Footnote 5 continued)
exchange market. For example, investors may hold the stock position for years, while currency traders typically close their positions daily, which may result in that BHM can capture the switching behaviour in stock market with yearly data while failing to capture the switching behaviour in foreign exchange market even with monthly data.

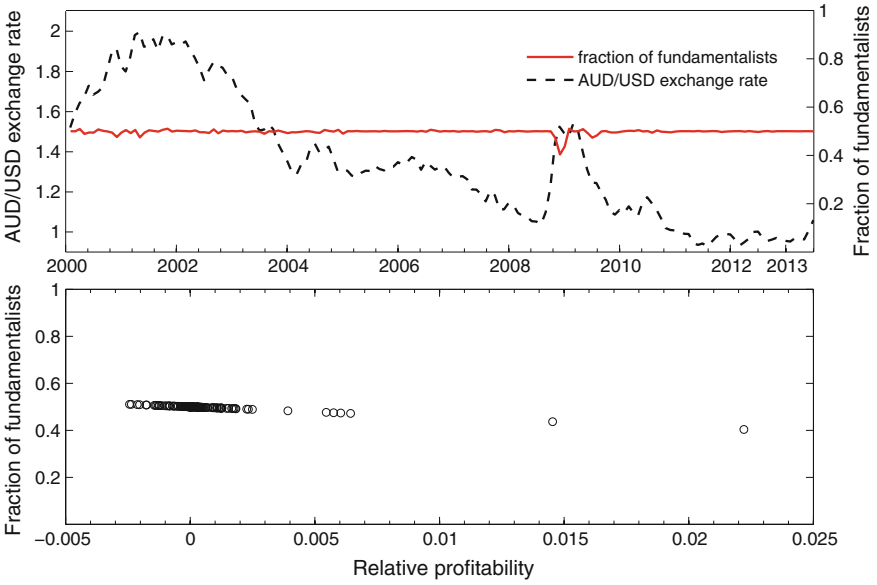

Fig. 2 Fraction of fundamentalists estimated from BHM

Table 3 Estimation results for the LSTR model

Parameter	$\gamma\alpha_f$	$\gamma\lambda$	τ	
Coefficient	−19.553	0.370	−0.933	
p-value	0.000	0.000	0.605	
Parameter	a_1	a_2	a_3	c
Coefficient	3.558	−0.283	−2.275	3.365
p-value	0.000	0.190	0.000	0.000

Notes a_1, a_2 and a_3 are the coefficients of the transitions variables REER, CPI and STY, respectively

The interpretation of $A = (a_1, a_2, a_3)^T$ reveals that fundamentalists dominate during the periods of relative lower growth in Australia compared to the US. Real effective exchange rate growth (REER) has a positive coefficient a_1 (3.558). This implies that a relative economic downturn in Australia depreciates the Australia real effective exchange rate and appreciates the US real effective exchange rate, and thus, increases REER, causing an increase in the fraction of fundamentalists. Inflation rate (CPI) has a negative but not significant coefficient a_2 (−0.283). A lower inflation rate in Australia indicates a relative economic downturn in Australia, and a larger fraction of fundamentalists in this model. Also the short-term government bond yield (STY) has a negative and significant coefficient a_3 (−2.275). A relative low yield on Australian government bonds (low-risk assets) implies high levels risk aversion, and in this model a high fraction of fundamentalists. In summary, fundamentalism is the dominant strategy during low economic growth periods in Australia.

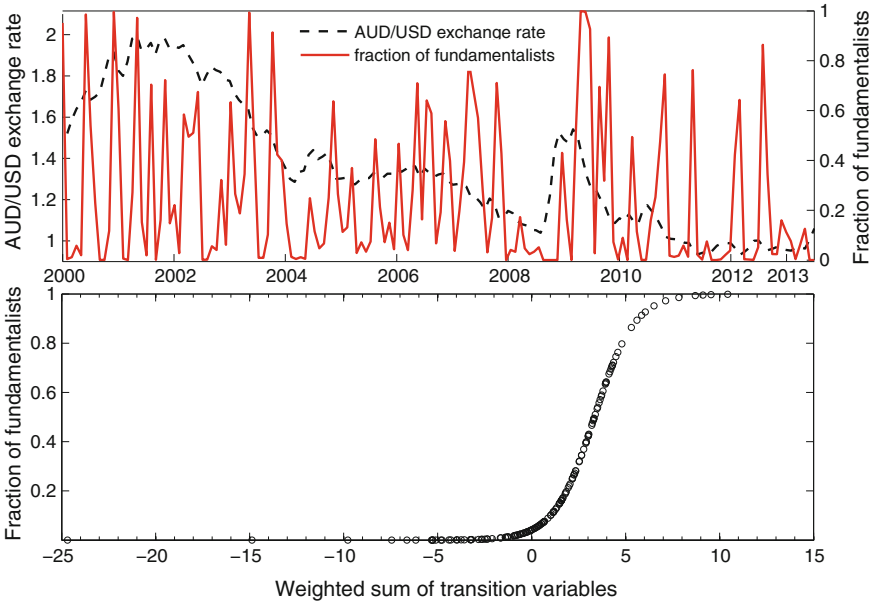

Fig. 3 Fraction of fundamentalists estimated from LSTR

The upper panel of Fig. 3 shows the plot of the fraction of fundamentalists and the AUD/USD nominal exchange rate overtime. The lower panel of Fig. 3 is the scatter plot of the fraction of fundamentalists $\omega_{f,t}$ against the weighted sum of the transition variables $X_{t-1}A$. Most of the time, the economy is represented by both fundamentalists and chartists, with $\omega_{f,t}$ ranges between 0 and 1. Overtime, chartists dominate since the fraction of fundamentalists are mostly below 0.5. The fraction of fundamentalists increases when S_t goes down, that is, the US dollar depreciates against the Australian dollar. In 2008, the market was dominated mostly by the chartists for a prolonged period, when the US dollar appreciated substantially against the Australian dollar. The scatter plot clearly shows a logistic curve, suggesting a smooth transition between fundamentalists and chartists. The curve is more dense in the lower part (below 0.5), indicating a relative smaller portion of fundamentalists in the market.

4.3 MS Estimation Results

The Markov regime switching model, defined by Eqs. (12)–(14), is estimated using maximum likelihood (Hamilton 1994, Chap. 22). The estimation results are summarised in Table 4. We use state 0 to denote a boom state where the Australia dollar is appreciating (decreasing S_t) against the US dollar with a relatively low volatility

Table 4 Estimation results for the MS model

Parameter	$\gamma\omega_f\alpha_{f0}$	$\gamma\omega_f\alpha_{f1}$	$\gamma\omega_c\lambda_0$	$\gamma\omega_c\lambda_1$	σ_0	σ_1	$P_{0,0}$	$P_{0,1}$
Coefficient	−2.231	2.043	0.223	0.439	0.029	0.062	0.986	0.069
p-value	0.010	0.592	0.021	0.007	0.000	0.000	0.000	0.167

(low risk) and state 1 to denote a bust state where the Australia dollar is depreciating (increasing S_t) against the US dollar with a relatively high volatility (high risk).

Since both γ and ω_f are positive, $\gamma\omega_f\alpha_{f0} < 0$ and $\gamma\omega_f\alpha_{f1} > 0$ suggest that $\alpha_{f0} < 0$ and $\alpha_{f1} > 0$. This implies the dominance of carry traders in the boom state (state 0) and UIP traders dominate in the bust state (state 1). In bust state (state 1), the large deviation of AUD/USD nominal exchange rate from its fundamental value as shown in Fig. 1 fades away quickly, as the UIP traders who are dominating are mean-reverting. As most of the sample period is covered by the boom state (state 0), we should expect an overall dominance of carry traders from 2000:1 to 2013:6. This is consistent with the observations from the estimation results of the BHM and LSTR models. The positive estimates of $\gamma\omega_c\lambda_0$ and $\gamma\omega_c\lambda_1$ suggest that the chartists expect the exchange rate movement to persist even though the intensity of the bandwagon effect may be different. This finding is also consistent with the results estimated based on the BHM and LSTR models. In the bust state, the degree of bandwagon expectation is larger than that in the boom state, suggesting a more pessimistic sentiment when the Australian dollar is depreciating against the US dollar. Both the different behaviour among the fundamentalists and the different behaviour among the chartists in different states provide evidence of within-group heterogeneity over time. The regime-dependent standard deviations are estimated to be $\sigma_0 = 0.029$ and $\sigma_1 = 0.062$ in states 0 and 1, respectively. The volatility in bust state is twice as much as that in the boom state, suggesting that the market is more sensitive to external news/shocks, and thus, exhibits higher volatility in the bust state than in the boom state.

The switching of the beliefs of the fundamentalists and chartists and the changing market volatilities between the two states are indicated by the transition probabilities. The fundamentalists do not fix their strategies over time. Instead, they switch from UIP traders to carry traders or vice versa with a time-varying probability. Similarly, the chartists adjust their degree of bandwagon expectation according to the market condition, which can be differentiated by the state variable m_t. In addition, the overall market sensitivity to external news/shocks are conditioned on the state variable. The results in Table 4 show that the probability of remaining in the boom state is 0.986 ($P_{0,0}$), suggesting that the boom state is persistent on an average of $1/(1-P_{0,0}) \approx 71$ months. It means that the probability for the fundamentalists to stay in the carry trading strategy, the chartists to have a smaller degree of bandwagon expectation and the market to be less volatile is 0.986. Given the two states identified, this also means that the probability for the fundamentalists and chartists to switch their beliefs from the boom state to the bust state is 0.014 ($P_{1,0}$). Similarly, the fundamentalists

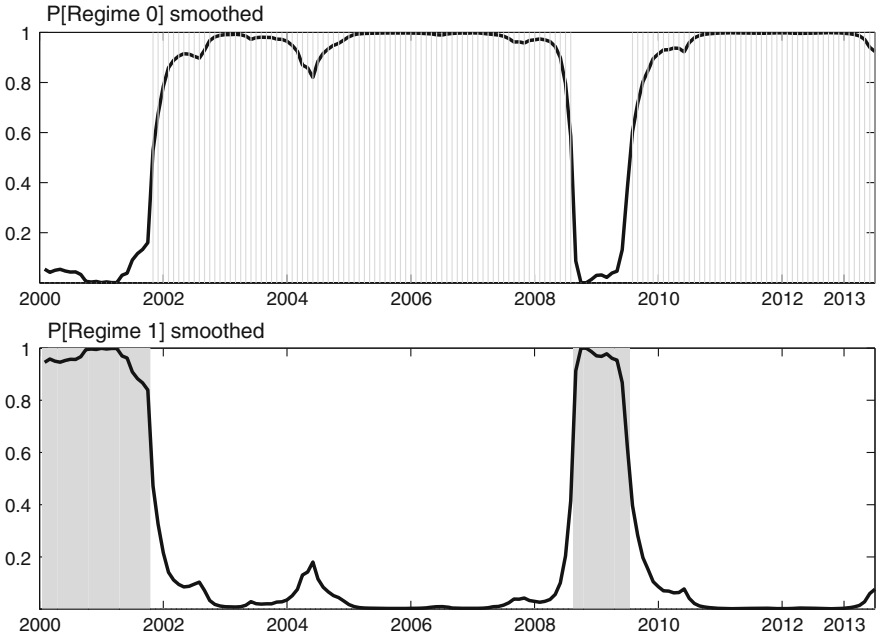

Fig. 4 Smoothed transition probability

maintain their UIP trading strategies and the chartists maintain higher degree of bandwagon expectation in the bust state with a probability 0.931 ($P_{1,1}$), and switch to the boom state with a probability of 0.069 ($P_{0,1}$).

With the estimates of the transition probabilities, we can further calculate the smoothed probability $P(m_t = j|p_1, \ldots, p_N)$ for each period, conditioning on the whole price series $\{p_1, \ldots, p_N\}$ (for details of the algorithm, see Kim and Nelson (1999)). Figure 4 shows the smoothed probability for α_{ft}, λ_t and σ_t to fall into the two states over the sample period of January 2000 to June 2013. The bust periods corresponding to state 1 cover the 2000–2001 falling market confidence in Australia and the 2008–2009 global financial crisis. The boom periods corresponding to state 0 include the time from late 2001 to early 2008, when the financial market was prospering in Australia. Recently, after the Australian dollar depreciates to the bottom in the late 2008, fundamentalists and chartists became optimistic with the global recovery and switched their estimates from regime 1 to regime 0.

Overall, the MS estimation results provide evidence on the coexistence of time-varying within-group heterogeneity and between-group heterogeneity. Most interestingly, although the dates of domestic economic instabilities and financial crises are not used in any way to estimate the parameters or form inference about transition probabilities, the classified regimes match well with the market booms and busts in actual episodes.

Table 5 Goodness-of-fit and misspecification tests

Model	Log-likelihood	AIC	Linearity LR	Jarque-Bera	AC(12)	ARCH(12)
BHM	291.939	−3.567	0.729	0.000	0.129	0.214
LSTR	318.272	−3.855	0.000	0.000	0.570	0.010
MS	305.850	−3.677	0.000	0.691	0.315	0.334

5 Efficiency Tests

In this section, we apply several measures to evaluate the goodness-of-fit of the three regime-switching models and aim to identify the most well-specified model. Table 5 presents the log-likelihood, Akaike Information Criterion (AIC),[6] the p-values of the linearity likelihood ratio test and the three misspecification diagnostics computed from the model residuals. Following Deschamps (2008), the three misspecification diagnostic tests are:

1. The Jarque-Bera statistic, used as an indicator of error non-normality;
2. A χ^2-statistic for the nullity of the autoregression coefficients in an AR(12) model of the residuals. This is used as an indicator of error autocorrelation, and is denoted by AC(12).
3. An F-statistic for the nullity of the autoregression coefficients in an AR(12) models of the squared residuals. This is used as an indicator of error conditional heteroscedasticity, and is denoted by ARCH(12).

In terms of in-sample estimation, the log-likelihood and AIC suggest that the LSTR model has a better fit than the BHM and MS models. The p-values for the likelihood ratio tests strongly reject the null hypothesis of linearity in favour of the LSTR and MS models. However, when comparing the BHM model with the linear model, the results fails to reject the linearity. This finding is consistent with the BHM estimation results that there is no significant switching between the chartists and fundamentalists. The estimated p-values of the Jarque-Bera statistics indicate non-normality of residuals for the BHM and LSTR models. This result could be attributed to the different volatilities of the foreign exchange market in different states, which are modelled explicitly in the MS model, but not the BHM or the LSTR models. The estimated p-values for error autocorrelation (AC(12)) and error conditional heteroscedasticity (ARCH(12)) are all larger than 0.01, with the exception of that for ARCH(12) in the LSTR model. Therefore, we conclude that there is weak evidence of conditional heteroscedasticity in this model.

[6] In comparing the in-sample fit, we use the AIC in addition to the log-likelihood. The advantage of AIC is that it can deal with the trade-off between the goodness of fit of the model and the complexity of the model. AIC not only rewards goodness of fit, but also includes a penalty that is an increasing function of the number of estimated parameters. This penalty discourages over-fitting.

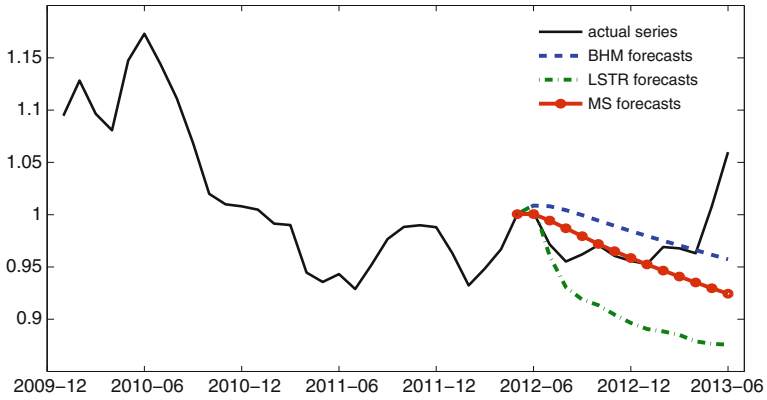

Fig. 5 Comparison of out-of-sample predictability

6 Predictive Power

Besides the in-sample efficiency, another important criterion to evaluate the different model specifications is the out-of-sample predictability. We compare the out-of-sample forecasting accuracy of the three regime-switching models, by dividing the data sample into two segments: one for the in-sample estimation and the other for the out-of-sample comparison. Specifically, we use data from 2000:1 to 2012:5 to estimate the models, on the basis of which we forecast the exchange rate from 2012:6 to 2013:6. The forecast series are then compared with the actual exchange rate as shown in Fig. 5. The model with switching between strategies based on macroeconomic fundamentals (LSTR) exhibits better forecasting accuracy in the short run (3-month horizon). However, the other two regime-switching models (BHM and MS) tend to outperform in a longer horizon. The regime-switching based on past performance (BHM) gives the best out-of-sample prediction accuracy in the long run.

In the above exercises, the separation of the sample is somewhat arbitrary, a natural question is whether the results are sensitive to the sample selection. As a robustness check, we apply a rolling forecasting technique with a fixed sample size and then compare the forecasting performance using the root-mean-squared forecast error (RMSE) and mean absolute error (MAE) for different forecasting horizons. The significance of the difference in forecasting differences is tested by the Diebold and Mariano (1995) test statistic.

Specifically, for each regime-switching model, we calculate the h-months-ahead forecasting error, for $h = 1, 2, \ldots, 12$. The forecasting recursion is based on a rolling estimation window with a fixed sample size, which is 108 (9 years of data) in our practice. First, the model is estimated for the period 2000:1 to 2008:12, that is, the first 108 observations ($S_1, S_2, \ldots, S_{108}$), and then obtain the first h-month-ahead forecasts ($\hat{S}_{108+1}, \ldots, \hat{S}_{108+h}$). Similarly, we obtain the second h-month-ahead forecasts ($\hat{S}_{109+1}, \ldots, \hat{S}_{109+h}$) from observations ($S_2, S_3, \ldots, S_{109}$). Such a process is

Table 6 Forecasting performance

Steps	Panel A: ratio of RMSE's			Panel B: ratio of MAE's			Panel C: Diebold-Mariano		
	BHM /LSTR	BHM /MS	LSTR /MS	BHM /LSTR	BHM /MS	LSTR /MS	BHM /LSTR	BHM /MS	LSTR /MS
1	1.396	0.928	0.665	1.515	0.937	0.618	3.506***	−2.057***	−3.175***
2	1.267	0.913	0.720	1.222	0.944	0.772	1.181	−1.113	−1.165
3	1.087	0.872	0.802	0.968	0.912	0.942	0.375	−1.220	−0.707
4	0.980	0.833	0.850	0.922	0.881	0.955	−0.086	−1.220	−0.476
5	0.884	0.803	0.909	0.857	0.846	0.987	−0.509	−1.218	−0.265
6	0.855	0.778	0.910	0.777	0.798	1.026	−0.556	−1.218	−0.237
7	0.821	0.769	0.937	0.764	0.784	1.025	−0.739	−1.237	−0.169
8	0.787	0.763	0.969	0.767	0.768	1.001	−1.090	−1.230	−0.094
9	0.721	0.753	1.044	0.721	0.760	1.055	−2.215**	−1.242	0.149
10	0.666	0.743	1.116	0.681	0.767	1.126	−6.296***	−1.259	0.463
11	0.612	0.725	1.184	0.614	0.733	1.192	−3.465***	−1.302	0.668
12	0.559	0.707	1.264	0.566	0.702	1.240	−27.52***	−1.402	1.095

Notes *, ** and *** indicate significance at the 10, 5 and 1 % level, respectively

repeated until we obtain the last h-month ahead forecasts $(\hat{S}_{162-h+1}, \ldots, \hat{S}_{162})$ from observations $(S_{55}, S_{56}, \ldots, S_{162-h})$ (162 is the total number of observations in the full sample). This recursion leads to a sequence of $k_h = 162 - 108 + 1 - h = 55 - h$ density forecasts. Based on recursive forecasting, we calculate h-step-ahead RMSE and MAE as:

$$\text{RMSE}_h = \sqrt{\frac{\sum_{t=108}^{t=162-h} (\hat{S}_{t+h} - S_{t+h})}{k_h}}, \qquad \text{MAE} = \frac{\sum_{t=108}^{t=162-h} |\hat{S}_{t+h} - S_{t+h}|}{k_h}$$

Table 6 presents the RMSE's (Panel A) and MAE's (Panel B) of the three models and the Diebold-Mariano test statistics (Panel C) for the 12 forecast horizons. Note that the smaller the value of the RMSE and MAE, the better the forecasting accuracy. Panel A contains the ratio of the RMSE of any two models. They are BHM versus LSTR, BHM versus MS and LSTR versus MS. A ratio <1 (>1) indicates a better forecasting performance for the model mentioned first (second). Panel C reports the Diebold-Mariano test statistics of equality of forecasting performance. We use the RMSE as loss function with a rectangular lag window with $h - 1$ sample autocovariances for the h-step-ahead forecast error. A negative (positive) number indicates better performance for the model mentioned first (second).

The results in Panels A and B suggest that the LSTR model generally outperforms both the BHM model and the MS model in short-run forecasts (up to 3-month horizon), while both the MS and BHM models outperform the LSTR model in the medium to long run. The BHM model always performs better than the MS model in terms of RMSE and MAE. For both the BHM and MS model, we observe that the forecasting power of the model vis-à-vis the LSTR model generally improves as the forecasting horizon increases. Similar observations are found in the the forecasting

power of the model BHM vis-à-vis the MS model, although the improvement is relatively smaller.

In terms of significant improvements as presented in Panel C, we observe a similar pattern. In terms of the 1-month-ahead forecasts, the differences of the three models are highly significant, where the LSTR model has the best performance while the MS has the worst performance. The good performance of the LSTR model dies out in the medium to long run. The BHM model performs significantly better than the LSTR in 9- to 12-month step ahead forecasts. The BHM model is always doing better than the MS model, however this difference is only significant for the 1-step-ahead forecast.

7 Conclusion

We evaluate the performance of three switching mechanisms, developed by Boswijk et al. (2007, BHM), Lof (2012, LSTR) and Chiarella et al. (2012, MS), in estimating behavioural heterogeneity. The BHM switching mechanism highlights the importance of past performance in determining the dynamic weight of heterogeneous trading rules. The LSTR switching model emphasizes the role of macroeconomic fundamental in shaping the trading behaviour and affecting the choice of the trading strategy. The MS model addresses how agents shift their expectations by extrapolating the financial market conditions from the price information and change their trading behaviour accordingly. Applying AUD/USD monthly exchange rate data from 2000:1 to 2013:6, we document empirical evidence on the comparative advantage of the three switching models. While the LSTR model provides better in-sample explanatory power than the other two, there is significant evidence to support the performance of the BHM model in terms of its out-of sample forecasting accuracy. There is, however, no significant evidence that either the BHM or the LSTR model outperforms the MS model in terms of predictive power. All the three models have consistently confirmed the presence of behaviour heterogeneity. More interestingly, both the LSTR and the MS models highlight a significant dominance of carry traders in the AUD/USD foreign exchange market.[7] The carry traders sell the currency of low-interest country and buy the currency of high-interest country expecting to profit from interest rate difference. This finding is in line with other findings in literature that the influence of carry traders on exchange rates is large and increasing (Galati et al. 2007; Pojarliev and Levich 2011). The role of carry traders has also been highlighted by Spronk et al. (2013) using simulations to generate stylised facts observed in empirical exchange rates.

In reality, the three different types of regime switching mechanisms can coexist. In this chapter, we isolate the three types and investigate which type can fit the

[7] The results however cannot rule out the presence of UIP traders. The BHM model also suggests the dominance of carry traders, even though it is not statistically significant. The conclusion, however, is restricted to AUD/USD exchange rate and hence should not be generalised.

AUD/USD exchange rate data better. But this superiority is not in absolute terms and we cannot rule out the coexistence of other types of switching in the market. The relative good performance of the LSTR model suggests that agents in the AUD/USD foreign exchange market are more responsive to the changes in the macroeconomic fundamentals than to the past profit and to the market states.

Future research in this area may include an investigation of the other financial markets, such as commodity market and stock market, in order to compare the relative performance of the three switching models in different markets. Another future research direction would be to improve the empirical estimation of the fundamental exchange rate, by taking into account the central bank's monetary policy. Other possible directions include innovating a model, which is able to integrate the three types of regime switching and check whether such an integration improves the overall empirical performance.

Acknowledgments We thank the two anonymous referees for their helpful comments and suggestions. All remaining errors and omissions are our own.

References

Allen, H., & Mark, P. T. (1990). Charts, noise and fundamentals in the London foreign exchange market. *The Economic Journal*, *100*, 49–59.

Boswijk, H. P., Hommes, C. H., & Manzan, S. (2007). Behavioral heterogeneity in stock prices. *Journal of Economic Dynamics and Control*, *31*, 1938–1970.

Brock, W. A., & Hommes, C. H. (1998). Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *Journal of Economic Dynamics and Control*, *22*, 1235–1274.

Cheung, Y. W., & Chinn, M. D. (2001). Currency traders and exchange rate dynamics: A survey of the US market. *Journal of International Money and Finance*, *20*, 439–471.

Chiarella, C., He, X., Huang, W., & Zheng, H. (2012). Estimating behavioural heterogeneity under regime switching. *Journal of Economic Behavior and Organization*, *83*, 446–460.

De Grauwe, P., & Grimaldi, M. (2006). Exchange rate puzzles: A tale of switching attractors. *European Economic Review*, *50*, 1–33.

de Jong, E., Verschoor, W., & Zwinkels, R. (2010). Heterogeneity of agents and exchange rate dynamics: Evidence from the EMS. *Journal of International Money and Finance*, *29*, 1652–1669.

Deschamps, P. J. (2008). Comparing smooth transition and Markov switching autoregressive models of US unemployment. *Journal of Applied Econometrics*, *23*, 435–462.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*, 253–263.

Frijns, B., Lehnert, T., & Zwinkels, R. (2010). Behavioral heterogeneity in the option market. *Journal of Economic Dynamics and Control*, *34*, 2273–2287.

Galati, G., Health, A., & McGuire, P. (2007). Evidence of carry trade activity. *BIS Quarterly Review*, *3*, 27–41.

Gehrig, T., & Menkhoff, L. (2004). The use of flow analysis in foreign exchange: Exploratory evidence. *Journal of International Money and Finance*, *23*, 573–594.

Gilli, M., & Winker, P. (2003). A global optimization heuristic for estimating agent based models. *Computational Statistics and Data Analysis*, *42*, 299–312.

Hamilton, J. D. (1994). *Time series analysis*. Princeton: Princeton University Press.

Hommes, C. (2011). The heterogeneous expectations hypothesis: Some evidence from the lab. *Journal of Economic Dynamics and Control*, *35*, 1–24.
Hsieh, D. A. (1988). The statistical properties of daily foreign exchange rates, 1974–1983. *Journal of International Economics*, *24*, 129–145.
Huang, W., & Zheng, H. (2012). Financial crises and regime-dependent dynamics. *Journal of Economic Behavior and Organization*, *82*, 445–461.
Huang, W., Zheng, H., & Chia, W. M. (2010). Financial crises and interacting heterogeneous agents. *Journal of Economic Dynamics and Control*, *34*, 1105–1122.
Huang, W., Zheng, H., & Chia, W. M. (2012). Asymmetric returns, gradual bubbles and sudden crashes. *The European Journal of Finance*, *19*, 420–437.
Jongen, R., Verschoor, W., Wolff, C., & Zwinkels, R. (2012). Explaining dispersion in foreign exchange expectations: A heterogeneous agent approach. *Journal of Economic Dynamics and Control*, *36*, 719–735.
Kim, C. J., & Nelson, C. R. (1999). *State space models with regime switching: Classical and Gibbs sampling approaches with applications*. MA: MIT Press.
Lof, M. (2012). Heterogeneity in stock prices: A STAR model with multivariate transition function. *Journal of Economic Dynamics and Control*, *36*, 1845–1854.
Luukkonen, R., Saikkonen, P., & Teräsvirta, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika*, *75*, 491–499.
Mark, N. C. (1995). Exchange rates and fundamentals: evidence on long-horizon predictability. *American Economic Review*, *85*, 201–218.
Manzan, S., & Westerhoff, F. (2005). Representativeness of news and exchange rate dynamics. *Journal of Economic Dynamics and Control*, *29*, 677–689.
Manzan, S., & Westerhoff, F. (2007). Heterogeneous expectations, exchange rate dynamics and predictability. *Journal of Economic Behavior and Organization*, *64*, 111–128.
Meese, R. A., & Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, *14*, 3–24.
Menkhoff, L., Rebitzky, R., & Schroder, M. (2009). Heterogeneity in exchange rate expectations: Evidence on the chartist-fundamentalist approach. *Journal of Economic Behavior and Organization*, *70*, 241–252.
Menkhoff, L., & Taylor, M. P. (2007). The obstinate passion of foreign exchange professionals: Technical analysis. *Journal of Economic Literature*, *45*, 936–972.
Neely, C. J., & Sarno, L. (2002). How well do monetary fundamentals forecast exchange rates? *Review, Federal Reserve Bank of St. Louis*, *84*(5), 51–74.
Pojarliev, M., & Levich, R. (2011). Detecting crowded trades in currency funds. *Financial Analyst Journal*, *67*, 26–39.
Spronk, R., Verschoor, W., & Zwinkels, R. (2013). Carry trade and foreign exchange rate puzzles. *European Economic Review*, *60*, 17–31.
Taylor, M. P. (1995). The economics of exchange rates. *Journal of Economic Literature*, *33*, 13–47.
ter Ellen, S., Verschoor, W., & Zwinkels, R. (2013). Dynamic expectation formation in the foreign exchange market. *Journal of International Money and Finance*, *37*, 75–97.
Teräsvirta, T. (1994). Specification, estimation and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, *89*, 208–218.
The Economist. (2013). *Buoyant, the Australian dollar*. The Economist, December 14th.
Westerhoff, F., & Reitz, S. (2005). Commodity price dynamics and the nonlinear market impact of technical traders: Empirical evidence for the US corn market. *Physica A: Statistical Mechanics and its Applications*, *349*, 641–648.
Yamamoto, R., Hirata, H. (2013). *Strategy switching in the Japanese stock market*. Japan Center for Economic Research Discussion Paper No. 135.

Time-Varying Cross-Speculation in Currency Futures Markets: An Empirical Analysis

Andreas Röthig and Andreea Röthig

1 Introduction

The past 20 years have seen significant changes in financial derivatives markets, especially regarding new types of market participants and new trading strategies using automated trading algorithms, but also regarding new regulation in the aftermath of financial shocks and crises, which were at least to some extent blamed on or related to financial derivatives. New traders such as index investors have entered derivatives markets and today hold a significant portion of open positions in futures markets. The rise of index traders is linked to new financial products, and in particular to exchange traded funds, which allow traders to participate from the performance of a benchmark or an index. These traders appear to constitute a new trader category since they act generally long-term, passive and long only. In addition, two large financial crises in 2001 and 2008 have rattled derivatives markets and led to a number of international regulatory initiatives. The most important of these initiatives include central clearing obligations and reporting to trade repositories which could have lasting effects on transparency, trading behaviors of market participants as well as on trading activities in general. Understanding the linkages between derivatives markets in this changing environment is of utmost importance for market participants and

The views expressed in this paper are the views of the authors and not necessarily those of the Deutsche Bundesbank or its staff.

A. Röthig (✉)
Deutsche Bundesbank, Wilhelm-Epstein-Strasse 14, 60431 Frankfurt am Main, Germany
e-mail: andreas.roethig@bundesbank.de

A. Röthig
Institute for Automatic Control and Mechatronics, Technische Universität Darmstadt, Landgraf-Georg Straße 4, 64283 Darmstadt, Germany
e-mail: aroethig@rmr.tu-darmstadt.de

R. Dieci et al. (eds.), *Nonlinear Economic Dynamics and Financial Modelling*,
DOI: 10.1007/978-3-319-07470-2_13,

regulators who are concerned with potential channels of contagion and systemic risk, but also with financial stability in general.

Linkages between different markets are often analyzed using price data. Relations between different price dynamics provide information about arbitrage opportunities, the appropriateness of hedging strategies or about basis risk in derivatives markets. In addition to price data, open position data for different types of traders are an important source of information. For instance, hedgers link different markets, most commonly the spot market for a specific asset with the respective derivatives market, in which the spot position is hedged. In addition to this type of hedging, where only a single asset type is involved, some hedging strategies involve several asset classes.[1] For example, a trader may decide to hedge an open spot position in a specific currency with futures contracts in gold. With this cross-hedge, the trader links to some degree the spot currency market to the gold derivatives market. Index traders by definition link a number of markets since they trade a benchmark or an index which generally include different assets. In addition, speculation plays a very important role, where speculative activities may link different futures markets. Cross-market speculative activities can arise from trading strategies involving fundamental and non-fundamental factors. Especially in currency futures markets, where market participants are generally active in different currencies, fundamental factors (e.g. macroeconomic data) which for example would show a weakening of the US dollar, could induce traders to simultaneously increase their positions in a number (or a portfolio) of different currencies, such as the Swiss Francs (CHF), the British Pound (GBP), the Canadian Dollar (CAD) and the Japanese Yen (JPY). While these fundamental factors should play a dominant role in the medium to long term, non-fundamental factors, which include overreaction to news, technical analysis and herding, could have a significant impact on traders' behavior over shorter horizons.[2]

Open interest data have been widely used since the mid-20th century to analyze the functioning and stability of futures markets.[3] The Commodity Futures Trading Commission's (CFTC) Commitments of Traders (COT) reports provide weekly information on reporting traders' open positions in futures markets. The CFTC classifies these positions as either commercial or non-commercial. A position is classified as commercial if the trader uses the futures contracts for hedging. The CFTC can re-classify the trader if the futures position is not entered into, in order to offset spot price risk. The distinction between speculation (i.e. non-commercials) and hedging might at times be fuzzy for specific markets.[4] However, in general the CFTC classification of traders' open interest has been regarded as accurate, and the COT

[1] See Fleming et al. (1998), Treepongkaruna and Gray (2009).

[2] See Manzan and Westerhoff (2007), Reitz and Westerhoff (2007) and Westerhoff (2008).

[3] See Houthakker (1957) and Working (1953), Working (1962).

[4] Ederington and Lee (2002) report for example, that commercial traders in the heating oil futures market might engage in speculation as well. However, Ederington and Lee (2002) note that the non-commercials are indeed speculators who do not engage in hedging activities. Sanders et al. (2004) stress that there are no obvious incentives for traders to self-classify as a speculator. Therefore, reporting non-commercials most likely represent speculative positions.

reports have been widely used in the literature.[5] Most of the literature concentrates on the relation between open interest and price dynamics. A number of studies for example analyze the forecasting performance of different types of traders or traders' impact on price volatility.[6] In addition, the theories of normal backwardation and hedging pressure can be investigated with open interest data.[7] Recently, studies have employed open interest data to analyze information flows and linkages between different types of traders. Röthig (2011) studies lead-lag relationships between hedging and speculation in currency futures markets and finds that speculators lead hedgers in these markets. Röthig and Chiarella (2011) find that non-reporting traders in currency futures markets behave more like speculators, rather than like hedgers. Röthig (2012) investigates the linkages between speculative activities in different currency futures markets. Using fixed parameter VAR models and impulse response analysis, Röthig (2012) shows that an increase in speculative activity in one currency futures market generally leads to an increase in speculative activity in another currency futures market. These empirical findings therefore point to cross-market herding activities of speculators in currency futures markets.

The current study builds on this previous research by allowing for time-variation in the response of speculative trading activity in one currency futures market to an increase in speculative trading activity in another currency futures market. Based on the findings of Röthig (2012) and the results of a number of bivariate fixed parameter VAR models for different currency pairs (the results not shown in this paper), we choose to analyze the impacts of an increase in speculative trading activity in the CHF futures market on speculative activities in the GBP, CAD and the JPY futures markets, using bivariate Bayesian time-varying VAR models. The investigation is carried out for (i) total speculation (long plus short speculative open interest), (ii) long speculation, and (iii) short speculation over the time period from January 1994 to September 2013. The results point to positive responses of total/long/short speculative activities in the GBP, CAD and JPY futures markets to an increase in total/long/short speculation in the CHF futures market. This indicates the presence of cross-market herding activities of speculators in these currency futures markets. Moreover, the cross-market linkages between speculative activities are relatively stable over the time period from 1994 to 2013. For all pairs of speculative activities investigated except one (CAD long), the signs and the durations of the responses to an increase in speculative activity in the CHF futures market do not change substantially over time. The results therefore do not suggest that changes in regulation, new market participants or new trading strategies had a significant and lasting impact on cross-market speculative activities in these futures markets.

---

[5] See Wang (2003) for a survey of the literature.

[6] See Adrangi and Chatrath (1998), Chatrath and Song (1999), Chatrath et al. (2003), Hartzmark (1987), Sanders et al. (2007), Schwarz (2012), Stein and Hong (1990), Tornell and Yuan (2012) and Wang (2002), Wang (2004).

[7] See for example Bryant et al. (2006).

The paper is organized as follows. In Sect. 2, the data and the estimation algorithm are presented. Section 3 reports the results of the impulse response analysis. Section 4 concludes.

2 Data and Model Description

This study investigates the relation between changes (i.e. first differences) in speculative (i.e. non-commercial) open interest in the CHF, GBP, CAD and JPY futures markets from 4 January 1994 to 17 September 2013.[8] The effects of an increase in speculative activity in the CHF futures market on speculative activities in the GBP, CAD and the JPY futures markets are investigated using bivariate Bayesian VAR models with time-varying parameters. The relations between (i) total speculation (long plus short), (ii) long speculation, and (iii) short speculation are investigated. The weekly open interest data are obtained from the CFTC COT report.[9]

The VAR model can be expressed as follows[10]:

$$Y_t = c_t + \sum_{l=1}^{L} B_{l,t} Y_{t-l} + \varepsilon_t, \tag{1}$$

where L denotes the lag length, ε_t is the error term with the covariance matrix $VAR(\varepsilon_t) = R$, and Y_t is the vector of the variables. In contrast to a fixed parameter VAR, this model allows for time variation in the parameters following the transition equation

$$\beta_t = \beta_{t-1} + e_t, \tag{2}$$

with $\beta_t = \{c_t, B_{1,t}, \ldots, B_{L,t}\}$ and the covariance matrix $\mathrm{VAR}(e_t) = Q$. We consider time-varying VAR models with two lags. The choice of the lag lengths is based on the Akaike, Hannan-Quinn and Schwartz information criteria, using fixed parameter VAR models as guidance.

The models are estimated using Bayesian methods similar to those described in Mumtaz and Sunder-Plassmann (2013). The first 64 observations of the sample are used as a training sample to estimate a fixed parameter VAR model, which provides initial starting values for the time-varying VAR model as well as for the prior dis-

[8] In fact, the investigation uses data from 13 October 1992 to 17 September 2013. However, the data from 13 October 1992 to 28 December 1993 are used as a training sample to obtain starting values for the estimation algorithm.

[9] The data are available on the CFTC's website at www.cftc.gov. For more information on the COT report, see Ederington and Lee (2002), Chatrath et al. (2003), Röthig (2007), Röthig and Chiarella (2011).

[10] The time series have been checked for stationarity and cointegration. Johansen tests show that the time series in levels are not cointegrated. Augmented Dickey Fuller (ADF) test results suggest that the time series in first differences are stationary.

tributions of R and Q, which are both assumed to be inverse Wishart. The training sample is then removed from the data and the remaining sample (4 January 1994 to 17 September 2013, 1029 observations) is used to estimate the time-varying VAR models. The basic algorithm for the estimation of the parameters of the Bayesian time-varying VAR model involves the sampling of the parameters from the conditional posterior distribution $H(\tilde{\beta_T} \backslash R, Q, \tilde{Y_t})$, with $\tilde{\beta_T} = [vec(\beta_1)', \ldots, vec(\beta_T)']$ and $\tilde{Y_t} = [Y_1, \ldots, Y_T]$, using the Carter and Kohn (1994) algorithm. Conditional on $\tilde{\beta_T}$, the covariance matrices Q and R are sampled from their conditional posterior distributions, which are both inverse Wishart. This study uses 110,000 Gibbs sampling iterations of which the last 1000 are used for inference.[11] The impulse response functions are calculated for each point in time using in each case the median impact value over the last 1000 samples, saved from the Gibbs sampling algorithm. This approach therefore allows for the computation of the impulse responses to a shock in speculative trading activity in CHF futures markets at each point in time and to analyze whether the responses of speculative trading activities in GBP, CAD and JPY futures markets have changed over the time period from 1994 to 2013.

3 Time-Varying Responses to a Shock to CHF Speculative Activity

The impulse response functions presented in Fig. 1 show positive initial responses of speculative activities in the GBP, CAD and JPY futures markets to a positive shock to speculative activity in the CHF future market. These results hold true for total, long and short speculation. In addition, the durations of the impacts of the shocks appear to be similar for the different pairs of speculative activity. The responses to the shocks tend to die out after about three to four time periods (i.e. three to four weeks).

The Bayesian time-varying VAR analysis shows that this behavior does not seem to change substantially over time. The impulse response functions show little evidence of substantial variation over the time period from 1994 to 2013, especially regarding the initial responses to the shocks. These initial responses are positive for all impulse response functions over the entire time horizon, except for the response of CAD long speculation to CHF long speculation between 1994 and the end of the 1990s. The response of CAD long speculation changes from initially negative in the 1990s to positive in the 2000s.

In addition, some impulse response functions show smaller changes over time. The responses of CAD and JPY short speculation to a positive shock to CHF short speculation are stronger and last longer in the second half of the sample, from about 2005 to 2013. This might indicate that the interrelations between speculators' trading activities in these markets have intensified to some extent. Overall however, these

[11] For more information on the algorithm employed in this study, see Bianchi et al. (2009), Carriero et al. (2013), Mumtaz and Surico (2012).

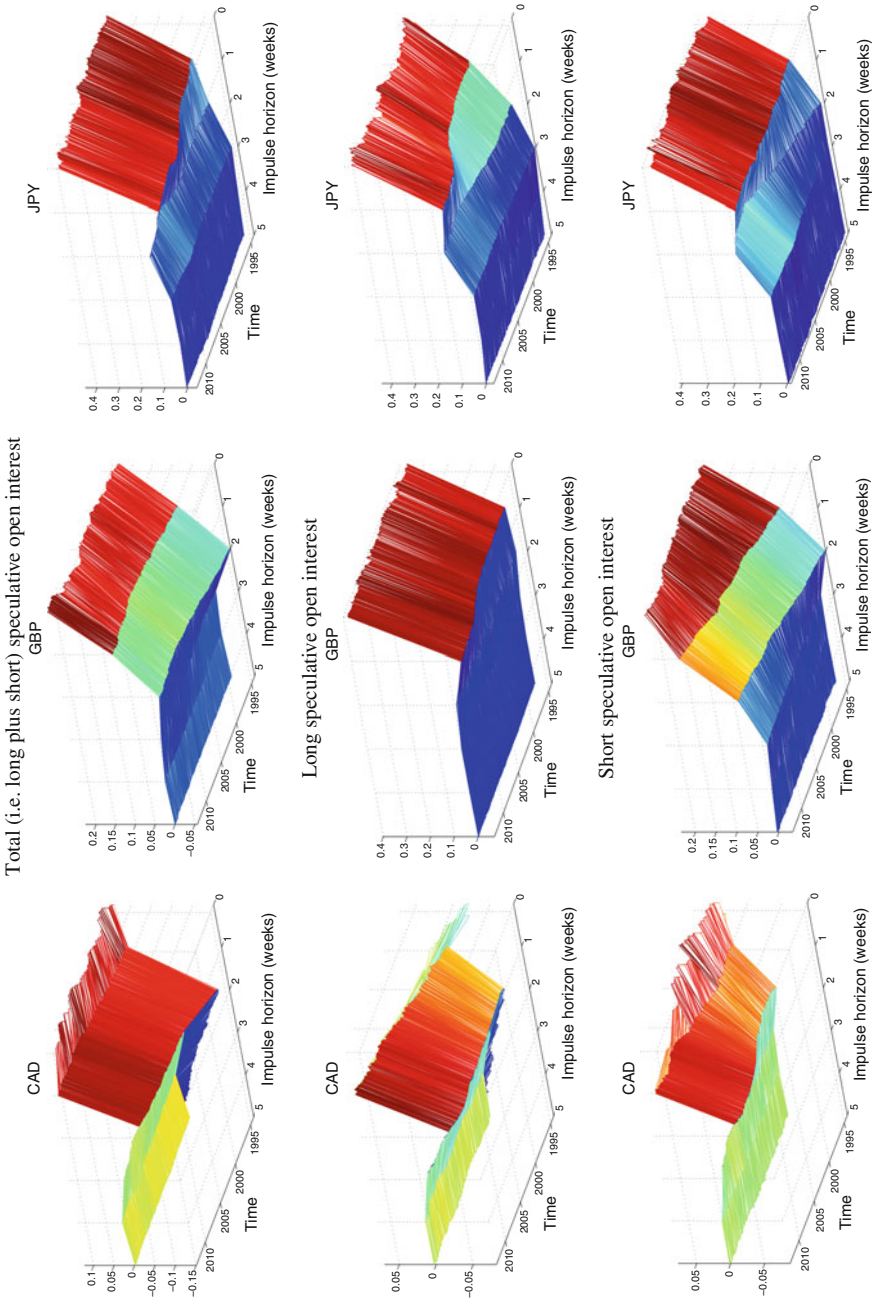

Fig. 1 Impulse response functions for total (i.e. long plus short), long and short speculative open interest: Effects of shock to CHF open interest on CAD, GBP and JPY open interest

findings suggest that the linkages between speculative activities in these currency futures markets are relatively stable over time.

4 Conclusion

This study investigates time variation in cross-market speculative activities in currency futures markets from 1994 to 2013. Using a Bayesian VAR model with time-varying parameters, the results point to positive relations between speculative activities in the CHF futures market and speculative activities in the GBP, CAD and JPY futures markets over the entire time period. The investigation was carried out for total (i.e. long plus short), long and short speculative open interest. A positive shock to total speculative activity in the CHF futures market leads to an increase of total speculative activity in the other three futures markets. In the same way, positive shocks to long and short speculative activities lead to increases in long and short speculative activities in the other futures markets. The findings point to correlated trades between speculators across different currency futures markets and, therefore, to cross-market herding activities of speculators.

Regarding time variation, the findings suggest that the cross-market interrelations between speculative activities in these markets are relatively stable over the entire time period from 1994 to 2013. In only one case (CAD long) does the sign of the initial response to the shock change from negative in the mid-1990s to positive after the year 2000. In all the other cases investigated, the signs of the initial impact and the durations of the impacts, which for all pairs of speculative activities are around three to four weeks, do not change substantially over time. Changes in regulation, new market participants and trading strategies do not seem to have had a significant and lasting impact on cross-market speculative activities in these futures markets, at least regarding the impact of speculative activities in the CHF futures market on speculative activities in the CAD, GBP and JPY futures markets. In a way, this result might be regarded as a sign of stability and maturity of these futures markets, and as a sign of their resilience to economic and financial crises, to changes in composition and behavior of market participants or to changes in regulation.

References

Adrangi, B., & Chatrath, A. (1998). Futures commitments and exchange rate volatility. *Journal of Business Finance and Accounting*, *25*, 501–520.

Bianchi, F., Mumtaz, H., & Surico, P. (2009). The great moderation of the term structure of UK interest rates. *Journal of Monetary Economics*, *56*, 856–871.

Bryant, H., Bessler, D., & Haigh, M. (2006). Causality in futures markets. *Journal of Futures Markets*, *26*, 1039–1057.

Carriero, A., Clark, T., & Marcellino, M. (2013). Bayesian VARs: Specification choices and forecast accuracy. Journal of Applied Econometrics. http://onlinelibrary.wiley.com/doi/10.1002/jae.2315/pdf

Carter, C., & Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, *81*, 541–553.

Chatrath, A., & Song, F. (1999). Futures commitments and commodity price jumps. *Financial Review*, *34*, 95–112.

Chatrath, A., Song, F., & Adrangi, B. (2003). Futures trading activity and stock price volatility: Some extensions. *Applied Financial Economics*, *13*, 655–664.

Ederington, L., & Lee, J. (2002). Who trades futures and how: Evidence from the heating oil futures market. *Journal of Business*, *75*, 353–373.

Fleming, J., Kirby, C., & Ostdiek, B. (1998). Information and volatility linkages in the stock, bond, and money markets. *Journal of Financial Economics*, *49*, 111–137.

Hartzmark, M. (1987). Returns to individual traders of futures: Aggregate results. *Journal of Political Economy*, *95*, 1292–1306.

Houthakker, H. (1957). Can speculators forecast prices? *Review of Economics and Statistics*, *39*, 143–151.

Manzan, S., & Westerhoff, F. (2007). Heterogeneous expectations, exchange rate dynamics and predictability. *Journal of Economic Behavior and Organization*, *64*, 111–128.

Mumtaz, H., & Sunder-Plassmann, L. (2013). Time-varying dynamics of the real exchange rate: An empirical analysis. *Journal of Applied Econometrics*, *28*, 498–525.

Mumtaz, H., & Surico, P. (2012). Evolving international inflation dynamics: World and country-specific factors. *Journal of the European Economic Association*, *10*, 716–734.

Reitz, S., & Westerhoff, F. (2007). Commodity price cycles and heterogeneous speculators: A STAR-GARCH model. *Empirical Economics*, *33*, 231–244.

Röthig, A. (2011). On speculators and hedgers in currency futures markets: Who leads whom? *International Journal of Finance and Economics*, *16*, 63–69.

Röthig, A. (2012). Cross-speculation in currency futures markets. *International Journal of Finance and Economics*, *17*, 272–278.

Röthig, A., & Chiarella, C. (2007). Investigating nonlinear speculation in cattle, corn, and hog futures markets using logistic smooth transition regression models. *Journal of Futures Markets*, *27*, 719–737.

Röthig, A., & Chiarella, C. (2011). Small traders in currency futures markets. *Journal of Futures Markets*, *31*, 898–913.

Sanders, D., Boris, K., & Manfredo, M. (2004). Hedgers, funds, and small speculators in the energy futures markets: An analysis of the CFTC's Commitments of Traders reports. *Energy Economics*, *26*, 425–445.

Sanders, D., Irwin, S., & Merrin, R. (2007). *Smart money? The forecasting ability of CFTC large traders. Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*. Chicago.

Schwarz, K. (2012). Are speculators informed? *Journal of Futures Markets*, *32*, 1–23.

Stein, J., & Hong, B. (1990). Price volatility and speculation. *Journal of Accounting, Auditing and Finance*, *5*, 277–300.

Tornell, A., & Yuan, C. (2012). Speculation and hedging in the currency futures markets: Are they informative to the spot exchange rate. *Journal of Futures Markets*, *32*, 122–151.

Treepongkaruna, S., & Gray, S. (2009). Information and volatility links in the foreign exchange market. *Accounting and Finance*, *49*, 385–405.

Wang, C. (2002). The effect of net positions by type of trader on volatility in foreign currency futures markets. *Journal of Futures Markets*, *22*, 427–450.

Wang, C. (2003). The behavior and performance of major types of futures traders. *Journal of Futures Markets*, *23*, 1–31.

Wang, C. (2004). Futures trading activity and predictable foreign exchange market movements. *Journal of Banking and Finance*, *28*, 1023–1041.

Westerhoff, F. (2008). The use of agent-based financial market models to test the effectiveness of regulatory policies. *Journal of Economics and Statistics*, *228*, 195–227.
Working, H. (1953). Futures trading and hedging. *American Economic Review*, *43*, 314–343.
Working, H. (1962). New concepts concerning futures markets and prices. *American Economic Review*, *52*, 431–459.

Computational Issues in the Stochastic Discount Factor Framework for Equity Risk Premium

Ramaprasad Bhar and A. G. Malliaris

1 Introduction

Individual stock returns at daily frequency are very noisy. When daily returns are aggregated in an index, such as the S&P 500 Index, we would expect such noisiness to persist and even become much more complex. If noisiness is present in the daily returns of an aggregate stock index, it is also there in numerous financial and economic variables either in their levels such as interest rates, volatility indexes, inflation rates, transactions volumes, or their returns as in the case of investing in foreign currencies or commodities.

Can we, however, argue that at this frequency returns of the S&P 500 Index contain no signals? Independent of our inability to decompose daily returns into the sum of a signal and noise component, empirical evidence that stock returns are pro-cyclical suggests that daily returns contain a signal component. The pro-cyclicality of returns does not coincide precisely with the performance of the real economy since returns are driven by future expectations and may anticipate turning points, often by several months in advance. However, the overlap of good times in the economy with higher than average returns as well as periods of bad times with below average returns is sufficiently long to offset possible decoupling of returns and the state of the macro-economy.

We are now ready to describe our goal of this research article: Numerous studies have deliberated the determinants of excess equity returns also called the equity premium defined as the difference between the S&P 500 Index and the risk free return

R. Bhar (✉)
School of Risk and Actuarial Studies, The University of New South Wales,
Sydney, NSW, Australia
e-mail: R.Bhar@unsw.edu.au

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

R. Dieci et al. (eds.), *Nonlinear Economic Dynamics and Financial Modelling*,
DOI: 10.1007/978-3-319-07470-2_14,

of a 3-month Treasury Bill. We are interested in explaining equity premium at daily frequency by studying drivers at the daily dimension. There are both fundamental and behavioral variables that drive the daily equity premium. These include the degree of uncertainty faced by investors and captured by the daily CBOE VIX and the spread between the daily short-term 3-month T-Bill interest rate and the long-term 10-year Note interest rate. This spread is usually called the yield curve and captures among other variables, Central bank policy, inflationary expectations and long-run growth prospects. To the VIX and yield curve we also add a measure of trading liquidity that became important during the current global financial crisis. This measure is a spread between the no risk Treasury Bill rate (T-Bill) and the low risk Eurodollar rate. Another similar measure of liquidity is the spread between the T-Bill and the LIBOR. To these three macroeconomic variables of daily frequency we add a behavioral variable to be defined below and called momentum. We will use these factors in the SDF framework to identify significant determinants of daily equity premium.

This article concentrates on setting up the SDF model using the factors outlined above and solves several implementation issues particularly with respect to the MATLAB software environment. We leave the actual implementation to other empirical researchers.

2 Bibliographical Review

Financial economists appear to agree that individual stocks are driven by news about their cash flows. When we aggregate among numerous stocks that are included in a certain index, such as the S&P 500 Index, this index moves by the news of cash flows of the individual stock according to their weights. However, there are several other factors that influence the aggregate index such as macroeconomic news about GDP growth, inflation, Fed policies and several others. Yan (2010) explores theoretically the behavior of individual and aggregate stock prices without any reference to the size of the trading interval. He develops a dynamic general equilibrium model with incomplete information and derives several important conclusions. In particular he finds that the correlation between stock returns and earnings is, on average, positive at the individual stock level but lower or even negative at the aggregate level. The daily frequency approach of our paper does not allow us to include either quarterly earnings or a monthly dividend yield which are the primary fundamental variables for equities so we consider daily macroeconomic variables such as interest rates and uncertainty.

The influence of interest rates, both short-term as Fed funds or T-Bills and long-term 10-year Treasury Notes on stock market returns has been investigated extensively. The relationships are complex and there are no definitive answers. For example, Fed funds are almost exclusively determined by the Fed to reflect its decisions on monetary policy. As it is well known Fed policy is driven by its dual mandate to promote price stability and economic growth and thus Fed funds are determined

by Taylor rules reflecting deviations of inflation and output from desirable targets. However, the longer-run 10-year T-Notes rates are determined by market conditions and incorporate both assessments of future inflation and real growth. Fed policies such as the three rounds of quantitative easing since the spring of 2009, may impact the 10-year T-Notes also.

The yield curve defined as the difference between the T-Note rate and T-Bill rate is just one measurement across time and gives an estimate of inflationary expectations that may be driven either by robust prospects of economic growth or inflation fears generated by an easy monetary policy. These topics have been investigated recently by Chen (2008), Craine and Martin (2003), Bernanke and Kuttner (2005), Caporale and Caporale (2003), Rigobon and Sack (2003) and others.

Equity risk and volatility as a topic of research has also attracted a great amount of interest. Standard CAPM models suggest that the risk premium is positively related to market volatility. This theoretical principle is often confirmed but also rejected in empirical studies which have motivated further theoretical hypotheses. One hypothesis emphasizes the leverage effect that says that a drop in an equity price reduces net worth and increases the debt to equity ratio, making the stock riskier that in turn causes the equity risk premium to increase. A second hypothesis called the volatility feedback effect argues that positive volatility shock increases the future required return on equity and therefore stock prices are expected to fall. Campbell and Hentschel (1992) find some evidence of a positive relation between the conditional volatility and stock returns, while Low and Zhang (2005) and Zhao (2008) found a negative correlation. Schwert (1989) finds that the leverage effects are more pronounced during financial recessions while Bekaert and Wu (2000) find volatility feedback effects dominate leverage effects.

Behavioral finance has offered valuable explanations on several asset pricing puzzles and has earned an increasing rate of academic acceptance. For the purpose of our hypotheses formulation we need to go beyond daily frequency macroeconomic variables and also consider the role of certain behavioral variables. In a macroeconomic environment of rapid real economic growth with high employment and low inflation causing real interest rates to be also low, stock price increases may signal further increases, thus generating positive feedback. As investors observe positive fundamentals that persist, they gradually become overconfident that the recently favourable macroeconomic regime will continue. Thus they extrapolate the limited recent sample of favourable conditions into the distant future. This motivates them to buy more and increases in return attract more buyers, thus generating a positive momentum. During periods of asset booms, momentum builds up slowly over several quarters. However, when sufficiently negative news occurs, reversals are most often faster with sharper declines.

Prior to the emergence of behavioral finance, positive deviations of asset returns from levels justified by fundamentals would produce arbitrage opportunities. Arbitrageurs acting rationally would sell short the overvalued asset and contribute to eliminating deviations away from fundamentals. Behavioral finance authors have produced numerous papers explaining the limits to arbitrage. For example Shleifer and Vishny (1997) have argued that arbitrageurs face both fundamental risk as well as

noise trader risk and these risks may discourage them from taking a position. What if, future fundamentals improve from being good to becoming even better? What if noise traders being overconfident continue to drive returns even higher? Shleifer and Vishny argue convincingly that under certain reasonable conditions prices may deviate from fundamentals for some time.

Chordia and Shivakumar (2002) consider the continuation of short-term returns called momentum. They ask the question: are momentum returns due to investor irrationality or can they be explained rationally. They show that certain lagged macroeconomic variables can explain future stock returns and payoffs from momentum strategies disappear once stock returns are adjusted for their predictability.

A number of recent models show that both momentum and contrarian investor behaviour may arise, be sustained in a financial market and may also be profitable depending upon the horizon of the strategy. Barberis et al. (1998), Daniel et al. (1998) and Hong and Stein (1999) each develop models of investor behaviour that show how common psychological heuristics, if used by market participants, may lead to both mean reverting and persistent patterns in asset prices. Goetzmann and Massa (2002) carefully study both positive feedback (momentum) and negative feedback (contrarian) behaviour for a large sample of individual investors with daily trading activities.

Momentum trading is not without its share of efficient-markets-based explanations. Conard and Kaul (1998) and Berk et al. (1999) have argued that stocks with high (low) realized returns will be those that have high (low) expected returns, suggesting that momentum strategy's profitability is a result of cross-sectional variability in expected returns due to macroeconomic variables such as dividend yields, default spread, three month T-Bills and term structure spread.

This limited and selective bibliographical review supported by the integrative and perceptive survey of Barberis and Thaler (2005) suggests that momentum in the aggregate market in the form of temporary persistence of above average returns during periods of boom and its eventual reversal during asset busts can be an important behavioral variable to be included in the formulation of our hypotheses. Although the concept of momentum in behavioral finance often describes continuation of short-term returns for individual stocks, the idea is used here broadly to describe continuation of short-term returns for the aggregate market.

3 Equity Risk Premium in Stochastic Discount Factor Framework

In this section we briefly outline the basic concepts in the SDF framework. This will help us to establish the econometric methodology for testing the influence of the selected observable factors determining the equity premium observed at daily frequency.

The SDF approach posits a very simple notion in asset pricing. It proposes that the price of an asset at time t is the expected discounted value of the payoff from the

asset in period $t+1$. Discount factor is, therefore, stochastic and that encapsulates all the uncertainties. One advantage in this approach is that we do not require the knowledge of investors' preferences. Various aspects of this framework have been analyzed in detail in Smith and Wickens (2002). In addition, the celebrated book by Cochrane (2005) relies solely on this methodology. In empirical asset pricing models the researchers select the SDF as a function of observable factors as well as of model parameters.

In order to formalize the concepts, we define, as the price of an asset at time t, X_{t+1} is the payoff from the asset at time $t+1$, M_{t+1} is the discount factor for the period $t+1$, E_t is the expectation operation at time t consistent with all available information at that time. In general both X_{t+1} and M_{t+1} would be stochastic. The SDF framework suggests the following pricing relation:

$$P_t = E_t[M_{t+1} \times X_{t+1}], 0 \leq M_{t+1} \leq 1. \tag{1}$$

It is customary to express the above relationship in term of gross asset return R_{t+1} as,

$$1 = E_t[M_{t+1} \times \frac{X_{t+1}}{P_t}] = E_t[M_{t+1} \times R_{t+1}]. \tag{2}$$

From standard statistical properties Eq. (2) may be expanded as:

$$E_t[M_{t+1}R_{t+1}] = E_t[M_{t+1}] \times E_t[R_{t+1}] + \sigma_{t,M_{t+1},R_{t+1}}, \tag{3}$$

where, the symbol $\sigma_{t,x,y}$ is used to denote covariance at time t between the two variables x and y. Combining the Eqs. (2) and (3) we have,

$$E_t[R_{t+1}] = \frac{1 - \sigma_{t,M_{t+1},R_{t+1}}}{E_t[M_{t+1}]}. \tag{4}$$

Thus, Eq. (4) relates the covariance of the error terms in the joint process of R_{t+1} and M_{t+1} with the expected values. If the payoff at $t+1$ is known with certainty, then the return becomes the risk-free rate, ($R_{t+1} \equiv 1 + r_{f,t}$), and the covariance term becomes zero. If we assume that the payoff is \$1 then Eq. (4) becomes,

$$1 = (1 + r_{f,t})E_t[M_{t+1}]. \tag{5}$$

Combining Eqs. (4) and (5) and denoting gross return R_{t+1} as $(1 + r_{t+1})$ yields,

$$E_t[r_{t+1}] - r_{f,t} = -(1 + r_{f,t})\sigma_{t,M_{t+1},(r_{t+1}-r_{f,t})}. \tag{6}$$

This is referred to as the no-arbitrage relation for all correctly priced securities (see Smith and Wickens (2002) page 400 for details). Since the right hand side in Eq. (6) is the risk premium, and we expect it to be positive, the covariance term must, therefore, be non-positive.

It is customary in SDF analysis to assume that the joint distribution of the gross return and the discount factor is log-normal. This has come about due to some empirical justification as well as the fact that the discount factor requires positive support. With this assumption, we let $m_{t+1} = \ln M_{t+1}$ and $r'_{t+1} = \ln R_{t+1}$, and by applying Jensen's inequality, we can express Eqs. (6) as, (see Smith and Wickens (2002) page 402 for the steps),

$$E_t[r'_{t+1} - r_{f,t}] + 0.5\sigma^2_{t,r_{t+1}} = -\sigma_{t,m_{t+1},r'_{t+1}}. \tag{7}$$

Apart from the term representing time varying Jensen effect, Eq. (7) relates the expected excess return or premium in terms of the covariance of the return with the SDF.

Equation (7) is quite general and suggests that the risk premium is generated by the asset specific covariance with the SDF. Cochrane (2005) points out that this analysis can be used for both linear and nonlinear asset pricing models. The choice of the SDF is important for different asset pricing applications and its generality is attributed to the property that only the covariance is required. This also forces the researchers to look for the correct specification of $m_{t+1} = \ln M_{t+1}$. It is quite common in the SDF framework for the researcher to specify an affine structure to the SDF in terms of relevant observable variables. It is also clear that since risk is measured by the covariance, the choice of variables and also observation frequency should be such that there is enough time variation in the covariance structures. This implies that the variables should exhibit conditional heteroscedasticity.

In this paper we use daily observations that exhibit high conditional heteroscedasticity and we employ two observable variables to define the SDF so we can study their impact on excess equity return. In this context, we express, (see also Kizys and Spencer (2007)),

$$-m_{t+1} = \beta_1 x_1 + \beta_2 x_2, \tag{8}$$

where, x_1 and x_2 are the two observable variables determining the SDF. Equation (7) thus becomes,

$$E_t[r'_{t+1} - r_{f,t}] + 0.5\sigma^2_{t,r'_{t+1}} = \beta_1 \sigma_{t,x_{1,t+1},r'_{t+1}} + \beta_2 \sigma_{t,x_{2,t+1},r'_{t+1}}. \tag{9}$$

With the above specification the sign of the parameters β_1 and β_2 would determine how these observable variables affect the risk premium.

4 Observable Determinants for SDF

The SDF method has been employed in studying the effect of macroeconomic variables on equity premium. Kizys and Spencer (2007) discuss several research articles in this context and the inflation and output growth variables feature prominently as the choice of observable variables determining the SDF.

The use of macroeconomic variables implies that the study is conducted with monthly or quarterly observable frequency. In this study we are concerned in analyzing the contribution to the equity risk premium at the daily frequency. This restricts our choice of candidate observable variables for the SDF. Pena and Rodriguez (2006) demonstrate the importance and the connection between the term spread and the stock prices. This guides us to choose one of the observable variables to be the term spread, measured by the difference between the yields on the 10-Year Treasury Notes and T-Bill. We label this variable as s_t.

It also makes sense to investigate the power of output gap to explain equity premium since output gap is a prime business cycle indicator and does not depend on the level of market prices. This has been done by Cooper and Priestley (2009). They find that output gap has predictive power for equity premium for all the G7 countries. Obviously, then the empirical investigation needs to be carried out in frequency compatible with business cycle. Cooper and Priestley work mainly in the linear regression framework. We are not aware of such an investigation in the SDF framework. We leave this for a future study.

With respect to our second observable variable determining the SDF we take cue from the literature in behavioral finance. Shefrin (2008) provides an excellent overview of the issues that arise in dealing with behavioral aspects of constructing SDF. Investor's behavioral aspects enter into the SDF framework via sentiment which relates to a belief system. This should affect the way expectation is taken, e.g. with reference to Eq. (9). If investor's subjective belief matches with the objective reality then the prices are said to reflect zero sentiment. Otherwise there will be non-zero sentiment. Most studies in this area have focused on explaining the equity premium puzzle by incorporating preference parameters in asset pricing model, as in Abel (2002). Shefrin (2005) shows that historically investors have been predominantly pessimistic, and that their pessimism was time varying as well.

In this paper, however, our objective is to choose another observable variable for the SDF that has some behavioral aspect. We focus on the concept of momentum return in the equity market. This simply states that the return performance in the recent past would continue for some time in the future. We define a proxy for the behavioral variable to represent recent performance or momentum return. We follow the definition used in Koijen et al. (2008). In their continuous time set up for strategic asset allocation problem, they allow the short-term performance of the equity market as a weighted function of the past returns. Given that S_t represents the index level at time t, then the momentum return g_t is given by:

$$g_t = \int_0^t e^{-(t-u)} \frac{\mathrm{d}S_u}{S_u} \tag{10}$$

where $e^{-(t-u)}$ is the weighting scheme. Koijen et al. (2008) show that there is no need to consider any more general weighting scheme since this simple approach is capable of matching the short-term and long-term autocorrelations of stock returns.

Since we are working in discrete time setting, the approximate discrete performance variable corresponding to Eq. (10) is given below:

$$g_t \approx \sum_{i=1}^{t} e^{-i} r_{t-i+1}. \tag{11}$$

This completes our choice of observable variables as required by Eq. (8). In the next section we show the empirical structures needed to implement Eq. (9) together with the SDF specification given by Eq. (8).

5 Empirical Setup and Econometric Issues

As we are dealing with three observable series, that is, the equity risk premium, the term spread and the performance variable representing momentum return, it is straightforward to set up a Vector Auto-Regressive (VAR) framework to generate conditional expectation. This also helps to incorporate feedback from conditional variance and conditional covariance as needed by Eq. (9). The VAR coefficient matrix may be constrained to impose the no-arbitrage condition suggested by the SDF based pricing relation given in Eq. (6).

Another important aspect of the VAR specification is the conditional covariance structure. The associated numerical computation of the likelihood function requires the guarantee of positive definiteness of the covariance while the elements of this matrix are unknown parameters. At the same time, a well documented empirical finding in the finance literature is the asymmetric impact of news on the volatility transmission (see Bae and Karolyi (1994); Koutmos and Booth (1995) and Booth et al. (1997)). The asymmetric phenomenon in combination with the observed volatility clustering in equity market returns validates the use of an EGARCH framework. The EGARCH model, as developed by Nelson (1991), captures the potential asymmetric behavior of equity market returns and avoids imposing non-negativity constraints in GARCH modelling—by specifying the logarithm of the variance. In a univariate problem it is no longer necessary to restrict parameters in order to avoid negative variances. We, therefore, need to be mindful of the efficacy of EGARCH framework and the need for ensuring positive definiteness of the covariance matrix in our problem.

Following the suggestions in Kizys and Spencer (2007), we define the VAR model that captures the essence of our SDF based model as follows:

$$\underbrace{\begin{bmatrix} r_t - r_{f,t} \\ s_t \\ g_t \end{bmatrix}}_{Y_t} = \underbrace{\begin{bmatrix} 0 \\ a_{2,1} \\ a_{3,1} \end{bmatrix}}_{A} + \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ b_{2,1} & b_{2,2} & b_{2,3} \\ b_{3,1} & b_{3,2} & b_{3,3} \end{bmatrix}}_{B} Y_{t-1} + \underbrace{\begin{bmatrix} \gamma_{1,1} & \gamma_{1,2} & \gamma_{1,3} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{B} \Omega_t \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \varepsilon_t \tag{12}$$

$$\varepsilon_t \sim N(0, \Omega_t). \tag{13}$$

By constraining the top element of the parameter vector, and the first row of the parameter matrix we are implementing the no-arbitrage condition of the SDF pricing relation, and the fact that the excess return or risk premium in the equity market is not predictable by any of the lagged variables. The first row of the parameter matrix captures the time varying Jensen effect as well as the time varying covariance effects. The other two rows of this matrix are constrained to be zeros.

We next consider the full specification of the conditional variance as it ultimately determines the time varying covariance that is essential for the SDF framework to identify the influence of the conditioning variables on the equity premium. Following Tsay (2002), we define the covariance matrix Ω_t in Eq. (9) as:

$$\Omega_t = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ p_{2,1} & 1 & 0 \\ p_{3,1} & p_{3,2} & 1 \end{bmatrix}}_{P} \underbrace{\begin{bmatrix} q_{1,1,t} & 0 & 0 \\ 0 & q_{2,2,t} & 0 \\ 0 & 0 & q_{3,3,t} \end{bmatrix}}_{Q} P' \tag{14}$$

The attractive feature of this triangular decomposition is that we only need to ensure only the diagonal elements of the matrix Q are positive for all time periods, and the elements of the matrix P are unconstrained. This automatically ensures the complete covariance matrix is positive definite. The full expanded form is given below:

$$\Omega_t = \begin{bmatrix} q_{1,1,t} & p_{2,1}q_{1,1,t} & p_{3,1}q_{1,1,t} \\ p_{2,1}q_{1,1,t} & p_{2,1}^2 q_{1,1,t} + q_{2,2,t} & p_{2,1}p_{3,1}q_{1,1,t} + p_{3,2}q_{2,2,t} \\ p_{3,1}q_{1,1,t} & p_{2,1}p_{3,1}q_{1,1,t} + p_{3,2}q_{2,2,t} & p_{3,1}^2 q_{1,1,t} + p_{3,2}^2 q_{2,2,t} + q_{3,3,t} \end{bmatrix}. \tag{15}$$

In addition to the above, the elements of the matrix P may be given economic interpretation depending on the problem at hand, as we will see later.

To complete the specification we allow the diagonal elements of the $Q, q_{j,j}, j = 1, 2, 3$ to have the exponential GARCH form as:

$$\ln q_{j,j,t} = \phi_{j,0} + \phi_{j,1} \ln(q_{j,j,t-1}) + \phi_{j,2}\hat{\varepsilon}_{j,t-1} + \phi_{j,3}\left(\mid \hat{\varepsilon}_{j,t-1} \mid -\sqrt{\frac{2}{\pi}} \right), \tag{16}$$

where $\hat{\varepsilon}_{j,t-1} = \frac{\varepsilon_{j,t-1}}{\sqrt{q_{j,j,t-1}}}$, the standardized innovation for the jth element.

This EGARCH specification follows from Nelson (1991) and has been found to be superior to a number of stochastic volatility model structures as studied by Chernov et al. (2003). In the structural conditional variance Eq. (16), the parameter $\phi_{j,0}$ represents the logarithm of the unconditional variance of the jth process and is assumed constant. The parameter $\phi_{j,1}$ determines the influence of the past conditional volatility on the current conditional volatility. For the conditional volatility process to be stationary it is required that, $\mid \phi_{j,1} \mid < 1$. The persistence of volatility may also be quantified by examining the half-life, which indicates the time period required for the shocks to reduce to one-half their original size. Defined as:

$$\text{HL} = \frac{\ln(0.5)}{\ln \mid \phi_{j,1} \mid}. \tag{17}$$

The leverage effect is captured in two parts by the parameters, $\phi_{j,2}$ (sign effect) and $\phi_{j,3}$ (size effect). If $\phi_{j,2}$ is negative then a negative realisation of standardised innovation will increase volatility by more than a positive realisation of equal magnitude. Similarly, if the past absolute value of standardised innovation is greater than its expected value then the current volatility will rise. However, the resultant impact of the size effect will be determined by the sign of the parameter $\phi_{j,3}$. The asymmetric effect (Relative Asymmetry) of standardised innovations on volatility may be measured by:

$$\text{RA} = \frac{\mid -1 + \phi_{j,2} \mid}{(1 + \phi_{j,2})}. \tag{18}$$

This quantity is greater than, equal to or less than one for negative asymmetry, symmetry and positive asymmetry, respectively.

The conditional variance specification in Eq. (16) is quite flexible. Depending on the problem being investigated, it allows us to incorporate additional exogenous variables that may be influencing the time variation of the variance. For example, if the conditioning variable in the SDF is related to business cycle e.g. output gap then lagged inflation may be thought of as an explanatory variable in Eq. (16) determining the conditional variance. This idea originates from the studies analysing impact of inflation on real stock return as part of the variability hypothesis (see for example Buono (1989)).

The variables selected for this exposition may be changed to test other candidates based on different economic insight. The set up above is general enough to accommodate such variations.

In the econometric set up discussed in this section, we do model time varying correlations between the residuals of the VAR variables explicitly. In a multivariate set up it is probably straightforward to include constant correlations. This also reduces the computational complexity. However, once the triangular decomposition we have used in defining the conditional covariance is estimated, we are able to infer time variation of correlation. For example, with reference to Eq. (15), the conditional correlation between excess return and term spread is given by:

$$\rho_{t,\text{Excess } Re\ t,\text{TermSpread}} = \frac{\Omega_t[1,2]}{\sqrt{\Omega_t[1,1]\Omega_t[2,2]}}. \tag{19}$$

In the same fashion the other conditional correlations may be inferred. For a sketch of the proof that it lies between -1 and $+1$, (see Kizys and Spencer (2007)). Once the model is estimated for a particular data set, the time variation of these conditional correlations could be further analyzed for their economic significance.

From the estimated parameters of the model set up above, we can infer the equity premium via the following equation:

$$rp_t = \gamma_{1,2} \times \Omega_t[2,1] + \gamma_{1,2} \times \Omega_t[3,1]. \tag{20}$$

The approach outlined in this section delivers a multifactor model where the equity premium is explained by lagged values of conditioning variables in the SDF set up. This should allow empirical researchers to develop this further. In the next section we will outline an alternative way to approach the empirical set up of the SDF framework.

6 Alternative Empirical Setup for SDF

The joint distribution of gross return and the SDF as shown in Eq. (2) is the basic condition that needs to be correctly dealt with for any empirical testing based on the SDF framework. Under the assumption joint log-normal distribution we have linearized the expression by invoking Jensen's inequality condition in term of covariance between the SDF and the gross return. We use this as the starting point and in Sect. 5 show the eventual econometric realization of the model in term of observable determinants of the SDF.

In this section, we demonstrate an alternative approach that models the joint distribution of the gross return and the SDF via copula. A copula is a distribution function with known marginals. Copulas provide a more detailed description of the dependency structure between two random variables, since they represent bivariate functions that link marginal probability distributions (and density) functions of the random variables to their joint probability distributions (and density) functions See for example Durrani and Zeng (2007). Copulas, therefore, offer interesting insights into the dependence structures between the distributions of random variables. Depending on the available economic understanding, different copula function may be employed.

However, in this article we are going to focus on log-normal copulas to be consistent with the earlier discussions. In order to keep the main part of the paper straightforward for readers we place the mathematical details in the appendix following Liu (2010). For empirical testing the researchers need to establish the correspondence between the variables used in the appendix and the SDF set up. The functions involved in the analytical structures in the appendix are easily implementable in software like MATLAB.

7 Concluding Remarks

In this methodological essay on the SDF approach to asset pricing we have outlined the main ideas with reference to the econometric modeling of the equity risk premium. In particular, we have shown how economically meaningful observable determinants could be employed to generate the discount factor. This requires the standard assumption of joint log-normal distribution of the discount factor and the gross return. The econometric realization of the model and the related implementation

issues are also discussed within the standard framework of EGARCH specification. This is necessary to build the time variation of the covariance between the discount factor and the equity premium.

In addition to this, we have contributed to the literature by introducing the copula based dependence modeling. Although we have only analyzed bivariate log-normal copula, the framework paves the way for empirical testing of other copula functions that may be more appropriate in other situations. This copula based approach requires computing double integrals, but we have shown how this may be simplified in the log-normal case to single integrals. These are easily implemented in MATLAB with the built-in functions available.

Appendix A: Bivariate Log-Normal Copula

The intrinsic relations between bivariate distributions and their marginal distributions can be clearly characterised by copulas. The bivariate log-normal distributions play important roles in areas other than being described here. In this appendix, log-normal copula is derived. These formulas are in terms of the Gaussian Q-function, being supported by MATLAB. Thus the copula evaluation process can be expedited both analytically and numerically.

The bivariate log-normal probability density function for a pair of random variables X and Y, is given by (A.3), where (x, y) are transformed from bivariate normal distribution variables (x', y') as:

$$x = A\exp(mx'), \tag{A.1}$$

$$y = B\exp(ny'). \tag{A.2}$$

With ρ as correlation between (x', y') and $(x > 0, y > 0, A > 0, B > 0)$:

$$f_{x,y}(x, y) = \frac{1}{2mn\pi\sigma_X\sigma_Y xy\sqrt{1-\rho^2}} \times \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\left(\frac{\ln(x/A)-m\mu_X}{m\sigma_X}\right)^2 - 2\rho\left(\frac{\ln(x/A)-m\mu_X}{m\sigma_X}\right)\left(\frac{\ln(y/B)-n\mu_Y}{n\sigma_Y}\right) + \left(\frac{\ln(y/B)-n\mu_Y}{n\sigma_Y}\right)^2\right]\right\} \tag{A.3}$$

Referring to (A.1) and (A.2), x and y are log-normal variables. With respect to the SDF formulation, we may consider x as the SDF and y as the gross return as discussed in Sect. 3. In this formulation, x' may be represented as a function of observable determinants and similarly for y'. The additional parameters (A, m, B, n) in (A.1) and (A.2) are used to make these transformations as general as possible.

The marginal PDFs associated with (A.3) take the following forms:

$$f_X(x) = \frac{1}{mx\sigma_X\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln(x/A) - m\mu_X}{m\sigma_X}\right)^2\right], \tag{A.4}$$

$$f_Y(y) = \frac{1}{my\sigma_Y\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln(y/B) - n\mu_Y}{n\sigma_Y}\right)^2\right]. \tag{A.5}$$

The corresponding marginal distributions are:

$$F_X(x) = 1 - Q\left(\frac{\ln(x/A) - m\mu_X}{m\sigma_X}\right), \tag{A.6}$$

$$F_Y(y) = 1 - Q\left(\frac{\ln(y/B) - n\mu_Y}{n\sigma_Y}\right). \tag{A.7}$$

Here, $Q(\cdot)$ is referred to as the Gaussian Q-function with the following definition:

$$Q(z) = \frac{1}{2\pi}\int_z^\infty \exp\left(-\frac{t^2}{2}\right)dt. \tag{A.8}$$

With change of variable, the Gaussian Q-function may be written as (over finite interval):

$$Q(z) = \frac{1}{\pi}\int_z^{\pi/2} \exp\left(-\frac{z^2}{2\sin^2\theta}\right)d\theta. \tag{A.9}$$

here $Q(\cdot)$ is computable using MATLAB built in function. For specific values of $m(=2)$ and $n(=2)$ the elements of the covariance matrix for the pair of random variables X and Y are given below:

$$\text{var}(X) = A^2\exp(4\mu_X)\exp(4\sigma_X^2)[\exp(4\sigma_X^2) - 1], \tag{A.10}$$

$$\text{var}(Y) = B^2\exp(4\mu_Y)\exp(4\sigma_Y^2)[\exp(4\sigma_Y^2) - 1], \tag{A.11}$$

$$\text{cov}(X, Y) = AB\exp(2\mu_X + 2\mu_Y)\exp(2\sigma_X^2 + 2\sigma_Y^2)[\exp(4\rho\sigma_X\sigma_Y) - 1]. \tag{A.12}$$

From (A.6) and (A.7) we can write:

$$x = F_X^{-1}(u) = A\exp[m\mu_X + m\sigma_X Q^{-1}(1-u)] := a_1, \tag{A.13}$$

$$y = F_Y^{-1}(w) = B\exp[n\mu_Y + n\sigma_Y Q^{-1}(1-w)] := a_2, \tag{A.14}$$

where $Q^{-1}(\cdot)$ is the inverse Gaussian Q-function and this is available as a built-in function in MATLAB. Now, we are in a position to write the bivariate log-normal

copula distribution function as follows:

$$C(u, w) = F_{XY}[F_X^{-1}(u), F_Y^{-1}(w)] = F_{XY}(a_1, a_2) = \int_0^{a_1}\int_0^{a_2} f_{X,Y}(x, y)dydx. \quad \text{(A.15)}$$

The simplification of the double integral in (A.15) is at the core issue in empirical implementation. The quantity $f_{X,Y}(x, y)$ is given by Eq. (A.3) and Liu (2010) shows how to convert this to a single integral which can be readily implemented in MATLAB. In this appendix we just quote the final expression for the copula distribution function.

$$C_{u,w} = \frac{1}{\sqrt{2\pi}} \int_{-x}^{Q^{-1}(1-u)} \exp\Big(-\frac{r^2}{2}\Big) Q\Big(\frac{\rho r - Q^{-1}(1-w)}{\sqrt{1-\rho^2}}\Big) dr \quad \text{(A.16)}$$

$$(0 < u < 1, 0 < w < 1).$$

Liu (2010) suggests further avenue to reduce computational burden.

References

Abel, A. B. (2002). An exploration of the effects of pessimism and doubt on asset returns. *Journal of Economic Dynamics and Control*, *26*, 1075–1092.

Bae, K. H., & Karolyi, G. A. (1994). Good news. *Bad News and International Spillovers of Stock Return Volatility Between Japan and the US, Pacific Basin Finance Journal*, *2*, 405–438.

Barberis, N., Shleifer, N., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, *49*, 307–345.

Barberis, N., & Thaler, R. (2005). A survey of behavioral finance. In R. Thaler (Ed.), *Advances in behavioral finance* (Vol. II). New Jersey: Princeton University Press, Princeton.

Bekaert, G., & Wu, G. (2000). Asymmetric volatility and risk in equity markets. *Review of Financial Studies*, *13*, 1–42.

Berk, J. B., Green, R. C., & Naik, V. (1999). Optimal investment. *Growth Options, and Security Returns, The Journal of Finance*, *54*, 1553–1607.

Bernanke, B., & Kuttner, K. (2005). What explains the stock market's reaction to federal reserve policy? *The Journal of Finance*, *60*, 1221–1257.

Booth, G. G., Martikainen, T., & Tse, Y. (1997). Price and volatility spillovers in Scandinavi an stock markets. *Journal of Banking and Finance*, *21*, 811–823.

Buono, M. J. (1989). The relationship between the variability of inflation and stock return: An empirical investigation. *The Journal of Financial Research*, *12*, 329–339.

Campbell, J. Y., & Hentschel, L. (1992). No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, *31*, 281–318.

Caporale, B., & Caporale, T. (2003). Investigating the effects of monetary regime shifts: The case of the federal reserve and the shrinking risk premium. *Economic Letters*, *80*, 87–91.

Chen, S., (2008). Predicting the Bear Stock Market: Macroeconomic Variables as Leading Indicators, National Taiwan University—Department of Economics, Working Paper Series.

Chernov, M., Gallant, R., Ghysels, E., & Tauchen, G. (2003). Alternative models for stock price dynamics. *Journal of Econometrics*, *116*, 225–257.
Chordia, T., & Shivakumar, L. (2002). Momentum, business cycle, and time-varying expected returns. *The Journal of Finance*, *57*(2), 985–1019.
Cochrane, J. H. (2005). *Asset pricing*. USA: Princeton University Press.
Conard, J., & Kaul, G. (1998). An anatomy of trading strategies. *Review of Financial Studies*, *11*, 489–519.
Craine, R. & Martin, V. (2003). Monetary Policy Shocks and Security Market Responses, University of California, Berkeley Economics Working Paper.
Cooper, I., & Priestley, T. (2009). Time-varying risk premia and the output gap. *The Review of Financial Studies*, *22*(7), 2801–2833.
Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under and overreactions. *The Journal of Finance*, *53*, 1839–1885.
Durrani, T. S., & Zeng, X. (2007). Copulas for bivariate probability distributions. *Economics Letters*, *43*(4), 248–249.
Goetzmann, W. N., & Massa, M. (2002). Daily momentum and contrarian behavior of index fund investors. *The Journal of Financial and Quantitative Analysis*, *37*, 375–389.
Hong, H., & Stein, J. C. (1999). A unified theory of underreaction. *Momentum Trading, and Overreaction in Asset Markets, The Journal of Finance*, *54*, 2143–2184.
Kizys, R. & Spencer, P. (2007). Assessing the Relation between Equity Risk Premium and Macroeconomic Volatilities in the UK. Discussion Papers in Economics, No. 13, UK: University of York.
Koutmos, G., & Booth, G. G. (1995). Asymmetric volatility transmission in international stock markets. *Journal of International Money and Finance*, *14*, 747–762.
Koijen, R. S. J., Rodriguez, J. C., & Sbuelz, A. (2008). Momentum Return and Mean-Reversion in Strategic Asset Allocation, Working Paper, Stern School of Business, New York University, May.
Liu, X. (2010). Copulas of bivariate rayleigh and log-normal distributions. *Electronics Letters*, *46*, 1669–1671.
Low, B. S., & Zhang, S. (2005). The volatility risk premium embedded in currency options. *Journal of Financial and Quantitative Analysis*, *40*, 803–832.
Nelson, D. B. (1991). Conditional heteroscedasticity in asset returns: A new approach. *Econometrica*, *59*, 347–370.
Pena, J. I., & Rodriguez, R. (2006). On the economic link between asset prices and real activity. *Journal of Business Finance and Accounting*, 34(5–6), 889–916.
Rigobon, R., & Sack, B. (2003). Measuring the reaction of monetary policy to the stock market. *The Quarterly Journal of Economics*, *118*, 639–669.
Schwert, G. W. (1989). Why does stock market volatility change over time? *Journal of Finance*, *44*, 1115–1153.
Shefrin, H. (2005). *A behavioral approach to asset pricing*. Boston: Elsevier Academic Press.
Shefrin, H. (2008). Risk and return in behavioral SDF-based asset pricing models. *Journal of Investment Management*, *6*(3), 1–18.
Shleifer, A., & Vishny, R. W. (1997). The limits of arbitrage. *The Journal of Finance*, *52*(1), 35–55.
Smith, P., & Wickens, M. (2002). Asset pricing with observable stochastic discount factors. *Journal of Economic Surveys*, *16*(3), 397–446.
Tsay, S. R. (2002). *Analysis of financial time series*. New York: Wiley.
Yan, H. (2010). Is noise trading cancelled out by aggregation? *Management Science*, *56*, 1047–1059.
Zhao, Y. (2008). Equity Risk Premium and Volatility: A Correlation Structure, Dalhousie University—School of Business Administration, Working Paper.

Part IV
Quantitative Finance

On the Risk Evaluation Method Based on the Market Model

Masaaki Kijima and Yukio Muromachi

1 Introduction

Risk management of interest-rate sensitive products has become more important than ever after the credit crunch, because the interest-rate market has changed drastically and traditional methods cannot be applied to evaluate risks involved in those products. For example, as many central banks have conducted the zero interest-rate policy (ZIRP) in order to support banks and corporate firms through a monetary policy, the low interest-rate environment has become common all over the world. Also, the multi-curve pricing approach has been adopted in the OTC interest-rate market, i.e. collateralised cash flows are discounted according to the OIS curve and non-collateralised cash flows are discounted by the appropriate funding rate.[1] Therefore, it is important to develop new methodology for the risk evaluation purpose under these circumstances. In this paper, we present a risk evaluation model for interest-rate sensitive products within the no-arbitrage framework.

Under the low interest-rate environment, Kijima et al. (2014) develop a risk evaluation methodology for mortgage-loan portfolios based on the single-factor quadratic Gaussian (QG) model. They take the QG spot-rate model under the observed probability measure to generate future scenarios of interest rates, and find the change of measure formula to derive the risk-neutral measure for evaluating future cash flows

[1] See, e.g., Bianchetti (2013) and Kijima et al. (2009) for details.

M. Kijima (✉) · Y. Muromachi
Tokyo Metropolitan University, 1-1 Minami-Ohsawa, Hachiohji, Tokyo 192-0397, Japan
e-mail: kijima@tmu.ac.jp

Y. Muromachi
e-mail: muromachi-yukio@tmu.ac.jp

R. Dieci et al. (eds.), *Nonlinear Economic Dynamics and Financial Modelling*,
DOI: 10.1007/978-3-319-07470-2_15,

associated with the mortgage-loan portfolios.[2] On the other hand, Kijima and Muromachi (2014) propose a simulation model to evaluate risks of interest-rate derivatives under the multi-curve setting. They adopt the QG spot-rate model to generate future scenarios of multi-curve interest rates that are consistent with the current multi-curve. The change of measure formula is also derived for the pricing of derivative securities at any future time within the no-arbitrage framework.

For the pricing of interest-rate derivatives, however, practitioners often use the market models such as Heath et al. (1992) and Brace et al. (1997), because they incorporate all current information in the yield curve, and arbitrage opportunities among bonds of different maturities are precluded. In this paper, we adopt the market model for risk evaluation purposes. To this end, we first consider a yield-curve model under the observed probability measure to generate future scenarios of interest rates, and then identify market prices of risk for the pricing of interest-rate derivatives under the risk-neutral measure at any future time.

The problem of modelling the real-world evolution of the term structure within the no-arbitrage framework has received little attention in the literature.[3] Most models of the term structure have been developed under the risk-neutral measure, simply because the major concern is to price interest-rate derivatives. Recently, Norman (2009) considers the problem of finding market prices of risk for the BGM model (1997) under the risk-neutral measure that are consistent with the term structure observed in the market. He proposes a form of the market prices of risk and illustrate an econometric method for calibrating them to historical forward (LIBOR) rates. It is shown through simulation studies that the model produces the yield curves with sensible shapes even over a long horizon. In this paper, we take the opposite way to address the risk evaluation problem for interest-rate sensitive products within the no-arbitrage framework.

Namely, we start with a yield-curve modelling under the observed probability measure based on the principal component analysis (PCA). Many empirical studies report that the changes in the yield curve can be well explained by the dominant three factors, i.e. the level, slope and curvature [see, e.g. Knez et al. (1994) for details]. Market participants have the consensus about this, and there is no other reason why we take this well-known result as the starting block of our risk evaluation problem for interest-rate sensitive products. We then investigate the conditions that market prices of risk must satisfy in order to preclude arbitrage opportunities. The yield-curve dynamics under the risk-neutral measure is given by the study of Brace et al. (1997). Hence, we can simply use this result for the pricing of derivative securities. Given the future scenarios of yield curve and the pricing results obtained so far in the literature, the prices of interest-rate sensitive products are calculated at any future time. Risk measures such as Value-at-Risk (VaR) of portfolios with interest-rate sensitive products can be evaluated through simple Monte Carlo simulation.

---

[2] See, e.g., Kijima and Muromachi (2000) for the general theory of risk evaluation methodology.

[3] The primary concern of many econometric papers for the term structures such as Diebold and Li (2006) is to provide point forecasts of yield-curve changes, not the arbitrage-free pricing. See also Norman (2009) and references therein for such models.

This paper is organised as follows. Section 2 starts with a yield-curve modelling under the observed probability measure, based on the principal component analysis (PCA). Some empirical results are reported for the cases that the PCA is applied to the forward-rate curve itself, the log-forward rate curve, and the square root of forward-rate curve. Section 3 investigates the conditions that market prices of risk must satisfy in order to preclude arbitrage opportunities. Examples are given to demonstrate that some market models often used in practice are not consistent with the no-arbitrage paradigm. Section 4 describes how to use the Monte Carlo simulation for the risk evaluation of interest-rate sensitive products. Finally, Section 5 concludes this paper. Appendix A provides an approximation method to generate a short-rate process under the risk-neutral measure. Throughout the paper, we assume that there exists a risk-neutral probability measure $\mathbb{Q}$. The observed probability measure is denoted by $\mathbb{P}$.

2 Yield-Curve Modelling under the Observed Measure

In this section, we construct a yield-curve model under the observed probability measure $\mathbb{P}$ based on the principal component analysis (PCA). Let $v(t, T)$ be the time-t price of the discount bond with maturity T, and let $D_t(x) = v(t, t + x)$, $x > 0$, denote the discount curve at time t. We want to generate the discount curve at any future time.

To this end, we consider the (instantaneous) forward-rate curve $R_t(x)$ defined by[4]

$$R_t(x) = -\frac{\partial}{\partial x} \log D_t(x); \qquad D_t(x) = \mathrm{e}^{-\int_0^x R_t(u)\mathrm{d}u}, \qquad x > 0, \tag{1}$$

and suppose that $R_t(x)$ follows the stochastic differential equation (SDE for short)

$$\mathrm{d}R_t(x) = \mu^r(t, x)\mathrm{d}t + \sum_i \sigma_i^r(t, x)\mathrm{d}w_{i,t}, \qquad t \geq 0, \tag{2}$$

for each $x > 0$ under the observed probability measure $\mathbb{P}$, where $w_{i,t}$ are the independent standard Brownian motions under $\mathbb{P}$. The drift μ^r as well as the volatilities σ_i^r is smooth enough to satisfy the regularity condition in order to guarantee the existence of $R_t(x)$ for all $x > 0$. Note that we can recover the discount function $D_t(x)$ at any future time t from that of $R_t(x)$.

In practice, it is well known that the changes in the forward-rate curve can be well explained by the dominant three factors, i.e. the level, slope and curvature. Let

[4] Norman (2009) considers a forward LIBOR rate of tenor δ and maturity T defined by $\frac{1}{\delta}\left(\frac{v(t,T)}{v(t,T+\delta)} - 1\right)$, rather than the instantaneous forward-rate curve $R_t(x)$, $x > 0$.

us denote the factors by $F_i(x)$, $i = 1, 2, 3$. Also, it seems reasonable to assume a mean reversion process for the evolution of interest rates over a *long period* of time. Therefore, a plausible specification of the forward-rate curve $R_t(x)$ in the SDE (2) is given by

$$\mathrm{d}R_t(x) = a(m(x) - R_t(x))\mathrm{d}t + F_1(x)\mathrm{d}w_{1,t} + F_2(x)\mathrm{d}w_{2,t} + F_3(x)\mathrm{d}w_{3,t}, \qquad t \geq 0, \tag{3}$$

where the mean-reverting level $m(x)$ is defined as the long-run average of the forward rate $R_t(x)$.[5] The speed of mean reversion represented by the parameter a is assumed to be independent of the term x just for the sake of simplicity.

The solution of the SDE (3) is given by

$$R_t(x) = m(x) + (R_0(x) - m(x))\mathrm{e}^{-at} + \sum_{i=1}^{3} F_i(x) \int_0^t \mathrm{e}^{-a(t-s)}\mathrm{d}w_{i,s}, \qquad t \geq 0, \tag{4}$$

for each $x > 0$. Note that, if the initial curve $R_0(x)$, the mean-reverting level $m(x)$, and the factors $F_i(x)$ are all continuous (differentiable, respectively), the solution $R_t(x)$ is also continuous (differentiable) in x for all $t \geq 0$. Throughout the paper, we use the model (3) as the basis of our risk evaluation methodology.

Note, however, that the right-hand side of (4) can be negative with positive probability. In order to avoid negative interest rates, some authors consider $L_t(x) = \log R_t(x)$ and assume that

$$\mathrm{d}L_t(x) = a^{\ell}(m^{\ell}(x) - L_t(x))\mathrm{d}t + F_1^{\ell}(x)\mathrm{d}w_{1,t} + F_2^{\ell}(x)\mathrm{d}w_{2,t} + F_3^{\ell}(x)\mathrm{d}w_{3,t}, \qquad t \geq 0. \tag{5}$$

Alternatively, consider the SDE

$$\mathrm{d}y_t(x) = -a^y y_t(x)\mathrm{d}t + F_1^y(x)\mathrm{d}w_{1,t} + F_2^y(x)\mathrm{d}w_{2,t} + F_3^y(x)\mathrm{d}w_{3,t}, \qquad t \geq 0, \tag{6}$$

and define $R_t(x) = (y_t(x) + m^y(x))^2$. The latter model is a generalisation of the quadratic Gaussian (QG) model considered in Kijima et al. (2014).

2.1 Principal Component Analysis

The models (3), (5) and (6) are directly connected to the principal component analysis (PCA). Let R be the correlation matrix of some *key* forward rates, and let β_j be its j-th eigenvector. That is, we have

[5] The formal derivation of (3) can be found in Jamshidian and Zhu (1997). In fact, they consider the model (5) below with the forward price as the drift term, not the mean reversion drift.

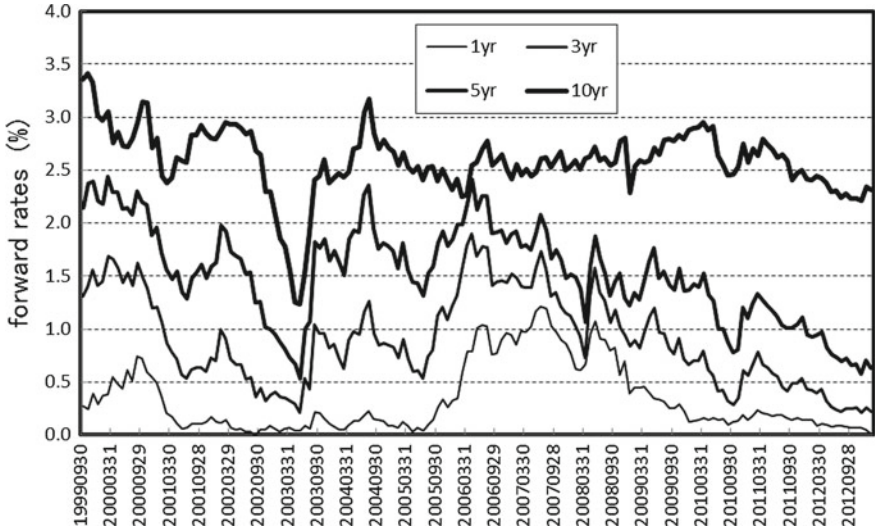

Fig. 1 Time series data of JGB forward rates ($x = 1, 3, 5, 10$ years)

$$R\beta_j = \eta_j \beta_j,$$

where η_j is the j-th eigenvalue. Since R is a symmetric, non-negative definite matrix, we have $\eta_j \geq 0$ and the eigenvectors are orthogonal to each other. Assuming that η_j are non-increasing in j, the vector β_j after normalisation (i.e. $\|\beta_j\|^2 = \eta_j$) is called the j-th *principal component*. The empirical analysis of historical yield-curve data (i.e. under the observed probability measure $\mathbb{P}$) demonstrates that all but the first three principal components are small in magnitude.

In this subsection, we perform the PCA for the data of the forward rates of Japanese government bond (JGB) at the end of months for the period of September, 1999–January, 2013. Some selected forward rates $R_t(x), x = 1, 3, 5, 10$, during this period are depicted in Fig. 1.[6] The forward-rate dynamics $R_t(x)$ for each x seems stationary during this period.

Figure 2 reports the results of PCA; the top panel shows the PCA results for the model (3), the middle panel for (5), and the bottom panel for (6), where we depict the first three principal components. Surprisingly, the three components are very similar for all the cases. Of course, this happens because the correlation matrices for the three cases are very similar for the data period. The first component (level) is not flat, because the yield curves under the ZIRP are typically S-shaped.[7] The second and third components are interpreted to be the slope and curvature as usual.

[6] The data at the end of August, 2002 are deleted because $R_t(1)$ was negative.

[7] See Gorovoi and Linetsky (2004) and Kabanov et al. (2007) for spot-rate models that can fit the S-shaped yield curves.

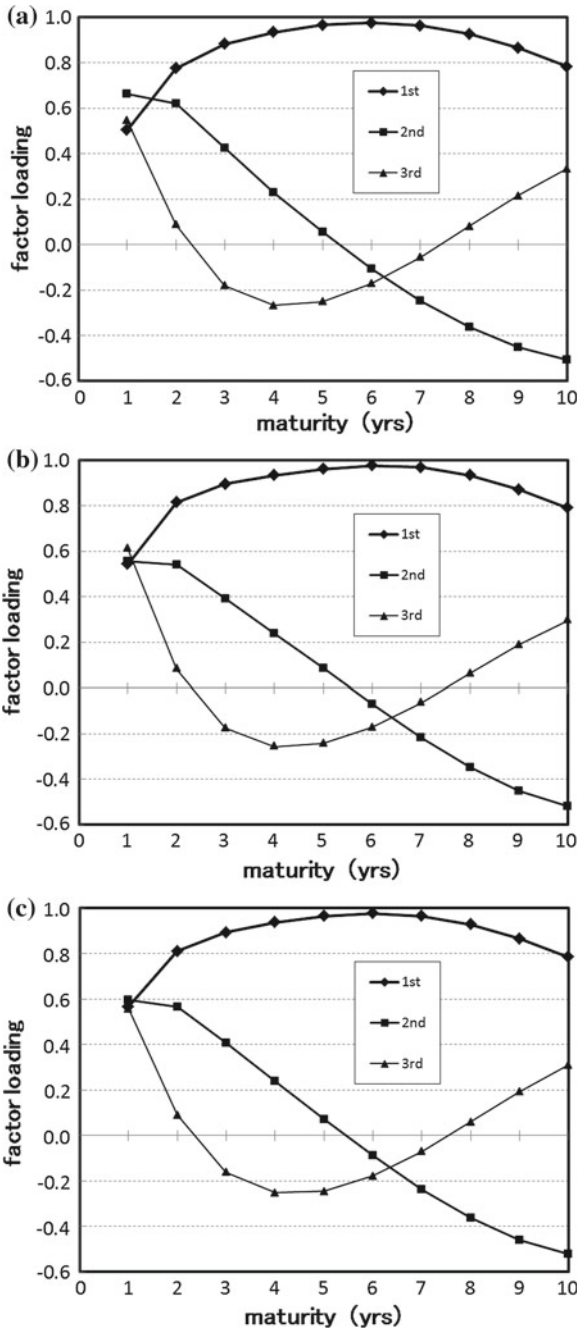

Fig. 2 Dominant three components of forward rates movements. **a** Forward rates, **b** Log-forward rates, **c** Square root of forward rates

2.2 Estimation of Parameters

According to Jamshidian and Zhu (1997), the factors $F_i(x)$ in the SDE (3) are related to the principal components β_j as

$$F_j(x) = \sigma(x)\beta_{xj}, \qquad j = 1, 2, 3,$$

where $\sigma(x)$ is the volatility of the forward rate $R_t(x)$ and β_{xj} is the x-th component of the vector β_j. The parameter a, the speed of mean reversion, as well as the mean reverting level $m(x)$ and the volatility $\sigma(x)$ for each x can be estimated by the ordinary GMM (generalised method of moments); see Kijima et al. (2014) for details. Recall that a is independent of x for the sake of simplicity. As we shall see later, however, the mean-reverting level $m(x)$ must satisfy some conditions in order to preclude arbitrage opportunities.

3 The Change of Measure

In this section, we consider the change of measure to derive the yield-curve dynamics under the risk-neutral measure $\mathbb{Q}$.

First, as in Brace et al. (1997), we assume that the forward rate $R_t(x)$ follows the SDE

$$\mathrm{d}R_t(x) = \frac{\partial}{\partial x}\left(\left(R_t(x) + \frac{1}{2}\sigma^2(t,x)\right)\mathrm{d}t + \sum_i \sigma_i(t,x)\mathrm{d}w^*_{i,t}\right) \tag{7}$$

under $\mathbb{Q}$, for some volatilities $\sigma_i(t,x)$ with $\sigma_i(t,0) = 0$ for all i, where $w^*_{i,t}$ are the independent standard Brownian motions under $\mathbb{Q}$ and $\sigma^2(t,x) = \sum_i \sigma_i^2(t,x)$. From (1), since

$$\log v(t,T) = -\int_0^{T-t} R_t(u)\mathrm{d}u,$$

by differentiating both sides with respect to t, we obtain

$$\mathrm{d}\log v(t,T) = R_t(T-t)\mathrm{d}t - \int_0^{T-t} [\mathrm{d}R_t(u)]\mathrm{d}u. \tag{8}$$

By substituting (7) into (8), we then have

$$\mathrm{d}\log v(t,T) = \left(R_t(0) - \frac{1}{2}\sigma^2(t,T-t)\right)\mathrm{d}t - \sum_i \sigma_i(t,T-t)\mathrm{d}w^*_{i,t}.$$

It follows that

$$\frac{\mathrm{d}v(t,T)}{v(t,T)} = R_t(0)\mathrm{d}t - \sum_i \sigma_i(t,T-t)\mathrm{d}w^*_{i,t}, \qquad 0 \le t \le T, \tag{9}$$

which implies that the denominated discount bond price is a martingale; hence, the risk-neutrality of the measure $\mathbb{Q}$ is justified.

3.1 Forward-Rate Dynamics

Consider the model (2) under the observed probability measure $\mathbb{P}$, and let $\lambda_i(t)$ be the market price of risk associated with $w_{i,t}$, i.e.

$$\mathrm{d}w^*_{i,t} = \mathrm{d}w_{i,t} - \lambda_i(t)\mathrm{d}t, \qquad i = 1, 2, \ldots, n. \tag{10}$$

The SDE (2) can be written as

$$\mathrm{d}R_t(x) = \nu^r(t,x)\mathrm{d}t + \sum_i \sigma^r_i(t,x)\mathrm{d}w^*_{i,t}, \tag{11}$$

where

$$\nu^r(t,x) = \mu^r(t,x) + \sum_i \sigma^r_i(t,x)\lambda_i(t).$$

The next result provides a condition that assures the model (2) under $\mathbb{P}$ to be consistent with the arbitrage-free model (7) under $\mathbb{Q}$.

Theorem 3.1 *The model (2) under $\mathbb{P}$ can be consistent with the arbitrage-free model (7) under $\mathbb{Q}$, if there exist $\lambda_i(t)$, the market prices of risk, that satisfy the condition*

$$\sum_i \lambda_i(t)\sigma^r_i(t,x) = \frac{\partial}{\partial x}R_t(x) - \mu^r(t,x) + \sum_i \sigma_i(t,x)\sigma^r_i(t,x) \tag{12}$$

for all $x > 0$, where

$$\sigma_i(t,s) = \int_0^s \sigma^r_i(t,u)\mathrm{d}u.$$

Proof By substituting (11) into (8), we have

$$\mathrm{d}\log v(t,T) = \left(R_t(T-t) - \int_0^{T-t} v^r(t,u)\mathrm{d}u \right) \mathrm{d}t - \sum_i \sigma_i(t,T-t)\mathrm{d}w^*_{i,t}.$$

It follows that

$$\frac{\mathrm{d}v(t,T)}{v(t,T)} = \left(R_t(T-t) - \int_0^{T-t} v^r(t,u)\mathrm{d}u + \frac{1}{2}\sigma^2(t,T-t) \right) \mathrm{d}t - \sum_i \sigma_i(t,T-t)\mathrm{d}w^*_{i,t}, \qquad 0 \le t \le T, \tag{13}$$

where

$$\sigma^2(t,T) = \sum_i \sigma_i^2(t,T).$$

Therefore, the model (2) under $\mathbb{P}$ can be consistent with the SDE (7) only if

$$\sum_i \lambda_i(t)\sigma_i(t,T-t) = R_t(T-t) - R_t(0) - \int_0^{T-t} \mu^r(t,u)\mathrm{d}u + \frac{1}{2}\sigma^2(t,T-t)$$

is satisfied for all $0 \le t \le T$. Differentiating both sides with respect to T, and then replacing $T-t$ by x, we obtain (12), completing the proof. □

Note that the market prices of risk, $\lambda_i(t)$, are independent of x. The x terms on the both sides of (12) must be matched by the remaining terms other than $\lambda_i(t)$.

When the PCA model (3) is considered, the condition (12) can be rewritten as follows.

Corollary 3.1 *The PCA model (3) under $\mathbb{P}$ can be consistent with the arbitrage-free model (7) under $\mathbb{Q}$, if there exist $\lambda_i(t)$, the market prices of risk, that satisfy the condition*

$$\begin{aligned} \sum_{i=1}^3 \lambda_i(t)F_i(x_k) = {} & \frac{\partial}{\partial x}\left(m(x_k) + (R_0(x_k) - m(x_k))\mathrm{e}^{-at} \right) \\ & + a(R_0(x_k) - m(x_k))\mathrm{e}^{-at} + \sum_{i=1}^3 F_i(x_k) \int_0^{x_k} F_i(u)\mathrm{d}u \\ & + \sum_{i=1}^3 (f_i(x_k) + aF_i(x_k)) \int_0^t \mathrm{e}^{-a(t-s)}\mathrm{d}w_{i,s}, \qquad k = 1,2,3, \end{aligned} \tag{14}$$

for all $x_1 < x_2 < x_3$.

Proof Consider the PCA model (3), and take $x_1 < x_2 < x_3$. It follows from (12) that

$$\begin{pmatrix} F_1(x_1) & F_2(x_1) & F_3(x_1) \\ F_1(x_2) & F_2(x_2) & F_3(x_2) \\ F_1(x_3) & F_2(x_3) & F_3(x_3) \end{pmatrix} \begin{pmatrix} \lambda_1(t) \\ \lambda_2(t) \\ \lambda_3(t) \end{pmatrix} \tag{15}$$
$$= \begin{pmatrix} \frac{\partial}{\partial x} R_t(x_1) - a(m(x_1) - R_t(x_1)) + \sum_{i=1}^{3} F_i(x_1) \int_0^{x_1} F_i(u)\mathrm{d}u \\ \frac{\partial}{\partial x} R_t(x_2) - a(m(x_2) - R_t(x_2)) + \sum_{i=1}^{3} F_i(x_2) \int_0^{x_2} F_i(u)\mathrm{d}u \\ \frac{\partial}{\partial x} R_t(x_3) - a(m(x_3) - R_t(x_3)) + \sum_{i=1}^{3} F_i(x_3) \int_0^{x_3} F_i(u)\mathrm{d}u \end{pmatrix}$$

for all $x_1 < x_2 < x_3$. From (4), we have

$$\frac{\partial}{\partial x} R_t(x) = \frac{\partial}{\partial x} \left(m(x) + (R_0(x) - m(x))\mathrm{e}^{-at} \right) + \sum_{i=1}^{3} f_i(x) \int_0^t \mathrm{e}^{-a(t-s)} \mathrm{d}w_{i,s},$$

where $f_i(x) = F_i'(x)$, which leads to the condition (14), completing the proof. □

Example 3.1 (Single-factor model) Consider the single-factor case, i.e. the model (3) with $n = 1$. In this case, the condition (14) is reduced to

$$\lambda_1(t) = \frac{1}{F_1(x)} \left[\frac{\partial}{\partial x} \left(m(x) + (R_0(x) - m(x))\mathrm{e}^{-at} \right) + a(R_0(x) - m(x))\mathrm{e}^{-at} \right] + \int_0^x F_1(u)\mathrm{d}u + \left(\frac{f_1(x)}{F_1(x)} + a \right) \int_0^t \mathrm{e}^{-a(t-s)} \mathrm{d}w_s,$$

and the right-hand side must be independent of x. Suppose that $F_1(x) = c\mathrm{e}^{bx}$ for some constants b and c. Then, the condition can be simplified as

$$\frac{\partial}{\partial x} \left(m(x) + (R_0(x) - m(x))\mathrm{e}^{-at} \right) + a(R_0(x) - m(x))\mathrm{e}^{-at} + \frac{c^2}{b} \mathrm{e}^{bx} \left(\mathrm{e}^{bx} - 1 \right) = 0.$$

Given the initial forward-rate curve $R_0(x)$, the mean-reverting level $m(x)$ can be solved to satisfy the above equation. In particular, when $R_0(x) = m(x)$, we have

$$m(x) = R_0(0) - \frac{c^2}{2b^2} \left(1 - \mathrm{e}^{bx} \right)^2, \qquad x \geq 0.$$

The market price of risk $\lambda_1(t)$ becomes a deterministic function of time t only if $a + b = 0$.

3.2 The Log-Forward Rate Model

Next, we consider the model (5) under the observed probability measure $\mathbb{P}$, whose solution is given by

$$L_t(x) = m^\ell(x) + (L_0(x) - m^\ell(x))\mathrm{e}^{-a^\ell t} + \sum_{i=1}^{3} F_i^\ell(x) \int_0^t \mathrm{e}^{-a^\ell(t-s)} \mathrm{d}w_{i,s}, \qquad t \geq 0. \tag{16}$$

Let $\lambda_i(t)$ be the market price of risk as in (10). Then, the SDE (5) can be written as

$$\mathrm{d}L_t(x) = a^\ell(m^\ell(t,x) - L_t(x))\mathrm{d}t + F_1^\ell(x)\mathrm{d}w_{1,t}^* + F_2^\ell(x)\mathrm{d}w_{2,t}^* + F_3^\ell(x)\mathrm{d}w_{3,t}^*, \tag{17}$$

where

$$m^\ell(t,x) = m^\ell(x) + \sum_i \frac{F_i^\ell(x)}{a^\ell}\lambda_i(t).$$

From (17), the forward rate $R_t(x) = \exp(L_t(x))$ follows the SDE

$$\frac{\mathrm{d}R_t(x)}{R_t(x)} = \left(\nu_F(t,x) + \frac{1}{2}\sigma_F^2(x)\right)\mathrm{d}t + \sum_{i=1}^{3} F_i^\ell(x)\mathrm{d}w_{i,t}^*, \tag{18}$$

where

$$\nu_F(t,x) = a^\ell(m^\ell(t,x) - \log R_t(x)), \qquad \sigma_F^2(x) = \sum_{i=1}^{3}(F_i^\ell(x))^2.$$

By substituting (18) into (8), we then have

$$\begin{aligned}\mathrm{d}\log v(t,T) = &\left(R_t(T-t) - \int_0^{T-t} \nu_F(t,u)R_t(u)\mathrm{d}u - \frac{1}{2}\int_0^{T-t} \sigma_F^2(u)R_t(u)\mathrm{d}u\right)\mathrm{d}t \\ &- \sum_{i=1}^{3} \sigma_{FR,i}(t,T-t)\mathrm{d}w_{i,t}^*,\end{aligned}$$

where

$$\sigma_{FR,i}(t,s) = \int_0^s F_i^\ell(u)R_t(u)\mathrm{d}u.$$

It follows that

$$\frac{\mathrm{d}v(t,T)}{v(t,T)} = M^{\ell}(T-t)\mathrm{d}t - \sum_{i=1}^{3}\sigma_{FR,i}(t,T-t)\mathrm{d}w_{i,t}^{*}, \tag{19}$$

where

$$M^{\ell}(x) = R_t(x) - \int_0^x v_F(t,u)R_t(u)\mathrm{d}u - \frac{1}{2}\int_0^x \sigma_F^2(u)R_t(u)\mathrm{d}u + \frac{1}{2}\sigma_{FR}^2(t,x)$$

and

$$\sigma_{FR}^2(t,x) = \sum_{i=1}^{3}\sigma_{FR,i}^2(t,x).$$

Therefore, the model (5) under $\mathbb{P}$ can be consistent with the SDE (7) only if

$$\sum_{i=1}^{3}\lambda_i(t)\sigma_{FR,i}(t,T-t) = R_t(T-t) - R_t(0) - \int_0^{T-t}\left(a^{\ell}(m^{\ell}(u) - \log R_t(u))\right)R_t(u)\mathrm{d}u - \frac{1}{2}\int_0^{T-t}\sigma_F^2(u)R_t(u)\mathrm{d}u + \frac{1}{2}\sigma_{FR}^2(t,T-t)$$

is satisfied for all $0 \le t \le T$. Differentiating both sides with respect to T, and then replacing $T-t$ by x, we obtain

$$\sum_{i=1}^{3}\lambda_i(t)F_i^{\ell}(x) = \frac{\partial L_t(x)}{\partial x} - a^{\ell}(m^{\ell}(x) - L_t(x)) - \frac{1}{2}\sigma_F^2(x) + \sum_{i=1}^{3}\sigma_{FR,i}(t,x)F_i^{\ell}(x) \tag{20}$$

for all $x > 0$, by using the relationship between $L_t(x)$ and $R_t(x)$.

Note from (16) that

$$\frac{\partial}{\partial x}L_t(x) = \frac{\partial}{\partial x}\left(m^{\ell}(x) + (L_0(x) - m^{\ell}(x))\mathrm{e}^{-a^{\ell}t}\right) + \sum_{i=1}^{3}f_i^{\ell}(x)\int_0^t \mathrm{e}^{-a^{\ell}(t-s)}\mathrm{d}w_{i,s},$$

where $f_i^{\ell}(x) = \mathrm{d}F_i^{\ell}(x)/\mathrm{d}x$. By the same argument as in Sect. 3.1, it follows that the condition (20) is equivalent to

$$\sum_{i=1}^{3}\lambda_i(t)F_i^\ell(x_k)=\frac{\partial}{\partial x}\left(m^\ell(x_k)+(L_0(x_k)-m^\ell(x_k))\mathrm{e}^{-a^\ell t}\right)$$
$$+a^\ell(L_0(x_k)-m^\ell(x_k))\mathrm{e}^{-a^\ell t}-\frac{1}{2}\sigma_F^2(x_k)+\sum_{i=1}^{3}\sigma_{FR,i}(t,x_k)F_i^\ell(x_k)$$
$$+\sum_{i=1}^{3}(f_i^\ell(x_k)+a^\ell F_i^\ell(x_k))\int_0^t\mathrm{e}^{-a^\ell(t-s)}\mathrm{d}w_{i,s} \tag{21}$$

for all $x_1<x_2<x_3$. If this holds true, we can use both the PCA model (5) under the observed probability measure $\mathbb{P}$ and the SDE (7) with $\sigma_i(t,x)=F_i^\ell(x)$ under $\mathbb{Q}$ as the forward-rate dynamics.

Example 3.2 (Single-factor model) Consider the single-factor case in the model (5). In this case, the condition (21) is reduced to

$$\lambda_1(t)=\frac{1}{F_1^\ell(x)}\left[\frac{\partial}{\partial x}\left(m^\ell(x)+(L_0(x)-m^\ell(x))\mathrm{e}^{-a^\ell t}\right)+a^\ell(L_0(x)-m^\ell(x))\mathrm{e}^{-a^\ell t}\right]$$
$$-\frac{1}{2}F_1^\ell(x)+\sigma_{FR,1}(t,x)+\left(\frac{f_1^\ell(x)}{F_1^\ell(x)}+a^\ell\right)\int_0^t\mathrm{e}^{-a^\ell(t-s)}\mathrm{d}w_s,$$

and the right-hand side must be independent of x. Suppose that $F_1^\ell(x)=c\mathrm{e}^{bx}$ for some constants b and c. Then, the condition can be simplified as

$$\frac{\partial}{\partial x}\left(m^\ell(x)+(L_0(x)-m^\ell(x))\mathrm{e}^{-a^\ell t}\right)+a^\ell(L_0(x)-m^\ell(x))\mathrm{e}^{-a^\ell t}$$
$$-\frac{c^2}{2}\mathrm{e}^{2bx}+c^2\mathrm{e}^{bx}\int_0^x\mathrm{e}^{L_t(u)+bu}\mathrm{d}u=0.$$

Since $L_t(x)$ is a stochastic process, there does not exist a deterministic solution of $m^\ell(x)$ for the non-trivial case ($c\neq 0$).

3.3 The QG Model

Finally, we consider the model (6) under the observed probability measure $\mathbb{P}$, whose solution is given by

$$y_t(x)=y_0(x)\mathrm{e}^{-a^y t}+\sum_{i=1}^{3}F_i^y(x)\int_0^t\mathrm{e}^{-a^y(t-s)}\mathrm{d}w_{i,s},\qquad t\geq 0. \tag{22}$$

As before, let $\lambda_i(t)$ be the market price of risk. Then, the SDE (6) can be written as

$$\mathrm{d}y_t(x) = a^y(v^y(t,x) - y_t(x))\mathrm{d}t + \sum_{i=1}^{3} F_i^y(x)\mathrm{d}w_{i,t}^*, \tag{23}$$

where

$$v^y(t,x) = \frac{1}{a^y}\sum_{i=1}^{3} F_i^y(x)\lambda_i(t).$$

From (23), the forward rate $R_t(x) = (y_t(x) + m^y(x))^2$ follows the SDE

$$\begin{aligned}\mathrm{d}R_t(x) = &\left[2a^y(y_t(x) + m^y(x))(v^y(t,x) - y_t(x)) + \sigma_F^2(x)\right]\mathrm{d}t \\ &+ 2(y_t(x) + m^y(x))\sum_{i=1}^{3} F_i^y(x)\mathrm{d}w_{i,t}^*,\end{aligned} \tag{24}$$

where $\sigma_F^2(x) = \sum_{i=1}^{3}(F_i^y(x))^2$. By substituting (24) into (8), we then have

$$\begin{aligned}\mathrm{d}\log v(t,T) = \Bigg[R_t(T-t) &- \int_0^{T-t} 2a^y(y_t(u) + m^y(u))(v^y(t,u) - y_t(u))\mathrm{d}u \\ &- \int_0^{T-t} \sigma_F^2(u)\mathrm{d}u\Bigg]\mathrm{d}t - \sum_{i=1}^{3}\sigma_{Fy,i}(t, T-t)\mathrm{d}w_{i,t}^*,\end{aligned}$$

where

$$\begin{aligned}\sigma_{Fy,i}(t,s) &= 2\int_0^s F_i^y(u)(y_t(u) + m^y(u))\mathrm{d}u \\ &= 2\mathrm{e}^{-a^y t}\int_0^s F_i^y(u)y_0(u)\mathrm{d}u + 2\int_0^s F_i^y(u)m^y(u)\mathrm{d}u \\ &\quad + 2\sum_{j=1}^{3}\Sigma_{ij}(s)\int_0^t \mathrm{e}^{-a^y(t-s)}\mathrm{d}w_{i,s}\end{aligned}$$

with

$$\Sigma_{ij}(s) = \int_0^s F_i^y(u)F_j^y(u)\mathrm{d}u.$$

It follows that

$$\frac{\mathrm{d}v(t,T)}{v(t,T)} = \left[R_t(T-t) - \int_0^{T-t} 2a^y(y_t(u)+m^y(u))(v^y(t,u)-y_t(u))\mathrm{d}u \right.$$
$$\left. - \int_0^{T-t} \sigma_F^2(u)\mathrm{d}u + \frac{1}{2}\sigma_{Fy}^2(t,T-t) \right] \mathrm{d}t - \sum_{i=1}^{3} \sigma_{Fy,i}(t,T-t)\mathrm{d}w_{i,t}^*,$$

where

$$\sigma_{Fy}^2(t,s) = \sum_{i=1}^{3} \sigma_{Fy,i}^2(t,s).$$

Therefore, the model $R_t(x) = (y_t(x)+m^y(x))^2$ with (6) under $\mathbb{P}$ can be consistent with the SDE (7) only if

$$\sum_{i=1}^{3} \lambda_i(t)\sigma_{Fy,i}(t,T-t) = R_t(T-t) - R_t(0) + \int_0^{T-t} 2a^y(y_t(u)+m^y(u))y_t(u)\mathrm{d}u$$
$$- \int_0^{T-t} \sigma_F^2(u)\mathrm{d}u + \frac{1}{2}\sigma_{Fy}^2(t,T-t)$$

is satisfied for all $0 \le t \le T$. Differentiating both sides with respect to T, and then replacing $T-t$ by x, we obtain

$$2(y_t(x)+m^y(x))\sum_{i=1}^{3}\lambda_i(t)F_i^y(x) = \frac{\partial}{\partial x}R_t(x) + 2a^y(y_t(x)+m^y(x))y_t(x) - \sigma_F^2(x)$$
$$+ 2(y_t(x)+m^y(x))\sum_{i=1}^{3}\sigma_{Fy,i}(t,x)F_i^y(x) \tag{25}$$

for all $x > 0$. Using $R_t(x) = (y_t(x)+m^y(x))^2$ and (22), we can rewrite the condition (25) as

$$\sum_{i=1}^{3}\lambda_i(t)F_i^y(x) = \frac{\partial}{\partial x}(y_t(x)+m^y(x)) + a^y y_t(x) - \frac{\sigma_F^2(x)}{2(y_t(x)+m^y(x))}$$
$$+ \sum_{i=1}^{3}\sigma_{Fy,i}(t,x)F_i^y(x)$$

$$= \frac{\partial}{\partial x}\left(y_0(x)\mathrm{e}^{-a^y t} + m^y(x)\right) + a^y y_0(x)\mathrm{e}^{-a^y t} - \frac{\sigma_F^2(x)}{2(y_t(x) + m^y(x))}$$
$$+ \sum_{i=1}^{3} \sigma_{Fy,i}(t,x) F_i^y(x) + \sum_{i=1}^{3}\left(f_i^y(x) + a^y F_i^y(x)\right) \int_0^t \mathrm{e}^{-a^y(t-s)} \mathrm{d}w_{i,s}, \tag{26}$$

where $f_i^y(x) = \mathrm{d}F_i^y(x)/\mathrm{d}x$.

Example 3.3 (Single-factor model) Consider the single-factor case. In this case, the condition (26) is reduced to

$$\lambda_1(t) = \frac{1}{F_1^y(x)}\left(\frac{\partial}{\partial x}\left(y_0(x)\mathrm{e}^{-a^y t} + m^y(x)\right) + a^y y_0(x)\mathrm{e}^{-a^y t}\right)$$
$$+ 2\mathrm{e}^{-a^y t}\int_0^x F_1^y(u) y_0(u)\mathrm{d}u + 2\int_0^x F_1^y(u) m^y(u)\mathrm{d}u$$
$$- \frac{1}{2}\left(\frac{y_0(x)\mathrm{e}^{-a^y t} + m^y(x)}{F_1^y(x)} + \int_0^t \mathrm{e}^{-a^y(t-s)}\mathrm{d}w_s\right)^{-1}$$
$$+ \left(\frac{f_1^y(x)}{F_1^y(x)} + 2\Sigma_{11}(x) + a^y\right)\int_0^t \mathrm{e}^{-a^y(t-s)}\mathrm{d}w_s. \tag{27}$$

In the right-hand side of (27), the first, second and third terms are deterministic, while the fourth and fifth terms are stochastic. Since the right-hand side of (27) must be independent of x, it seems difficult to find the model which satisfies the condition (27).

4 Monte Carlo Simulation

According to Kijima and Muromachi (2000), any risk evaluation model comprises the following three components:

1. Generation of stochastic scenarios for uncertainty,
2. Valuation of the present values, and
3. Valuation of the future distribution.

The input data in the model are the present discount curve, either observed in the market or evaluated if necessary, and parameters of the stochastic structure. In this section, we consider the model (3) for the sake of simplicity, although the forward rates become negative with positive probability.

4.1 Generation of Scenarios

Suppose that the functions $m(x)$, $F_i(x)$, $i = 1, 2, 3$, as well as the parameter a in the SDE (3) are all estimated and the current forward-rate curve $R_0(x)$ is observed. Then, we have all the information needed to generate future scenarios of forward rates. Let us denote the ith scenario by $R_t^i(x)$, $x > 0$, at future time $t > 0$.

4.2 Valuation of Derivative Prices

In order to obtain the future derivative prices, we assume that the market prices of risk $\lambda_i(t)$ satisfying the condition (14) always exist. In other words, the change of measure to derive the SDE (7) is always possible. Hence, at some future time $u > 0$, given the scenario $R_u^i(x)$ as the initial condition, the forward-rate dynamics after time u can be generated by the SDE (7). Note that the scenario $R_u^i(x)$ is generated by (3) under the observed probability measure $\mathbb{P}$, not under the risk-neutral measure $\mathbb{Q}$.

Example 4.1 (Swap rate) Let t be some future time, and we simulate a scenario $R_t^i(x)$, $x > 0$, according to the SDE (3). Then, from (1), we obtain a scenario $D_t(x)$, $x > 0$, of the (secured) discount curve. The swap rate $S_t(T)$ with payment schedule $\{T_k\}$, $k = 1, 2, \ldots, n$, where $T_n = T$ and $\delta = T_k - T_{k-1}$, is determined as

$$S_t(T) = \frac{\sum_{k=1}^n v(t, T_k) L_k(t)}{\sum_{k=1}^n v(t, T_k)}, \tag{28}$$

under the multi-curve environment, where $L_k(t)$ denotes the (risky) forward LIBOR at time t [see, e.g. Bianchetti (2013) for details]. Of course, if the forward LIBOR is secured, we have

$$L_k(t) = \frac{1}{\delta}\left(\frac{v(t, T_{k-1})}{v(t, T_k)} - 1\right), \tag{29}$$

and the swap rate $S_t(T)$ is given by the well known *telescopic* form

$$S_t(T) = \frac{v(t, T_0) - v(t, T_n)}{\delta \sum_{k=1}^n v(t, T_k)}.$$

Note that, in order to generate future swap rate given by (28), we also need to generate the forward LIBOR $L_k(t)$ for any future time. This could be done by considering the risky discount curve $D_t^L(x)$, $x > 0$, associated with the LIBOR interest rates. The methodology is the same and omitted.

Example 4.2 (Caplet) Let $L_i(t)$ be the (secured) forward LIBOR defined by (29). According to Brace et al. (1997), if we start from the model (7), the forward LIBOR $L_i(t)$ follows the SDE

$$\frac{\mathrm{d}L_i(t)}{L_i(t)} = \gamma(t, T_i - t)\mathrm{d}w_t^{T_{i+1}}$$

for some $\gamma(t, x)$, where $w_t^{T_{i+1}}$ denotes the standard Brownian motion under the forward-neutral measure $\mathbb{Q}^{T_{i+1}}$ [see, e.g. Kijima (2013) for details]. It is common to assume for practitioners that $\gamma(t, x)$ is a deterministic function. In this case, the caplet price is given by the Black formula with the volatility $v_i(t)$, where

$$v_i^2(t) = \int_t^{T_i} \gamma^2(s, T_i - s)\mathrm{d}s.$$

Example 4.3 (Swaption) The swaption price can be evaluated by, e.g. a formula given by Brace et al. (1997).

Example 4.4 (General Cash flow) Consider a mortgage loan and suppose that, as in Kijima et al. (2014), the prepayment rate depends on the swap rate. More precisely, let $C(T) = C(T, S_T(\tau))$ be the (secured) cashflow caused by the prepayment at future time T for some $\tau > T$. The future swap rate can be generated as in Example 4.1 under $\mathbb{P}$. However, in order to evaluate the cash flow within the no-arbitrage paradigm, we need to discount it with respect to the short rate $R_t(0)$ under $\mathbb{Q}$. That is, the present value of the cash flow is given by

$$\mathrm{PV}(C) = \mathbb{E}^{\mathbb{Q}}\left[\exp\left\{-\int_0^T R_t(0)\mathrm{d}t\right\} C(T, S_T(\tau))\right],$$

where $\mathbb{E}^{\mathbb{Q}}$ denotes the expectation under $\mathbb{Q}$. Note that the short-rate dynamics $R_t(0)$ follows the SDE

$$\mathrm{d}R_t(0) = \frac{\partial}{\partial x}\left(\left(R_t(x) + \frac{1}{2}\sigma^2(t, x)\right)\mathrm{d}t + \sum_i \sigma_i(t, x)\mathrm{d}w_{i,t}^*\right)\Big|_{x=0}. \tag{30}$$

Hence, in general, the evolution of the spot-rate process is non-Markovian with respect to a finite-dimensional state space.[8] This means that a significantly complex computation scheme and a large memory storage space are required to price contingent claims against the term structure. In Appendix A, we propose an approximate scheme to generate the short rates.

[8] See, e.g. Chiarella and Kwon (2003) and Inui and Kijima (1998) for finite-dimensional Markovian HJM term structure models.

5 Concluding Remarks

In this paper, we present a risk evaluation model for interest-rate sensitive products within the no-arbitrage framework. A yield-curve dynamics is modelled, based on the results of the principal component analysis (PCA), to generate future scenarios of interest rates under the observed probability measure $\mathbb{P}$. The market model is adopted for the pricing of interest-rate derivatives under the risk-neutral measure $\mathbb{Q}$ by identifying market prices of risk that are consistent with the yield-curve model. Given the future scenarios of yield curve and the market prices of risk, the prices of interest-rate sensitive products are calculated at any future time. Risk measures such as Value-at-Risk (VaR) of portfolios with interest-rate sensitive products can be evaluated through simple Monte Carlo simulation.

However, as shown by the single-factor examples, it seems difficult to judge whether or not the given PCA model (under $\mathbb{P}$) is consistent with the no-arbitrage paradigm. One way to check it may be to determine the market prices of risk, $\lambda_i(t)$, and the mean-reverting level, $m(x)$, approximately by assuming their parametric forms, as in Norman (2009), and calibrating them to the market data. This method may sound, because the *mean* level, $m(x)$, and the *latent* factors, $\lambda_i(t)$, are difficult to estimate in a stable way. This is the subject left to the future research in conjunction with empirical research.

Acknowledgments The authors are grateful for the research grant funded by the Grant-in-Aid (A) (#26242028) from Japan's Ministry of Education, Culture, Sports, Science and Technology.

Appendix: A Spot-Rate Model under $\mathbb{Q}$

In this appendix, we explain how to generate the short rates $R_t(0)$ under $\mathbb{Q}$. To this end, take $h > 0$ sufficiently small, and assume that

$$\frac{\partial}{\partial x} R_t(x)\Big|_{x=0} = \frac{1}{h}[R_t(h) - R_t(0)].$$

Then, from (30), we have

$$\mathrm{d}R_t(0) = \left(\frac{1}{h}[R_t(h) - R_t(0)] + \frac{1}{2}\frac{\partial}{\partial x}\sigma^2(t,0)\right)\mathrm{d}t + \sum_i \frac{\partial}{\partial x}\sigma_i(t,0)\mathrm{d}w^*_{i,t} \quad \text{(A.1)}$$

and

$$\mathrm{d}R_t(h) = \left(\frac{1}{h}[R_t(h) - R_t(0)] + \frac{1}{2}\frac{\partial}{\partial x}\sigma^2(t,h)\right)\mathrm{d}t + \sum_i \frac{\partial}{\partial x}\sigma_i(t,h)\mathrm{d}w^*_{i,t}.$$

Let $S_t = R_t(h) - R_t(0)$, so that

$$\mathrm{d}S_t = \left(\frac{1}{2}\frac{\partial}{\partial x}\left[\sigma^2(t,h) - \sigma^2(t,0)\right]\right)\mathrm{d}t + \sum_i \frac{\partial}{\partial x}[\sigma_i(t,h) - \sigma_i(t,0)]\mathrm{d}w^*_{i,t}.$$

In particular, when $\sigma_i(t,x) = \int_0^x F_i(u)\mathrm{d}u$ as in the PCA model (3), we have

$$\mathrm{d}S_t = \sum_{i=1}^{3}\left(F_i(h)\int_0^h F_i(u)\mathrm{d}u\right)\mathrm{d}t + \sum_{i=1}^{3}[F_i(h) - F_i(0)]\mathrm{d}w^*_{i,t}.$$

Given the solution $\{S_t\}$ at hand, it follows from (A.1) that

$$\mathrm{d}R_t(0) = \left(\frac{1}{h}S_t + \frac{1}{2}\frac{\partial}{\partial x}\sigma^2(t,0)\right)\mathrm{d}t + \sum_i \frac{\partial}{\partial x}\sigma_i(t,0)\mathrm{d}w^*_{i,t}. \tag{A.2}$$

Using (A.2), the spot rates $R_t(0)$ can be generated approximately by Monte Carlo simulation.

References

Bianchetti, M. (2013). Modern pricing of interest rate derivatives including funding and collateral. In M. Bianchetti, M. Morini (Eds.), *Interest Rate Modelling after the Financial Crisis* (pp. 113–152). London: Risk Books.

Brace, A., Gątarek, D., & Musiela, M. (1997). The market model of interest rate dynamics. *Mathematical Finance*, *7*, 127–155.

Chiarella, C., & Kwon, O. K. (2003). Finite dimensional affine realisations of HJM models in terms of forward rates and yields. *Review of Derivatives Research*, *6*, 129–155.

Diebold, F. X., & Li, C. (2006). Forecasting the term structure of governmental bond yields. *Journal of Econometrics*, *130*, 337–364.

Gorovoi, V., & Linetsky, V. (2004). Black's model of interest rates as options, eigenfunction expansions and Japanese interest rates. *Mathematical Finance*, *14*, 49–78.

Heath, D., Jarrow, R., & Morton, A. (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica*, *60*, 77–105.

Inui, K., & Kijima, M. (1998). A Markovian framework in multi-factor Heath-Jarrow-Morton models. *Journal of Financial and Quantitative Analysis*, *33*, 423–440.

Jamshidian, F., & Zhu, Y. (1997). Scenario simulation: Theory and methodology. *Finance and Stochastics*, *1*, 43–67.

Kabanov, Y., Kijima, M., & Rinaz, S. (2007). A positive interest rate model with sticky barrier. *Quantitative Finance*, *7*, 269–284.

Kijima, M. (2013). *Stochastic Processes with Applications to Finance* (2nd ed.). London: Chapman & Hall.

Kijima, M., & Muromachi, Y. (2000). Evaluation of credit risk of a portfolio with stochastic interest rate and default processes. *Journal of Risk*, *3*, 5–36.

Kijima, M., & Muromachi, Y. (2014). Risk evaluation of interest-rate derivatives under the multi-curve pricing approach, in preparation.

Kijima, M., Suzuki, Y., & Tamba, Y. (2014). Risk evaluation of mortgage-Loan portfolios under low interest-rate environment. *Journal of Risk*, forthcoming.

Kijima, M., Tanaka, K., & Wong, T. (2009). A multi-quality model of interest rates. *Quantitative Finance*, *9*, 133–145.

Knez, P. J., Litterman, R., & Scheinkman, J. A. (1994). Explorations into factors explaining money market returns. *Journal of Finance*, *49*, 1861–1882.

Norman, J. P. (2009). Real world interest rate modelling with the BGM model, Working Paper, EMB Consultancy LLP, Cambridge.

On Multicurve Models for the Term Structure

Laura Morino and Wolfgang J. Runggaldier

1 Introduction

In the wake of the big crisis one has witnessed a significant increase in the spreads between LIBORs of different tenors as well as the spread between a LIBOR and the discount curve (LIBOR-OIS). This has led to the construction of *multicurve models* where, typically, future cash flows are generated through curves associated to the underlying rates, but are discounted by another curve.

The majority of the models that have been considered reflects the usual classical distinction between

(i) short rate models;
(ii) HJM setup;
(iii) BGM or LIBOR market models.

By analogy to credit risk we may call the first two categories of models as *bottom-up* models, while the third one could be classified as *top-down*. In addition, methodologies have appeared that are related to foreign exchange.

Here we consider only the first two setups. We begin by discussing some issues arising with the HJM methodology and concentrate then on short rate models. The third setup (top-down) is mainly present in work by F. Mercurio and co-authors (see e.g. Mercurio (2010a, b)), but also in other recent work such

L. Morino · W. J. Runggaldier (✉)
Dipartimento di Matematica Pura ed Applicata, Università di Padova,
Via Trieste 63, 35121 Padua, Italy
e-mail: runggal@math.unipd.it

Present address:
L. Morino
Deloitte Consulting Srl., Milan, Italy
e-mail: laura.morino88@gmail.com

R. Dieci et al. (eds.), *Nonlinear Economic Dynamics and Financial Modelling*,
DOI: 10.1007/978-3-319-07470-2_16,

as Keller-Ressel et al. (2013). There are advantages and disadvantages with each setup. Among the possible advantages of short rate models is the fact that they lead more easily to a Markovian setting, which is convenient for various calculations (see Crépey et al. (2013b)). On the other hand, one of the major advantages of HJM over a direct short rate modeling is that the model is automatically calibrated to the initial term structure. Short rate models in a multi-curve setup have already appeared in the literature, e.g. Kijima et al. (2009), Kenyon (2010), Filipović and Trolle (2013).

To present the basic ideas in a simple way, here we consider a *two-curve model*, namely with a curve for discounting and one for generating future cash flows. The choice of the discount curve is not unique; we follow the common choice of considering the *OIS swap curve*. For the risky cash flows without collateral we consider a single LIBOR (i.e. *for a given tenor structure*).

We present an approach for the pricing of some basic LIBOR-related derivatives, namely FRAs and CAPs (*linear/nonlinear*) and consider only *clean valuation* formulas, namely without counterparty risk. Although real pricing problems require a more global approach (see e.g. the discussions in Fuji et al. (2009, 2011), Piterbarg (2010), Crépey et al. (2013a) as well as in recent work by D. Brigo and co-authors such as Pallavicini and Brigo (2013), Brigo, Morini and Pallavicini (2013)), clean valuation formulas are nevertheless useful for various reasons: as pointed out in Crépey et al. (2013b), market quotes typically reflect prices of fully collateralized transactions so that clean price formulas may turn out to be sufficient for calibration also when using the model to compute possible value adjustments; furthermore (see Crépey et al. (2013b)), TVA adjustments are often computed on top of clean prices. Concerning methodology, since our approach is of the bottom-up type that considers short rate modeling, we heavily exploit the advantages of an affine term structure. This is in contrast with top-down approaches, where (see Mercurio (2010a, b)) log-normal models are common (see however Keller-Ressel et al. (2013) and Grbac et al. (2014) for affine LIBOR models with general distributions in a multi-curve context).

Traditionally, interest rates are defined to be coherent with the bond prices $p(t, T)$, which represent the expectation of the market concerning the future value of money. For the discrete compounding forward LIBORs, which we denote here by $L(t; T, S)$, this leads to ($t < T < S$)

$$L(t; T, S) = \frac{1}{S - T} \left(\frac{p(t, T)}{p(t, S)} - 1 \right) \tag{1}$$

which can also be justified as representing the fair value of the fixed rate in a FRA on the LIBOR. Since we consider only a single LIBOR that corresponds to a given tenor structure, we assume $S = T + \Delta$ (for tenor Δ). In this way one obtains a single curve for the term structure. The actual LIBOR rates, which in what follows we shall denote by $\bar{L}(t; T, T + \Delta)$, are determined by the LIBOR panel that takes into account various factors such as credit risk, liquidity, etc. (see the discussion in Filipović and Trolle (2013)). Following some of the recent literature, in particular Crépey et al. (2012) (see also Kijima et al. (2009)), we keep the formal relationship

(1) between LIBOR rates and bond prices, but replace the risk-free bond prices $p(t,T)$ by fictitious *"risky" bond prices* $\bar{p}(t,T)$ that are supposed to be affected by the same factors as the actual LIBORs and that, analogously to the risk-free bond prices, we define then as

$$\bar{p}(t,T) = E^Q\left\{\exp\left[-\int_t^T (r_u + s_u)du\right] \mid \mathscr{F}_t\right\} \tag{2}$$

where r_t is the classical short rate, whereas s_t represents the short rate spread (*hazard rate in case of only default risk*). Notice that in this way the spread is introduced from the outset. Notice also that the fictitious bond prices $\bar{p}(t,T)$ are not actual prices.

Since in what follows we are interested in FRAs and CAPs that are based on the $T-$spot LIBOR $\bar{L}(T;T,T+\Delta)$, we actually postulate the relationship (1) only at the inception time $t = T$. Our starting point is thus the following relationship

$$\bar{L}(T;T,T+\Delta) = \frac{1}{\Delta}\left(\frac{1}{\bar{p}(T,T+\Delta)} - 1\right) \tag{3}$$

where we have taken into account the fact that also for the "risky" bonds we have $\bar{p}(T,T) = 1$.

In addition to the pricing of FRAs and CAPs in our two-curve setup, our major goal here is to derive a relationship between theoretically risk-free and actual FRAs (possibly also CAPs) thereby exhibiting an *adjustment factor* which plays a role analogous to that of the quanto adjustments in the pricing of cross-currency derivatives or the "multiplicative forward basis" in Bianchetti (2012).

2 The Model

2.1 Preliminary Considerations

We start with some comments concerning HJM-like approaches to better motivate our short rate approach. Given the bond price processes $p(t,T)$ and $\bar{p}(t,T)$, in order to apply an HJM-approach, we need to introduce corresponding forward rate processes $f^T(t)$ and $\bar{f}^T(t)$ that lead to a forward rate spread expressed as $g^T(t) := \bar{f}^T(t) - f^T(t)$. One then also obtains corresponding short rates and a short rate spread, namely $r_t = f^t(t)$, $\bar{r}_t = \bar{f}^t(t)$, $s_t = g^t(t) = \bar{r}_t - r_t$. Notice that a consistent model should lead to $\bar{p}(t,T) \leq p(t,T)$, which implies $\bar{f}^T(t) \geq f^T(t)$ or, equivalently $g^T(t) \geq 0 \ \forall t < T \leq \bar{T}$, where $\bar{T}$ is a given maximal maturity.

An extensive study within the multicurve HJM approach has appeared in Crépey et al. (2012). The driving random process is a Levy and a corresponding *HJM drift condition* is derived. Conditions are given for the non-negativity of rates and spreads; explicit formulas are obtained for various interest rate derivatives. What may not be fully satisfactory in Crépey et al. (2012) is that:

(i) some difficulties arise when dealing not only with credit risk, but also other risks such as liquidity. In particular, when looking for a condition that corresponds to the *defaultable HJM drift condition*;
(ii) a fictitious default has to be considered explicitly *(with pre default bond prices)*.

The study in Crépey et al. (2012) is continued in the recent paper Crépey et al. (2013b) with the main purpose of taking into account also counterparty risk and funding costs and of determining various *valuation adjustments* on top of the clean prices. The methodology in Crépey et al. (2013b) is again based on an HJM approach, but with explicit ingredients for the induced short rate models in order to obtain a Markovian structure and to be able to actually perform the value adjustment calculations. In particular, the authors in Crépey et al. (2013b) use a Levy Hull & White extended Vasicek model for r_t and introduce an additional factor that can be interpreted as representing a *short rate spread*. In this latter sense it becomes analogous to the approach to be presented here.

Another HJM-based approach, limited to default risk, appears in Chiarella et al. (2007) with emphasis on obtaining Markovian models with state dependent volatilities. The driving processes are of the jump-diffusion type. The difficulties here appear to be given by the fact that, for convenient specifications of the volatilities, one obtains deterministic short rate spreads. For more general, stochastic volatilities the authors obtain only approximate Markovianity. These difficulties have been overcome in the subsequent paper Chiarella et al. (2010), where the authors obtain finite-dimensional Markovian realizations also with stochastic spreads and, in addition, obtain a correlation structure between credit spread, interest rate and the stochastic volatility. When trying to extend their approach to a multi curve setting, beyond that implied by credit risk alone, there appear though some computational difficulties due to the stochastic volatility.

Before coming now to describing our short rate model, we recall some basics concerning FRAs. We start from the

Definition 2.1 A FRA (forward rate agreement) is an OTC derivative that allows the holder to lock in at $t < T$ the interest rate between the inception date T and the maturity $T + \Delta$ at a fixed value K. At maturity $T + \Delta$, a payment based on K is made and one based on $\bar{L}(T; T, T + \Delta)$ is received.

We shall denote the value of the FRA at $t < T$ by $FRA^T(t, K)$. In our two-curve risky setup, the fair price of a FRA in $t < T$ with fixed rate K and notional N is

$$\begin{aligned} FRA^T(t, K) &= N\Delta p(t, T + \Delta) E^{T+\Delta}\left[\bar{L}(T; T, T + \Delta) - K \mid \mathscr{F}_t\right] \\ &= N p(t, T + \Delta) E^{T+\Delta}\left[\tfrac{1}{\bar{p}(T, T+\Delta)} - (1 + \Delta K) \mid \mathscr{F}_t\right] \end{aligned} \tag{4}$$

where $E^{T+\Delta}$ denotes expectation under the $(T + \Delta)-$ forward measure $Q^{T+\Delta}$. Notice that the simultaneous presence of $p(t, T + \Delta)$ and $\bar{p}(t, T + \Delta)$ does not allow for the convenient reduction of the formula to a simpler form as in the one-curve setup.

2.2 Description of the Model Itself

For the *short-rate model approach* we shall have to start by modeling directly the short rate r_t and the short rate spread s_t and we do it under the standard martingale measure Q (to be calibrated to the market) for the risk-free money market account as numeraire. In order to account for a possible (negative) correlation between r_t and s_t we introduce a *factor model*: given three independent affine factor processes $\Psi_t^i,\ i = 1, 2, 3$ let

$$\begin{cases} r_t = \Psi_t^2 - \Psi_t^1 \\ s_t = \kappa \Psi_t^1 + \Psi_t^3 \end{cases} \tag{5}$$

where κ is a constant that measures the instantaneous correlation between r_t and s_t (negative correlation for $\kappa > 0$). This setup could be generalized in various ways, in particular by using more factors to drive s_t. In view of the existing literature one could, instead of using an affine model structure as we do it here, consider e.g. ambit-type processes as presented in Corcuera, Farkas, Schoutens and Valkeila (2013). Such a model, which is not of the semimartingale type, allows also for analytical computations and gives the possibility to take into account long-range dependence. Remaining within the pure credit risk setting where, see the comment after (2), the spread is given by the default intensity, some of the factors affecting the spread could be given a specific meaning as in Douady and Jeanblanc (2002) where, using an HJM-type approach, the authors consider a spread field process with one of the variables representing the rating of the issuer. The approach in Douady and Jeanblanc (2002) could possibly be generalized also to the present setting.

A common approach to modeling the factors in an affine context is to assume them of the type of a square root diffusion. This guarantees positivity of the spread, but the negative correlation comes at the expense of possibly negative interest rates (even if only with small probability). With such a model, by passing to the $(T + \Delta)-$forward measure, one can compute the value of a FRA and of the fair fixed rate.

For various reasons, in particular in view of our main goal to obtain an *adjustment factor*, it is convenient to be able to have the same factor model for FRAs with different maturities. We therefore aim at performing the calculations under a single reference measure, namely the standard martingale measure Q. More precisely, for the factor processes we assume the following affine diffusions under Q that are of the Vasicek type, namely

$$\begin{cases} d\Psi_t^1 = (a^1 - b^1 \Psi_t^1)dt + \sigma^1 \, dw_t^1 \\ d\Psi_t^i = (a^i - b^i \Psi_t^i)dt + \sigma^i \sqrt{\Psi_t^i} \, dw_t^i, \quad i = 2, 3 \end{cases} \tag{6}$$

where a^i, b^i, σ^i are positive constants with $a^i \geq (\sigma^i)^2/2$ for $i = 2, 3$, and w_t^i independent Wiener processes. We have chosen a Vasicek-type model for simplicity, but the results below can be easily extended to the Hull and White version of the

Vasicek model. Notice that the factor Ψ_t^1 may take negative values implying that, not only r_t, but also s_t may become negative (see however later under "comments on the main result"). Results completely analogous to those that we shall obtain here for the above pure diffusion model may be derived also for affine jump-diffusions at the sole expense of more complicated notation.

3 Main Result (FRAs)

3.1 Preliminary Notions and Results

Recalling the expression for a FRA under the forward measure, namely

$$\mathrm{FRA}^T(t,K) = Np(t,T+\Delta)E^{T+\Delta}\left[\frac{1}{\bar{p}(T,T+\Delta)} - (1+\Delta K) \mid \mathscr{F}_t\right], \tag{7}$$

one has that the crucial quantity to compute is

$$\bar{v}_{t,T} := E^{T+\Delta}\left[\frac{1}{\bar{p}(T,T+\Delta)} \mid \mathscr{F}_t\right] \tag{8}$$

and that the fixed rate to make the FRA a fair contract at time t is

$$\bar{K}_t := \frac{1}{\Delta}(\bar{v}_{t,T} - 1) \tag{9}$$

In the classical single curve case we have instead

$$v_{t,T} := E^{T+\Delta}\left[\frac{1}{p(T,T+\Delta)} \mid \mathscr{F}_t\right] = \frac{p(t,T)}{p(t,T+\Delta)} \tag{10}$$

being $\frac{p(t,T)}{p(t,T+\Delta)}$ an $\mathscr{F}_t$—martingale under the $(T+\Delta)$—forward measure. The fair fixed rate in the single curve case is then

$$K_t = \frac{1}{\Delta}\left(v_{t,T} - 1\right) = \frac{1}{\Delta}\left(\frac{p(t,T)}{p(t,T+\Delta)} - 1\right) \tag{11}$$

and notice that, in order to compute K_t, no interest rate model is needed (contrary to $\bar{K}_t$).

Due to the *affine dynamics* of Ψ_t^i $(i = 1, 2, 3)$ under Q, we have for the risk-free bond

$$\begin{aligned} p(t,T) &= E^Q\left\{ \exp\left[-\int_t^T r_u du \right] \mid \mathscr{F}_t \right\} = E^Q\left\{ \exp\left[\int_t^T (\Psi_u^1 - \Psi_u^2) du \right] \mid \mathscr{F}_t \right\} \\ &= \exp\left[A(t,T) - B^1(t,T)\Psi_t^1 - B^2(t,T)\Psi_t^2 \right] \end{aligned} \tag{12}$$

The coefficients satisfy

$$\begin{cases} B_t^1 - b^1 B^1 - 1 = 0, & B^1(T,T) = 0 \\ B_t^2 - b^2 B^2 - \frac{(\sigma^2)^2}{2}(B^2)^2 + 1 = 0, & B^2(T,T) = 0 \\ A_t = a^1 B^1 - \frac{(\sigma^1)^2}{2}(B^1)^2 + a^2 B^2\,, & A(T,T) = 0 \end{cases} \tag{13}$$

leading, in particular, to

$$B^1(t,T) = \frac{1}{b^1}\left(e^{-b^1(T-t)} - 1 \right). \tag{14}$$

For the *risky bond* we have instead

$$\begin{aligned} \bar{p}(t,T) &= E^Q\left\{ \exp\left[-\int_t^T (r_u + s_u) du \right] \mid \mathscr{F}_t \right\} \\ &= E^Q\left\{ \exp\left[-\int_t^T ((\kappa - 1)\Psi_u^1 + \Psi_u^2 + \Psi_u^3) du \right] \mid \mathscr{F}_t \right\} \\ &= \exp\left[\bar{A}(t,T) - \bar{B}^1(t,T)\Psi_t^1 - \bar{B}^2(t,T)\Psi_t^2 - \bar{B}^3(t,T)\Psi_t^3 \right] \end{aligned} \tag{15}$$

This time the coefficients satisfy

$$\begin{cases} \bar{B}_t^1 - b^1 \bar{B}^1 + (\kappa - 1) = 0, & \bar{B}^1(T,T) = 0 \\ \bar{B}_t^2 - b^2 \bar{B}^2 - \frac{(\sigma^2)^2}{2}(\bar{B}^2)^2 + 1 = 0, & \bar{B}^2(T,T) = 0 \\ \bar{B}_t^3 - b^3 \bar{B}^3 - \frac{(\sigma^3)^2}{2}(\bar{B}^3)^2 + 1 = 0, & \bar{B}^3(T,T) = 0 \\ \bar{A}_t = a^1 \bar{B}^1 - \frac{(\sigma^1)^2}{2}(\bar{B}^1)^2 + a^2 \bar{B}^2 + a^3 \bar{B}^3\,, & \bar{A}(T,T) = 0 \end{cases} \tag{16}$$

leading, in particular, to

$$\bar{B}^1(t,T) = \frac{1-\kappa}{b^1}\left(e^{-b^1(T-t)} - 1 \right) = (1-\kappa)\, B^1(t,T) \tag{17}$$

From the above 1st order equations it follows that

$$\begin{cases} \bar{B}^1(t,T) = (1-\kappa)\, B^1(t,T) \\ \bar{B}^2(t,T) = B^2(t,T) \\ \bar{A}(t,T) \;\; = A(t,T) - a^1\kappa \int_t^T B^1(u,T)du \\ \qquad\qquad + \frac{(\sigma^1)^2}{2}\kappa^2 \int_t^T (B^1(u,T))^2 du - (\sigma^1)^2\kappa \int_t^T B^1(u,T)du \\ \qquad\qquad - a^3 \int_t^T \bar{B}^3(u,T)du \end{cases} \tag{18}$$

Letting then

$$\tilde{A}(t,T) := \bar{A}(t,T) - A(t,T) \tag{19}$$

we obtain

$$\begin{aligned} \bar{p}(t,T) = &\exp\Big[\bar{A}(t,T) - B^1(t,T)\Psi_t^1 - B^2(t,T)\Psi_t^2 \\ &- \bar{B}^3(t,T)\Psi_t^3 + \kappa B^1(t,T)\Psi_t^1\Big] \\ = &p(t,T)\, \exp\Big[\tilde{A}(t,T) + \kappa B^1(t,T)\Psi_t^1 - \bar{B}^3(t,T)\Psi_t^3\Big] \end{aligned} \tag{20}$$

so that, putting for simplicity $\tilde{B}^1 := B^1(T, T+\Delta)$, one may write

$$\frac{p(T,T+\Delta)}{\bar{p}(T,T+\Delta)} = \exp\Big[-\tilde{A}(T,T+\Delta) - \kappa\tilde{B}^1\Psi_T^1 + \bar{B}^3(T,T+\Delta)\Psi_T^3\Big]. \tag{21}$$

3.2 The Result Itself

We introduce the

Definition 3.1 We call adjustment factor the process

$$\mathrm{Ad}_t^{T,\Delta} := E^Q\left\{\frac{p(T,T+\Delta)}{\bar{p}(T,T+\Delta)} \mid \mathscr{F}_t\right\}, \tag{22}$$

and shall prove the following

Proposition 3.1 We have

$$\bar{\nu}_{t,T} = \nu_{t,T} \cdot \mathrm{Ad}_t^{T,\Delta} \cdot \exp\left[\kappa \frac{(\sigma^1)^2}{2(b^1)^3}\left(1 - e^{-b^1\Delta}\right)\left(1 - e^{-b^1(T-t)}\right)^2\right] \tag{23}$$

with two adjustment factors on the right, of which the first one can be expressed as

$$\begin{aligned} \mathrm{Ad}_t^{T,\Delta} &= e^{-\tilde{A}(T,T+\Delta)} E^Q\left\{e^{-\kappa \tilde{B}^1 \Psi_T^1 + \tilde{B}^3(T,T+\Delta)\Psi_T^3} \mid \mathscr{F}_t\right\} \\ &:= A(\theta, \kappa, \Psi_t^1, \Psi_t^3) \end{aligned} \tag{24}$$

with $\theta := (a^i, b^i, \sigma^i,\ i = 1, 2, 3)$.

One may notice the analogy here with the multiplicative forward basis in Bianchetti (2012).

As a consequence of the previous proposition we have the following relation between the fair value $\bar{K}_t$ of the fixed rate in an actual FRA and the fair value K_t in a corresponding riskless one:

Corollary 3.1 The following relationship holds

$$\bar{K}_t = \left(K_t + \frac{1}{\Delta}\right) \cdot \mathrm{Ad}_t^{T,\Delta} \cdot \exp\left[\kappa \frac{(\sigma^1)^2}{2(b^1)^3}\left(1 - e^{-b^1\Delta}\right)\left(1 - e^{-b^1(T-t)}\right)^2\right] - \frac{1}{\Delta} \tag{25}$$

Notice that the factor given by the exponential is equal to 1 for zero correlation, i.e. for $(\kappa = 0)$.

3.3 Comments on the Main Result

3.3.1 Comments Concerning the Adjustment Factors

An easy intuitive interpretation of the main result can be obtained in the case of $\kappa = 0$ (independence of r_t and s_t): in this case we have $r_t + s_t > r_t$ implying $\bar{p}(T, T+\Delta) < p(T, T+\Delta)$ so that $Ad_t^{T,\Delta} \geq 1$ (the exponential adjustment factor is equal to 1). As expected, from Proposition 3.1 and Corollary 3.1 it then follows that

$$\bar{\nu}_{t,T} \geq \nu_{t,T}, \quad \bar{K}_t \geq K_t \tag{26}$$

To gain some intuition for the cases when $\kappa \neq 0$, let $\bar{p}^\kappa(t,T)$, $\bar{\nu}_{t,T}^\kappa$, $Ad_t^{T,\Delta,\kappa}$ denote the given quantities by stressing that the correlation parameter has value κ. Notice that $p(t,T)$ and thus also $\nu_{t,T}$ do not depend on κ. Consider then the case $\kappa > 0$, which is the standard case implying negative correlation between r_t and s_t. (The case $\kappa < 0$ is analogous/dual). For illustrative purposes we distinguish between the two events $\{\Psi_t^1 > 0,\ \forall t \in [T, T+\Delta]\}$, $\{\Psi_t^1 < 0,\ \forall t \in [T, T+\Delta]\}$ where the latter occurs only with small probability (in reality, Ψ_t^1 will be positive for certain values of t and negative for the remaining ones).

On $\{\Psi_t^1 > 0,\ t \in [T, T+\Delta]\}$ we now have

$$\begin{aligned} &\bar{p}^{\kappa}(T, T+\Delta) < \bar{p}^{0}(T, T+\Delta) \\ &\Rightarrow \bar{\nu}_{t,T}^{\kappa} > \bar{\nu}_{t,T}^{0} \quad \Rightarrow \quad \bar{\nu}_{t,T}^{\kappa}/\nu_{t,T} > \bar{\nu}_{t,T}^{0}/\nu_{t,T} \end{aligned} \tag{27}$$

Recalling then

$$\bar{\nu}_{t,T}^{\kappa} = \nu_{t,T} \cdot Ad_t^{T,\Delta,\kappa} \cdot \exp\left[\kappa \frac{(\sigma^1)^2}{2(b^1)^3}\left(1 - e^{-b^1\Delta}\right)\left(1 - e^{-b^1(T-t)}\right)^2\right] \tag{28}$$

the last inequality in (27) can be seen to be in line with the fact that, in this case, in (28) the exponential factor is >1 and $Ad_t^{T,\Delta,\kappa} > Ad_t^{T,\Delta,0}$ (recall Definition 3.1).

On the other hand, on $\{\Psi_t^1 < 0,\ t \in [T, T+\Delta]\}$, we have

$$\bar{p}^{\kappa}(T, T+\Delta) > \bar{p}^{0}(T, T+\Delta) \quad \Rightarrow \quad \bar{\nu}_{t,T}^{\kappa}/\nu_{t,T} < \bar{\nu}_{t,T}^{0}/\nu_{t,T} \tag{29}$$

This inequality can be seen to be in line with the fact that, here, $Ad_t^{T,\Delta,\kappa} < Ad_t^{T,\Delta,0}$, but the exponential factor is still >1. This can nevertheless be explained by noticing that, in this case, r_t is relatively large and $r_t + s_t$ is closer to r_t (may be even $<r_t$). This implies a push of $\bar{\nu}_{t,T}^{\kappa}/\nu_{t,T}$ towards smaller values than in the previous case.

3.3.2 Comments Concerning the Use of the Results for Calibration

For what concerns calibration of our model to FRA and other available market data, notice that the coefficients $a^1, a^2, b^1, b^2, \sigma^1, \sigma^2$ can be calibrated in the usual way on the basis of the observations of default-free bonds $p(t, T)$ (if we had a Hull & White extension of our Vasicek-type model (6) then also for this model the calibration could be performed as in the standard case). To calibrate a^3, b^3, σ^3, notice that, contrary to $p(t, T)$, the "risky" bonds $\bar{p}(t, T)$ are not observable (relation (3) does not imply a unique inverse relationship to determine $\bar{p}(t, T)$ from observations of the LIBORs). One can however observe $K_t = \frac{1}{\Delta}\left(\frac{p(t,T)}{p(t,T+\Delta)} - 1\right)$ as well as the "risky" FRA rate $\bar{K}_t$. Recalling then Corollary 3.1 and the fact that $Ad_t^{T,\Delta} = A(\theta, \kappa, \Psi_t^1, \Psi_t^3)$, notice that, having calibrated a^i, b^i, σ^i $(i = 1, 2)$, from the observations of K_t and $\bar{K}_t$ one could thus calibrate a^3, b^3, σ^3 as well as κ. If there is a way to determine directly $Ad_t^{T,\Delta}$ (e.g. by observing the FRA rates for uncorrelated r_t and s_t), then the relationship between K_t and $\bar{K}_t$ as expressed in Corollary 3.1 would allow to calibrate separately κ. We furthermore recall that, as pointed out in Crépey et al. (2013b), calibration of clean prices is sufficient also when using the model to compute possible value adjustments.

3.4 Proof of the Main Result

Since the quantities of interest, namely $\bar{\nu}_{t,T}$ and $\nu_{t,T}$ were defined under the forward measure (see (8) and (10)), as a first step we perform a change from the forward measure $Q^{T+\Delta}$ to the standard martingale measure Q. To this effect, putting $b_t := \exp\left[\int_0^t r_u du\right]$, the density process for changing from Q to $Q^{T+\Delta}$ is $L_t = \frac{p(t,T+\Delta)}{p(0,T+\Delta)b_t}$. We can thus write

$$\begin{aligned}\bar{\nu}_{t,T} &= E^{T+\Delta}\left\{\tfrac{1}{\bar{p}(T,T+\Delta)} \mid \mathscr{F}_t\right\} = L_t^{-1} E^Q\left\{\tfrac{L_{T+\Delta}}{\bar{p}(T,T+\Delta)} \mid \mathscr{F}_t\right\} \\ &= \tfrac{1}{p(t,T+\Delta)} E^Q\left\{\exp[-\int_t^T r_u du]\, \tfrac{p(T,T+\Delta)}{\bar{p}(T,T+\Delta)} \mid \mathscr{F}_t\right\}\end{aligned} \tag{30}$$

Recalling the expression for $p(T, T+\Delta)/\bar{p}(T, T+\Delta)$ (see (21)) this becomes

$$\begin{aligned}\bar{\nu}_{t,T} =& \frac{1}{p(t,T+\Delta)} E^Q\left\{e^{-\int_t^T r_u du} \cdot \exp\left[-\tilde{A}(T,T+\Delta) - \kappa\tilde{B}^1\Psi_T^1 + \bar{B}^3(T,T+\Delta)\Psi_T^3\right] \mid \mathscr{F}_t\right\} \\ =& \frac{1}{p(t,T+\Delta)} \exp\left[-\tilde{A}(T,T+\Delta)\right] E^Q\left\{e^{\bar{B}^3(T,T+\Delta)\Psi_T^3} \mid \mathscr{F}_t\right\} \\ & \cdot E^Q\left\{e^{-\int_t^T(-\Psi_u^1+\Psi_u^2)du} e^{-\kappa\tilde{B}^1\Psi_T^1} \mid \mathscr{F}_t\right\}\end{aligned} \tag{31}$$

To proceed, consider the process F_t given by the last factor in (31), namely

$$F_t := E^Q\left\{e^{-\int_t^T(-\Psi_u^1+\Psi_u^2)du} e^{-\kappa\tilde{B}^1\Psi_T^1} \mid \mathscr{F}_t\right\} \tag{32}$$

Due to the affine dynamics of Ψ_t^i, $i = 1, 2$, and the independence of Ψ_t^1 and Ψ_t^2, we may write

$$\begin{aligned}F_t :=& E^Q\left\{e^{\int_t^T \Psi_u^1 du} e^{-\kappa\tilde{B}^1\Psi_T^1} \mid \mathscr{F}_t\right\} E^Q\left\{e^{-\int_t^T \Psi_u^2 du} \mid \mathscr{F}_t\right\} \\ =& \exp\left[\alpha^1(t,T) - \beta^1(t,T)\Psi_t^1\right] \exp\left[\alpha^2(t,T) - \beta^2(t,T)\Psi_t^2\right]\end{aligned} \tag{33}$$

where the coefficients satisfy

$$\begin{cases} \beta_t^1 - b^1\beta^1 - 1 = 0, & \beta^1(T,T) = \kappa\tilde{B}^1 \\ \beta_t^2 - b^2\beta^2 - \frac{(\sigma^2)^2}{2}(\beta^2)^2 + 1 = 0, & \beta^2(T,T) = 0 \\ \alpha_t^1 = -\frac{(\sigma^1)^2}{2}(\beta^1)^2 + a^1\beta^1, & \alpha^1(T,T) = 0 \\ \alpha_t^2 = a^2\beta^2, & \alpha^2(T,T) = 0 \end{cases} \tag{34}$$

Recalling also (12)–(14), the solutions of the system (34) can be expressed as

$$\begin{cases} \beta^1(t,T) = \frac{1}{b^1}\left[(b^1\kappa\tilde{B}^1 + 1)e^{-b^1(T-t)} - 1\right] = B^1(t,T) + \kappa\tilde{B}^1 e^{-b^1(T-t)} \\ \beta^2(t,T) = B^2(t,T) \\ \alpha^1(t,T) = \frac{(\sigma^1)^2}{2}\int_t^T(\beta^1(u,T))^2du - a^1\int_t^T\beta^1(u,T)du \\ \qquad\quad = \frac{(\sigma^1)^2}{2}\int_t^T(B^1(u,T))^2du - a^1\int_t^T B^1(u,T)du \\ \qquad\qquad + \frac{(\sigma^1)^2}{2}(\kappa\tilde{B}^1)^2\int_t^T e^{-2b^1(T-u)}du \\ \qquad\qquad + \kappa\tilde{B}^1(\sigma^1)^2\int_t^T B^1(u,T)e^{-b^1(T-u)}du - a^1\kappa\tilde{B}^1\int_t^T e^{-b^1(T-u)}du \\ \alpha^2(t,T) = -a^2\int_t^T B^2(u,T)du \end{cases} \tag{35}$$

Consequently

$$\begin{aligned} F_t = \exp\Big[&\frac{(\sigma^1)^2}{2}\int_t^T(B^1(u,T))^2du - a^1\int_t^T B^1(u,T)du \\ &-a^2\int_t^T B^2(u,T)du - B^1(t,T)\Psi_t^1 - B^2(t,T)\Psi_t^2\Big] \\ &\cdot\exp\left[\frac{(\sigma^1)^2}{2}(\kappa\tilde{B}^1)^2\int_t^T e^{-2b^1(T-u)}du - a^1\kappa\tilde{B}^1\int_t^T e^{-b^1(T-u)}du\right. \\ &\left.-\kappa\tilde{B}^1 e^{-b^1(T-t)}\Psi_t^1\right] \\ &\cdot\exp\left[\kappa\tilde{B}^1(\sigma^1)^2\int_t^T B^1(u,T)e^{-b^1(T-u)}du\right] \end{aligned} \tag{36}$$

$$\begin{aligned} = p(t,T)\cdot\exp&\left[\frac{(\sigma^1)^2}{2}(\kappa\tilde{B}^1)^2\int_t^T e^{-2b^1(T-u)}du - a^1\kappa\tilde{B}^1\int_t^T e^{-b^1(T-u)}du\right. \\ &\left.-\kappa\tilde{B}^1 e^{-b^1(T-t)}\Psi_t^1\right] \\ &\cdot\exp\left[\kappa\tilde{B}^1(\sigma^1)^2\int_t^T B^1(u,T)e^{-b^1(T-u)}du\right] \end{aligned}$$

On the other hand, recalling (21), one obtains

$$\begin{aligned} &E^Q\left\{\tfrac{p(T,T+\Delta)}{\bar{p}(T,T+\Delta)} \mid \mathscr{F}_t\right\} \\ &= e^{-\tilde{A}(T,T+\Delta)} E^Q\left\{e^{\bar{B}^3(T,T+\Delta)\Psi_T^3} \mid \mathscr{F}_t\right\} E^Q\left\{e^{-\kappa\tilde{B}^1\Psi_T^1} \mid \mathscr{F}_t\right\} \end{aligned} \tag{37}$$

where, due to the affine dynamics of Ψ_t^1, we may write

$$E^Q\left\{e^{-\kappa\tilde{B}^1\Psi_T^1} \mid \mathscr{F}_t\right\} = \exp\left[\bar{\alpha}(t,T) - \bar{\beta}(t,T)\Psi_t^1\right] \tag{38}$$

with $\bar{\alpha}(\cdot)$ and $\bar{\beta}(\cdot)$ satisfying

$$\begin{cases} \bar{\beta}_t - b^1\bar{\beta} = 0, & \bar{\beta}(T,T) = \kappa\tilde{B}^1 \\ \bar{\alpha}_t = a^1\bar{\beta} - \frac{(\sigma^1)^2}{2}(\bar{\beta})^2, & \bar{\alpha}(T,T) = 0 \end{cases} \tag{39}$$

so that

$$\begin{aligned} \bar{\beta}(t,T) &= \kappa\tilde{B}^1 e^{-b^1(T-t)} \\ \bar{\alpha}(t,T) &= -a^1\kappa\tilde{B}^1\int_t^T e^{-b^1(T-u)}du + \tfrac{(\sigma^1)^2}{2}(\kappa\tilde{B}^1)^2\int_t^T e^{-2b^1(T-u)}du \end{aligned} \tag{40}$$

and, consequently,

$$\begin{aligned} &E^Q\left\{e^{-\kappa\tilde{B}^1\Psi_T^1} \mid \mathscr{F}_t\right\} = \exp\left[-\kappa\tilde{B}^1 e^{-b^1(T-t)}\Psi_t^1\right] \\ &\exp\left[-a^1\kappa\tilde{B}^1\int_t^T e^{-b^1(T-u)}du + \tfrac{(\sigma^1)^2}{2}(\kappa\tilde{B}^1)^2\int_t^T e^{-2b^1(T-u)}du\right] \end{aligned} \tag{41}$$

Combining (31) with (36) as well as with (37) together with (41), we obtain

$$\begin{aligned} \bar{\nu}_{t,T} &= \tfrac{1}{p(t,T+\Delta)}\exp\left[-\tilde{A}(T,T+\Delta)\right] E^Q\left\{e^{\bar{B}^3(T,T+\Delta)\Psi_T^3} \mid \mathscr{F}_t\right\}\cdot F_t \\ &= \tfrac{p(t,T)}{p(t,T+\Delta)} E^Q\left\{\tfrac{p(T,T+\Delta)}{\bar{p}(T,T+\Delta)} \mid \mathscr{F}_t\right\} \\ &\quad\cdot\exp\left[\kappa(\sigma^1)^2\tilde{B}^1\int_t^T B^1(u,T)e^{-b^1(T-u)}du\right]. \end{aligned} \tag{42}$$

The result then follows noticing that

$$\tilde{B}^1\int_t^T B^1(u,T)e^{-b^1(T-u)}du = \frac{1}{2(b^1)^3}\left(1-e^{-b^1\Delta}\right)\left(1-e^{-b^1(T-t)}\right)^2. \tag{43}$$

4 Aspects of CAP Pricing

4.1 Preliminary Comments

This part is related to work in progress, but we want nevertheless to present some ideas on how our results obtained for FRAs (linear derivatives) can be extended to nonlinear derivatives. To discuss a specific case, we concentrate here on the pricing of a *single Caplet*, with strike K, maturity T on the spot LIBOR for the period $[T, T+\Delta]$. Using the forward measure $Q^{T+\Delta}$, its price in $t < T$ is then given by

$$\begin{aligned} Capl^{T,\Delta}(t) &= \Delta p(t, T+\Delta) E^{T+\Delta}\left\{\left(\bar{L}(T; T, T+\Delta) - K\right)^+ \mid \mathscr{F}_t\right\} \\ &= p(t, T+\Delta) E^{T+\Delta}\left\{\left(\frac{1}{\bar{p}(T, T+\Delta)} - \tilde{K}\right)^+ \mid \mathscr{F}_t\right\} \end{aligned} \tag{44}$$

with $\tilde{K} := 1 + \Delta K$.

As model, we may use the same " risky" short rate model as for the FRAs that we may consider as already calibrated (for the standard martingale measure Q). It may thus suffice to derive just a pricing algorithm that need not also be used for calibration. The most convenient way to price a Caplet is, as in (44) and as we do it below, to compute the expectations under the forward measure. Notice however that expectations with respect to a forward measure can be easily computed by performing a change to the standard martingale measure (see e.g. (30)), namely the one for which we may already have calibrated the model. Besides pricing, it may be desirable to obtain also here an "adjustment factor".

4.2 A Possible Pricing Methodology

For the pricing, in the forward measure, we may use *Fourier transform methods* as in Crépey et al. (2012) and Crépey et al. (2013b) thereby representing the claim as

$$\left(e^X - \bar{K}\right)^+ \quad \text{with} \quad X := -\log \bar{p}(T, T+\Delta) \tag{45}$$

We then need only to compute the *moment generating function* of X, which is a linear combination of the factors (this computation is feasible thanks to the affine structure) and use the Fourier transform of $f(x) = \left(e^x - \bar{K}\right)^+$, which is well-known.

Notice that one could possibly also apply a Gram-Charlier expansion as in Kijima et al. (2009).

With the Fourier transform method the price in $t = 0$ of the Caplet can then be obtained in the form (see Crépey et al. (2013b))

$$\text{Capl}(0,T,T+\Delta)=\frac{p(0,T+\Delta)}{2\pi}\int\frac{\tilde{K}^{1-iv-R}\bar{M}_X^{T+\Delta}(R+iv)}{(R+iv)(R+iv-1)}dv \tag{46}$$

where $\bar{M}_X^{T+\Delta}(\cdot)$ is the moment generating function of X under the $(T+\Delta)$—forward measure and R is such that $\bar{M}_X^{T+\Delta}(R+iv)$ is finite. This moment generating function can be computed for each of the various forward measures in terms of the Q—characteristics of the factors, analogously to the computations in Sect. 3.4 (see, in particular, (30)). From these computations one can also see that the Radon-Nikodym-derivative to change from Q to $Q^{T+\Delta}$ can in fact be expressed in explicit form and it preserves the affine structure, see Corollary 10.2 in Filipović (2009) (For a recent account on conditions for an absolutely continuous measure transformation to preserve the affine structure see Fontana and Montes (2014)).

If $M_X^{T+\Delta}(z)$ is the moment generating function of X with $p(T,T+\Delta)$ instead of $\bar{p}(T,T+\Delta)$, then

$$\bar{M}_X^{T+\Delta}(z)=M_X^{T+\Delta}(z)A(z;\theta,\kappa,\Psi_0^1,\Psi_0^2,\Psi_0^3) \tag{47}$$

where $A(z;\theta,\kappa,\Psi_0^1,\Psi_0^2,\Psi_0^3)=E^{T+\Delta}\left\{\left(M_X^{T+\Delta}(z)\right)^{-1}e^{zX}\right\}$. Now, from the expression for $p(t,T)$ in (12) we obtain

$$\begin{aligned}M_X^{T+\Delta}(z)&=E^{T+\Delta}\left\{e^{zX}\right\}=E^{T+\Delta}\left\{e^{-z\log p(T,T+\Delta)}\right\}\\&=E^{T+\Delta}\left\{\exp\left[-zA(T,T+\Delta)+zB^1(T,T+\Delta)\Psi_T^1+zB^2(T,T+\Delta)\Psi_T^2\right]\right\}\end{aligned} \tag{48}$$

On the other hand, from the expression for $\bar{p}(t,T)$ in (15) (see also the variant in (20), where the parameter κ appears explicitly) we obtain

$$\begin{aligned}A(z;\theta,\kappa,\Psi_0^1,\Psi_0^2,\Psi_0^3)&=\left(M_X^{T+\Delta}(z)\right)^{-1}E^{T+\Delta}\left\{e^{-z\log\bar{p}(T,T+\Delta)}\right\}\\&=E^{T+\Delta}\left\{\exp\left[-z\bar{A}(\cdot)+z\bar{B}^1(\cdot)\Psi_T^1+z\bar{B}^2(\cdot)\Psi_T^2+z\bar{B}^3(\cdot)\Psi_T^3\right]\right\}\end{aligned} \tag{49}$$

where $(\cdot)$ stands for $(T,T+\Delta)$. Given the affine nature of the factors, both expressions in (48) and (49) can be explicitly computed as a function of the parameters of the model and the initial values $\Psi_0^1,\Psi_0^2,\Psi_0^3$ of the factors, as expressed by the symbol $A(z;\theta,\kappa,\Psi_0^1,\Psi_0^2,\Psi_0^3)$. We may now consider $A(z;\theta,\kappa,\Psi_0^1,\Psi_0^2,\Psi_0^3)$ as adjustment factor for this nonlinear example given by the Caplets. It is not as explicit as the adjustment factor for the FRAs in (23) and (25) and we are presently working on obtaining a more explicit form also in this case.

Acknowledgments We are grateful to Giulio Miglietta as well as to Claudio Fontana and Zorana Grbac for very constructive comments.

References

Bianchetti, M. (2012). Two curves, one price: pricing and hedging interest rate derivatives, decoupling forwarding and discounting yield curves. arXiv:0905.2770v4

Brigo, D., Morini, M., & Pallavicini, A. (2013). *Counterparty credit risk, collateral and funding with pricing cases for all asset classes*. New York: Wiley, Forthcoming.

Chiarella, C., Nikitopoulos Sklibosios, C., & Schloegl, E. (2007). A Markovian defaultable term structure model with state dependent volatilities. *International Journal of Theoretical and Applied Finance*, *10*, 155–202.

Chiarella, C., Maina, S. C., Nikitopoulos & Sklibosios, C. (2010). *Markovian defaultable HJM term structure models with unspanned stochastic volatility*. Quantitative Finance Research Centre Research Paper no. 283, Sydney: University of Technology.

Corcuera, J. M., Farkas, G., Schoutens, W., & Valkeila, E. (2013). A short rate model using ambit processes. In *Malliavin calculus and stochastic analysis: A festschrift in honor of David Nualart*. Springer proceedings in mathematics and statistics (Vol. 34, pp. 525–553). New York: Springer Science+Business Media.

Crépey, S., Gerboud, R., Grbac, Z., & Ngor, N. (2013a). Counterparty risk and funding: The four wings of the TVA. *International Journal of Theoretical and Applied Finance*, *16*(2), 1350006.

Crépey, S., Grbac, Z., & Nguyen, H.-N. (2012). A multiple-curve HJM model of interbank risk. *Mathematics and Financial Economics*, *6*, 155–190.

Crépey, S., Grbac, Z., Ngor, N., & Skovmand, D. (2013b). A Levy HJM multiple-curve model with application to CVA computation. (Preprint)

Douady, R., & Jeanblanc, M. (2002). A rating-based model for credit derivatives. *European Investment Review*, *1*, 17–29.

Filipović, D. (2009). In *Term structure models a graduate course*. Berlin: Springer Verlag.

Filipović, D., & Trolle, A. B. (2013). The term structure of interbank risk. *Journal of Financial Econometrics*, *109*, 707–733.

Fontana, C., & Montes, J. M. (2014). A unified approach to pricing and risk management of equity and credit risk. *Journal of Computational and Applied Mathematics*, *259*(B), 350–361.

Fuji, M., Shimada, Y., & Takahashi, A. (2009). *A note on the construction of multiple swap curves with and without collateral*. CARF Working Paper Series F-154.

Fuji, M., Shimada, Y., & Takahashi, A. (2011). A market model of interest rates with dynamic basis spreads in the presence of collateral and multiple currencies. *Wilmott Magazine*, *54*, 61–73.

Grbac, Z., Papapantoleon, A., Schoenmakers, J., & Skovmand, D. (2014). Affine LIBOR models with multiple curves: Theory, examples and calibration. (Preprint)

Keller-Ressel, M., Papapantoleon, A., & Teichmann, J. (2013). The Affine Libor Models. *Mathematical Finance*, *23*, 627–658.

Kenyon, C. (2010). Short-Rate Pricing After the Liquidity and Credit Shocks: Including the Basis. Retrieved August 18, 2010, from SSRN: http://ssrn.com/abstract=1558429 or http://dx.doi.org/10.2139/ssrn.1558429

Kijima, M., Tanaka, K., & Wong, T. (2009). A multi-quality model of interest rates. *Quantitative Finance*, *9*, 133–145.

Mercurio, F. (2010). Interest Rates and The Credit Crunch: New Formulas and Market Models. Bloomberg Portfolio Research Paper (2010–01).

Mercurio, F. (2010). LIBOR Market Models with Stochastic Basis. Bloomberg Education and Quantitative Research Paper (2010–05).

Pallavicini, A., & Brigo, D. (2013). Interest-Rate Modelling in Collateralized Markets: Multiple curves, credit-liquidity effects, CCPs. arXiv:1304.1397v1.

Piterbarg, V. (2010). Funding beyond discounting: collateral agreements and derivatives pricing. *Risk Magazine*, *24*, 97–102.

Pricing an American Call Under Stochastic Volatility and Interest Rates

Boda Kang and Gunter H. Meyer

1 Introduction

The option pricing goes back to the seminal paper of Black and Scholes (1973) which is pricing options under geometric Brownian motion (GBM) process. Since then a vast amount of literature discussed the problem of pricing both European and American options under GBM process.

However, derivative securities are commonly written on underlying assets with return dynamics that are not sufficiently well described by the GBM process with constant volatility. There have been numerous efforts to develop alternative asset return models that are capable of capturing the leptokurtic features found in financial market data, and subsequently to use these models to develop option prices that better reflect the volatility smiles and skews found in market traded options. One of the classical ways to develop option pricing models that are capable of generating such behaviour is to allow the volatility to evolve stochastically, for instance according to the square-root process introduced by Heston (1993) for which there are a number of papers discussing the American option pricing problem using either the method of lines and finite difference methods by Chiarella et al. (2009) or the integral transform approach by Chiarella et al. (2010) and Adolfsson et al. (2013).

Since we consider the pricing of American-type options, the early exercise premium of the option depends on the cost of carry determined by interest rates. Consequently, the volatility of interest rates does affect the decision to exercise this option at any given point in time. Hence the American options of the type that we consider in this chapter are sensitive not only to the volatility of the underlying but also to the

B. Kang (✉)
Department of Mathematics, University of York, Heslington, York YO10 5DD, UK
e-mail: boda.kang@york.ac.uk

G. H. Meyer
School of Mathematics, Georgia Institute of Technology, Atlanta, GA, USA
e-mail: meyer@math.gatech.edu

R. Dieci et al. (eds.), *Nonlinear Economic Dynamics and Financial Modelling*,
DOI: 10.1007/978-3-319-07470-2_17,

risk-free interest rate, and this is the motivation for considering American options under stochastic volatility and stochastic interest rates.

To the best of our knowledge, some authors discuss the American option pricing problem under these dynamics. Boyarchenko and Levendorski (2013) formulate the option pricing problem by a PDE approach and they calculate the option prices with the help of an iteration method based on Wiener-Hopf factorisation. Medvedev and Scaillet (2010) introduce a new analytical approach. After using an explicit and intuitive proxy for the exercise rule, they derive tractable pricing formulae using a short-maturity asymptotic expansion. Depending on model parameters, this method can accurately price options with time-to-maturity up to several years. Chiarella and Kang (2013) discuss the problem of pricing American compound option under this dynamics using the sparse grid approach in which case it is hard to obtain a nice and smooth early exercise boundary.

All of the above papers on price American put options instead of call options. However, with non-zero dividend yield and a stochastic interest rate it is more challenging to compute accurate early exercise boundaries for a call because the continuation region can become large when the interest rate moves further away from the dividend yield.

Also, the way of handling the boundary conditions in Boyarchenko and Levendroski (2013), especially when $v = 0$ or/and $r = 0$, seems not appropriate from the point of view of either finance or mathematics. In this chapter, we provide a complete discussion of the proper boundary conditions which should be imposed on the partial differential equation the option price satisfies.

The remainder of the chapter is structured as follows. Section 2 outlines the model and proper boundary conditions for American option prices when the underlying asset follows stochastic volatility and stochastic interest rate dynamics. Section 3 describes a method of lines approach to find the prices, free boundaries and hedge ratios of the American call option. In Sect. 4 we outline the basic idea of the sparse grid approach and implement a combination technique on a sparse grid to find the price profile of the American option. A number of numerical examples that demonstrate the computational advantages of the method of lines approach are provided in Sect. 5 before we draw some conclusions in Sect. 6.

2 Problem Statement-American Call Option with Stochastic Volatility and Stochastic Interest Rates

2.1 Model Description

Let $C(S, v, r, t)$ denote the price of an American call option written on a stock of price S at time t with maturity time T and strike price K. The variables v and r denote the variance of the stock price return and the risk-free rate at time t, respectively.

Analogous to the setting in Heston (1993) with addition of a stochastic interest rate of the Cox-Ingersoll-Ross (CIR) type, the dynamics for the share price S under the risk neutral measure is governed by the stochastic differential equation (SDE) system[1]

$$dS = (r - q)Sdt + \sqrt{v}SdZ_1, \tag{1}$$

$$dv = \kappa_v(\theta_v - v)dt + \sigma_v\sqrt{v}dZ_2, \tag{2}$$

$$dr = \kappa_r(\theta_r - r)dt + \sigma_r\sqrt{r}dZ_3, \tag{3}$$

where Z_1, Z_2 and Z_3 are standard Wiener processes and $\mathbb{E}(dZ_i dZ_j) = \rho_{ij}dt, i = 1, 2; j = i + 1, \ldots, 3$ with $\mathbb{E}$ being the expectation operator under the risk neutral measure. In Eq. (1), r is the risk-free rate of interest and q is the continuously compounded dividend yield. In Eq. (2) the parameter σ_v is the so-called vol-of-vol (in fact, $\sigma_v^2 v$ is the variance of the variance process v). The parameters κ_v and θ_v are, respectively, the rate of mean reversion and long run variance of the process for the variance v. In Eq. (3) the parameter σ_r is the volatility of the interest rate process (in fact, $\sigma_r^2 r$ is the variance of the interest rate process r). The parameters κ_r and θ_r are, respectively, the rate of mean reversion and long run interest rate of the process for the interest rate r. These parameters are under the risk-neutral measure and are related to the corresponding quantities under the physical measure (that we denote as $\kappa_v^{\mathbb{P}}, \theta_v^{\mathbb{P}}, \kappa_r^{\mathbb{P}}$ and $\theta_r^{\mathbb{P}}$) by two parameters that appear in the market prices of both volatility risk and interest rate risk.[2] We are also able to write down the above system (1)−(3) using independent Wiener processes W_1, W_2 and W_3 so that,

$$\begin{pmatrix} dZ_1 \\ dZ_2 \\ dZ_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \rho_{12} & \sqrt{1-\rho_{12}^2} & 0 \\ \rho_{13} & \frac{\rho_{23}-\rho_{13}\rho_{12}}{\sqrt{1-\rho_{12}^2}} & \sqrt{1-\rho_{13}^2-\left(\frac{\rho_{23}-\rho_{13}\rho_{12}}{\sqrt{1-\rho_{12}^2}}\right)^2} \end{pmatrix} \begin{pmatrix} dW_1 \\ dW_2 \\ dW_3 \end{pmatrix}.$$

The price of an American call option under stochastic volatility and interest rate at time t, $C(S, v, r, t)$, can be formulated as the solution to a free boundary PDE problem. We need to solve the PDE for the value of the call option $C(S, v, r, t)$ given as

[1] Of course, since we are using a numerical technique we could in fact use more general processes for S and v. The choice of the Heston processes is driven partly by the fact that this has become a very traditional stochastic volatility model and partly because the transform methods do not easily handle the more general variance processes.

[2] In fact, if it is assumed that the market prices of risk associated with the uncertainty driving the variance process and the interest rate process have the form $\lambda_v\sqrt{v}$ and $\lambda_r\sqrt{r}$, respectively, where λ_v is a constant (this was the assumption in Heston (1993)) and λ_r is a constant. In addition $\kappa_v^{\mathbb{P}}, \theta_v^{\mathbb{P}}$ and $\kappa_r^{\mathbb{P}}, \theta_r^{\mathbb{P}}$ are the corresponding parameters under the physical measure. Then $\kappa_v = \kappa_v^{\mathbb{P}} + \lambda_v\sigma_v, \theta_v = \frac{\kappa_v^{\mathbb{P}}\theta_v^{\mathbb{P}}}{\kappa_v^{\mathbb{P}}+\lambda_v\sigma_v}$; $\kappa_r = \kappa_r^{\mathbb{P}} + \lambda_r\sigma_r, \theta_r = \frac{\kappa_r^{\mathbb{P}}\theta_r^{\mathbb{P}}}{\kappa_r^{\mathbb{P}}+\lambda_r\sigma_r}$.

$$\mathcal{K}C - rC + \frac{\partial C}{\partial t} = 0, \tag{4}$$

on the interval $0 \leq t \leq T$ and subject to the terminal condition

$$C(S, v, r, T) = (S - K)^+, \tag{5}$$

the free (early exercise) boundary condition

$$C(d(v, r, t), v, r, t) = d(v, r, t) - K, \tag{6}$$

and the smooth-pasting conditions

$$\lim_{S \to d(v,r,t)} \frac{\partial C}{\partial S} = 1, \quad \lim_{S \to d(v,r,t)} \frac{\partial C}{\partial v} = 0, \quad \lim_{S \to d(v,r,t)} \frac{\partial C}{\partial r} = 0, \tag{7}$$

where $S = d(v, r, t)$ is the early exercise boundary for the call option at time t, variance v and interest rate r. The set $\{S : S > d(v, r, t)\}$ denotes the early exercise region at time t where the call assumes its intrinsic value

$$C(S, v, r, t) = S - K.$$

In Eq. (4) the Kolmogorov operator $\mathscr{K}$ is given as

$$\begin{aligned}\mathscr{K} = &\frac{vS^2}{2}\frac{\partial^2}{\partial S^2} + \frac{\sigma_v^2 v}{2}\frac{\partial^2}{\partial v^2} + \frac{\sigma_r^2 r}{2}\frac{\partial^2}{\partial r^2} + \rho_{12}\sigma_v v S \frac{\partial^2}{\partial S \partial v} + \rho_{13}\sigma_r \sqrt{rv} S \frac{\partial^2}{\partial S \partial r} \\ &+ \rho_{23}\sigma_v \sigma_r \sqrt{vr}\frac{\partial^2}{\partial v \partial r} + (\kappa_r(\theta_r - r) - \lambda_r r)\frac{\partial}{\partial r} + (r - q)\, S \frac{\partial}{\partial S} \\ &+ (\kappa_v(\theta_v - v) - \lambda_v v)\frac{\partial}{\partial v},\end{aligned} \tag{8}$$

where λ_v and λ_r are the constants appearing in the equation for the market prices of volatility risk and interest rate risk, which as stated in Footnote 2 are assumed to be of the form $\lambda_v\sqrt{v}$ and $\lambda_r\sqrt{r}$, respectively.

2.2 Boundary Conditions

Equation (4) is defined on $0 < S < d(v, r, t), 0 < v < \infty, 0 < r < \infty$ and $0 < t < T$ but will be solved numerically on a bounded domain. We shall restrict the "spatial" variables (S, v, r) to the set

$$D = \{(S, v, r) : 0 < S < S_{\max}, 0 < v < v_{\max}, 0 < r < r_{\max}\}.$$

In order to have a well-defined problem we need to impose boundary conditions on Eq. (4) at the boundary ∂D of D.

We shall assume that $S_{\max}$ is chosen so large that $d(v, r, t) < S_{\max}, \forall t \in (0, T]$ so that

$$C(S_{\max}, v, r, t) = S_{\max} - K. \tag{9}$$

At $S = 0, v = 0$ and $r = 0$ the parabolic Eq. (4) is degenerate which limits what boundary conditions can be set. Whether boundary conditions like Dirichlet data can be imposed, or the differential equation has to hold at points of degeneracy on ∂D is connected with the algebraic sign of the Fichera function for Eq. (4). A detailed exposition of the general theory may be found in Oleinik and Radkevic (1973), and its application to some pricing equations of finance is given in Meyer (2014). We shall adapt the exposition of Meyer (2014) for the American call considered here.

To simplify the arguments we consider Eq. (4) at a fixed t and think of $\frac{\partial C}{\partial t}$ as a known source term $f(S, v, r)$. Then the theory of Oleinik and Radkevic (1973) has to apply to the time independent (elliptic) equation

$$\mathscr{K} C - rC = f(S, v, r). \tag{10}$$

The matrix characterising Eq. (10) in (S, v, r)-space is

$$A = \frac{1}{2} \begin{pmatrix} vS^2 & \rho_{12}\sigma_v vS & \rho_{13}\sigma_r \sqrt{rv} S \\ \rho_{12}\sigma_v vS & \sigma_v^2 vS & \rho_{23}\sigma_v\sigma_r \sqrt{vr} \\ \rho_{13}\sigma_r \sqrt{rv} S & \rho_{23}\sigma_v\sigma_r \sqrt{vr} & \sigma_r^2 r \end{pmatrix}$$

At $S = 0$ the inward normal vector is $n = (1, 0, 0)^T$. At $v = 0$ we have $n = (0, 1, 0)^T$ and on $r = 0$ we have $n = (0, 0, 1)^T$. We see that

$$\langle An, n \rangle = 0$$

on these three faces of ∂D, where $\langle x, y \rangle$ denotes the dot product. According to Oleinik and Radkevic (1973) the Fichera theory applies. The Fichera function for Eq. (10) is

$$\begin{aligned} h(S, v, r) = &\left\{ (r - q)S - \frac{1}{2}\left[2vS + \rho_{12}\sigma_v S + \rho_{13}\sigma_r \frac{1}{2}\sqrt{\frac{v}{r}} S \right] \right\} n_1 \\ &+ \left\{ \kappa_v\theta_v - (\kappa_v + \lambda_v)v - \frac{1}{2}\left[\rho_{12}\sigma_v v + \sigma_v^2 + \rho_{23}\sigma_v\sigma_r \frac{1}{2}\sqrt{\frac{v}{r}} \right] \right\} n_2 \\ &+ \left\{ \kappa_r\theta_r - (\kappa_r + \lambda_r)r - \frac{1}{2}\left[\rho_{13}\sigma_r \sqrt{rv} + \sigma_r^2 + \rho_{23}\sigma_v\sigma_r \frac{1}{2}\sqrt{\frac{r}{v}} \right] \right\} n_3, \end{aligned}$$

hence

$$h(0, v, r) = 0,\ h(S, 0, r) = \kappa_v\theta_v - \frac{\sigma_v^2}{2},\ h(S, v, 0) = \kappa_r\theta_r - \frac{\sigma_r^2}{2}.$$

(i) The boundary condition at $S = 0$. According to Oleinik and Radkevic (1973) the pricing Eq. (4) should hold at $S = 0$. But if the spot price reaches 0 then according to Eq. (1) it will stay at 0 and the call will be worthless so that

$$C(0, v, r, t) = 0 \tag{11}$$

which is a solution of Eq. (4) when $S = 0$ provided that any additional boundary conditions imposed on Eq. (4) are consistent with Eq. (11). Looking ahead, the commonly chosen boundary condition

$$C_v(S, v_{\max}, r, t) = 0 \tag{12}$$

is consistent with Eq. (11) but, as simulations with the Heston model reported in Meyer (2014) indicate, may not be optimal compared to the Venttsel boundary conditions discussed below.
(ii) The boundary condition at $v = 0$ and $r = 0$. At $v = 0$ the pricing Eq. (4) reduces to

$$\frac{1}{2}\sigma_r^2 r\frac{\partial^2 C}{\partial r^2} + (r-q)S\frac{\partial C}{\partial S} + \kappa_v\theta_v\frac{\partial C}{\partial v} + (\kappa_r(\theta_r - r) - \lambda_r r)\frac{\partial C}{\partial r} - rC + \frac{\partial C}{\partial t} = 0. \tag{13}$$

If $h(S, 0, r) \geq 0$ then any solution of Eq. (4) with bounded derivatives would have to satisfy Eq. (13) at $v = 0$. However, a calibration of the Heston model may lead to parameters yielding $h(S, 0, r) \geq 0$ or $h(S, 0, r) < 0$ (see, e.g. Kjellin and Lovgren (2006)) and there does not seem to be any reason that the call price should change discontinuously with respect to κ_v, θ_v and σ_v as the Fichera function changes its algebraic sign. If one could impose Eq. (13) also when $h(S, 0, r) < 0$ one would reasonably expect the solution of Eq. (4) to be continuous with respect to the financial parameters as they cross the contour $h(S, 0, r, \kappa_v, \theta_v, \sigma_v) = 0$ for fixed S and r. The mathematical theory for well-posed parabolic problems cited in Meyer (2014), although not applicable in full technical detail, suggests that this is possible. Since the inward normal at $v = 0$ is $n = (0, 1, 0)^T$ we see that the convective terms of Eq. (13) satisfy

$$\langle((r - q)S, \kappa_v\theta_v, \kappa_r(\theta_r - r) - \lambda_r), n >= \kappa_v\theta_v\rangle 0.$$

This condition implies that Eq. (13) is an admissible (Venttsel) boundary condition for a non-degenerate diffusion equation regardless of the sign of the Fichera function. We shall assume that the problem remains well-posed when Eq. (13) is imposed on the degenerate diffusion Eq. (4) at $v = 0$.

An analogous argument shows that the restriction of Eq. (4) to $r = 0$, i.e.

$$\frac{vS^2}{2}\frac{\partial^2 C}{\partial S^2} + \frac{\sigma_v^2 v}{2}\frac{\partial^2 C}{\partial v^2} + \rho_{12}\sigma_v vS\frac{\partial^2 C}{\partial S \partial v} + (r-q)S\frac{\partial C}{\partial S} + (\kappa_v(\theta_v - v) - \lambda_v v)\frac{\partial C}{\partial v} + \kappa_r\theta_r\frac{\partial C}{\partial r} + \frac{\partial C}{\partial t} = 0, \tag{14}$$

likewise is an admissible boundary condition for all v, S and financial parameters.

The numerical treatment of the boundary conditions (13) and (14) in an iterative solution of the American call with finite differences is based on a very simple but apparently effective numerical approximation which entirely avoids solving these equations.

Suppose we choose uniform grids

$$0 = v_0 < v_1 < \cdots < v_M = v_{\max} \text{ with } \Delta v = v_{m+1} - v_m;$$

and

$$0 = r_0 < r_1 < \cdots < r_N = r_{\max} \text{ with } \Delta r = r_{n+1} - r_n;$$

and denote by C^k the numerical solution of Eq. (4) obtained in iteration k, We do not compute a solution of Eq. (13) at $v = 0$ and of Eq. (14) at $r = 0$ in iteration $k + 1$. Instead we impose the quadratic extrapolant of C^k through v_1, v_2 and v_3 as Dirichlet data at $v = 0$. Similarly, we impose the quadratic extrapolant through r_1, r_2 and r_3 as data at $r = 0$. For a smooth function $f(x)$ the quadratic extrapolant

$$f_e(0) = 3f(\Delta x) - 3f(2\Delta x) + f(3\Delta x)$$

satisfies

$$f_e(0) - f(0) = O(\Delta x^3)$$

and a Taylor expansion shows that

$$\frac{f(2\Delta x) - f_e(0)}{2\Delta x} = f'(\Delta x) + O(\Delta x^2), \quad \frac{f(2\Delta x) - 2f(\Delta x) + f_e(0)}{\Delta x^2} = f''(2\Delta x) + O(\Delta x^2).$$

It follows that a finite difference approximation of Eq. (4) at v_1 and r_1 with extrapolated values at $v = 0$ and $r = 0$ converges to a consistent approximation of Eqs. (13) and (14) as $v \to 0$ and $r \to 0$.

(iii) The boundary condition at $v = v_{\max}$. On the computational boundary $v = v_{\max}$ the Eq. (4) is not degenerate and the general theory cited in Meyer (2014) suggests that

$$\frac{vS^2}{2}\frac{\partial^2 C}{\partial S^2} + \frac{\sigma_r^2 r}{2}\frac{\partial^2 C}{\partial r^2} + \rho_{13}\sigma_r\sqrt{r v_{\max}}S\frac{\partial^2 C}{\partial S \partial r} + (r-q)S\frac{\partial C}{\partial S} + (\kappa_v(\theta_v - v_{\max}) - \lambda_v v_{\max})\frac{\partial C}{\partial v} + (\kappa_r(\theta_r - r) - \lambda_r r)\frac{\partial C}{\partial r} - rC + \frac{\partial C}{\partial t} = 0 \tag{15}$$

is an admissible (Venttsel) boundary condition for $C(S, v_{\max}, r, t)$ provided

$$\kappa_v(\theta_v - v_{\max}) - \lambda_v v_{\max} \leq 0.$$

Equation (15) is defined for $0 < S < d(v_{\max}, r, t)$, $0 < r < r_{\max}$, $t \in (0, T]$ and is subject to the initial condition

$$C(S, v_{\max}, r, T) = (S - K)^+$$

and the boundary conditions

$$C(0, v_{\max}, r, t) = 0,\ C(d(v_{\max}, r, t), v_{\max}, r, t) = d(v_{\max}, r, t) - K,$$
$$C_S(d(v_{\max}, r, t), v_{\max}, r, t) = 1.$$

At $r = 0$ we use the extrapolated solution described above. At $r = r_{\max}$ the pricing Eq. (15) without the C_{Sr} and C_{rr} terms provides an admissible Venttsel boundary condition for Eq. (15).

(iv) The boundary condition at $r = r_{\max}$. At $r = r_{\max}$ the (Venttsel) boundary condition

$$\frac{vS^2}{2}\frac{\partial^2 C}{\partial S^2} + \frac{\sigma_v^2 v}{2}\frac{\partial^2 C}{\partial v^2} + \rho_{12}\sigma_v v S\frac{\partial^2 C}{\partial S \partial v} + (r-q)S\frac{\partial C}{\partial S} + (\kappa_v(\theta_v - v) - \lambda_v v)\frac{\partial C}{\partial v} + (\kappa_r(\theta_r - r_{\max}) - \lambda_r r_{\max})\frac{\partial C}{\partial r} - rC + \frac{\partial C}{\partial t} = 0, \tag{16}$$

for $0 < S < d(v, r_{\max}, t)$, $0 < v_{\max} < v$ is imposed on $C(S, v, r_{\max}, t)$.

It is consistent with the well-posedness of the call (4) on the finite computational domain provided that

$$\kappa_r(\theta_r - r_{\max}) - \lambda_r r_{\max} < 0.$$

$C(S, v, r_{\max}, t)$ assumes its intrinsic value at $t = T$, the value matching and smooth pasting condition on the early exercise boundary $d(v, r_{\max}, t)$.

At $S = 0$ we set again $C(0, v, r_{\max}, t) = 0$ while at $v_{\max} = v$ we impose Eq. (16) without the $\frac{\partial^2 C}{\partial S \partial v}$ and $\frac{\partial^2 C}{\partial v^2}$ terms. We also note that at $(v, r) = (v_{\max}, r_{\max})$ we impose Equation (4) but retain only the $\frac{\partial^2 C}{\partial S^2}$ second derivative term.

In summary, we have for Eq. (4) a complete set of boundary conditions which allows us to compute $C(S, v, r, t)$ and $d(v, r, t)$ for $v \in [v_1, v_{\max}]$ and $r \in [r_1, r_{\max}]$. The boundary conditions are easy to implement simply by zeroing out terms in the full pricing Eq. (4) depending on $\{v_m, r_n\}$.

3 Method of Lines Implementation

In this section we discuss the implementation of the Method of Lines approach to price the American call option under stochastic volatility and interest rates by solving the PDE Eq. (4) subject to boundary conditions discussed above. The call was chosen as the more difficult pricing problem compared to a put because the early exercise boundary at expiration behaves like $\max\{K, Kr/q\}$ and can become large for large $r_{\max}$ and small q. The application of the method of lines to pricing problems in finance is discussed in detail in a forthcoming monograph (see Meyer (2014)).

The key idea behind the method of lines is to approximate the PDE with a system of ordinary differential equations (ODEs), whose solution is more readily obtained using numerical techniques. When volatility is constant, the system of ODEs is developed by discretising the time derivative. For the PDE (4), we must also discretise the derivative terms involving the variance, v and those involving the interest rate r, and to cater for the fact that the option price and free surface also depend on both the variance v and the interest rate r.

We begin by setting $v_m = m\Delta v, r_n = n\Delta r$, where $m = 0, 1, 2, \ldots, M, n = 0, 1, 2, \ldots, N$. Typically, we will set the maximum variance to be $v_M = 50\,\%$ and the maximum interest rate to be $r_N = 25\,\%$ as well. Furthermore, we discretise the time to expiry according to $\tau_l = l\Delta\tau$, where $\tau_L = T$. We denote the option price along the variance line v_m, the interet rate line r_n and time line τ_l by $C(S, v_m, r_n, \tau_l) \equiv C^l_{m,n}(S)$, and set

$$V(S, v_m, r_n, \tau_l) = \frac{\partial C(S, v_m, r_n, \tau_l)}{\partial S} \equiv V^l_{m,n}(S), \tag{17}$$

which is of course the option delta at the grid point.

We now select finite difference approximations for the derivative terms with respect to v and r. For the second order term, at the grid point (S, v_m, r_n, τ_l) we use the standard central difference scheme

$$\frac{\partial^2 C}{\partial v^2} = \frac{C^l_{m+1,n} - 2C^l_{m,n} + C^l_{m-1,n}}{(\Delta v)^2}, \quad \frac{\partial^2 C}{\partial r^2} = \frac{C^l_{m,n+1} - 2C^l_{m,n} + C^l_{m,n-1}}{(\Delta r)^2}, \tag{18}$$

and for the cross-derivative term at the grid point (S, v_m, r_n, τ_l) we use the central difference approximation

$$\frac{\partial^2 C}{\partial S \partial v} = \frac{V^l_{m+1,n} - V^l_{m-1,n}}{2\Delta v}, \quad \frac{\partial^2 C}{\partial S \partial r} = \frac{V^l_{m,n+1} - V^l_{m,n-1}}{2\Delta r}, \tag{19}$$

$$\frac{\partial^2 C}{\partial v \partial r} = \frac{C^l_{m+1,n+1} - C^l_{m+1,n-1} - C^l_{m-1,n+1} + C^l_{m-1,n-1}}{4\Delta v \Delta r}. \tag{20}$$

Numerical simulations with MOL suggest that upwinding of first order derivatives is not necessary. The results presented here were obtained with the central difference quotients

$$\frac{\partial C}{\partial v} = \frac{C^l_{m+1,n} - C^l_{m-1,n}}{2\Delta v}, \quad \frac{\partial C}{\partial r} = \frac{C^l_{m,n+1} - C^l_{m,n-1}}{2\Delta r}, \tag{21}$$

while backward quotients are used at $v_{\max}$ and $r_{\max}$.

Next we must select a discretisation for the time derivative. Initially we use a standard backward difference scheme, given at the grid point (S, v_m, r_n, τ_l) by

$$\frac{\partial C}{\partial \tau} = \frac{C^l_{m,n} - C^{l-1}_{m,n}}{\Delta \tau}. \tag{22}$$

This approximation is only first order accurate with respect to time. For the case of the standard American call option, Meyer and van der Hoek (1997) demonstrate that the accuracy of the method of lines increases considerably by using a second order approximation for the time derivative, specifically

$$\frac{\partial C}{\partial \tau} = \frac{3}{2}\frac{C^l_{m,n} - C^{l-1}_{m,n}}{\Delta \tau} - \frac{1}{2}\frac{C^{l-1}_{m,n} - C^{l-2}_{m,n}}{\Delta \tau}. \tag{23}$$

Thus we initiate the method of lines solution using Eq. (22) for the first several time steps, and then switching to Eq. (23) for all subsequent time steps.

Applying Eqs. (18)–(21) and the time discretisation to the PDE (4), we must now solve a system of second order ODEs at each time step, variance grid point and interest rate point. For the first three time steps, the ODE at the grid point $v = v_m$, $r = r_n$ and $\tau = \tau_l$ is

$$\begin{aligned}
&\frac{v_m S^2}{2}\frac{d^2 C^l_{m,n}}{dS^2} + \rho_{12}\sigma_v v_m S \frac{V^l_{m+1,n} - V^l_{m-1,n}}{2\Delta v} + \frac{\sigma_v^2 v_m}{2}\frac{C^l_{m+1,n} - 2C^l_{m,n} + C^n_{m-1,n}}{(\Delta v)^2} \\
&\quad + (\alpha_v - \beta_v v_m)\frac{C^l_{m+1,n} - C^l_{m-1,n}}{2\Delta v} + \rho_{13}\sigma_r\sqrt{v_m r_n} S \frac{V^l_{m,n+1} - V^l_{m,n-1}}{2\Delta r} \\
&\quad + \frac{\sigma_r^2 r_n}{2}\frac{C^l_{m,n+1} - 2C^l_{m,n} + C^n_{m,n-1}}{(\Delta r)^2} + (\alpha_r - \beta_r r_n)\frac{C^l_{m,n+1} - C^l_{m,n-1}}{2\Delta r} \\
&\quad + \rho_{23}\sigma_v\sigma_r\sqrt{v_m r_n}\frac{C^l_{m+1,n+1} - C^l_{m+1,n-1} - C^l_{m-1,n+1} + C^l_{m-1,n-1}}{4\Delta v \Delta r} \\
&\quad + (r - q)S\frac{dC^l_{m,n}}{dS} - rC^l_{m,n} - \frac{C^l_{m,n} - C^{l-1}_{m,n}}{\Delta \tau} = 0,
\end{aligned} \tag{24}$$

and for all subsequent time steps the ODE is

$$\frac{v_m S^2}{2}\frac{d^2 C^l_{m,n}}{dS^2} + \rho_{12}\sigma_v v_m S\frac{V^l_{m+1,n} - V^l_{m-1,n}}{2\Delta v} + \frac{\sigma_v^2 v_m}{2}\frac{C^l_{m+1,n} - 2C^l_{m,n} + C^n_{m-1,n}}{(\Delta v)^2}$$
$$+ (\alpha_v - \beta_v v_m)\frac{C^l_{m+1,n} - C^l_{m-1,n}}{2\Delta v} + \rho_{13}\sigma_r\sqrt{v_m r_n}S\frac{V^l_{m,n+1} - V^l_{m,n-1}}{2\Delta r}$$
$$+ \frac{\sigma_r^2 r_n}{2}\frac{C^l_{m,n+1} - 2C^l_{m,n} + C^n_{m,n-1}}{(\Delta r)^2} + (\alpha_r - \beta_r r_n)\frac{C^l_{m,n+1} - C^l_{m,n-1}}{2\Delta r}$$
$$+ \rho_{23}\sigma_v\sigma_r\sqrt{v_m r_n}\frac{C^l_{m+1,n+1} - C^l_{m+1,n-1} - C^l_{m-1,n+1} + C^l_{m-1,n-1}}{4\Delta v\Delta r}$$
$$+ (r-q)S\frac{dC^l_{m,n}}{dS} - rC^l_{m,n} - \frac{3}{2}\frac{C^l_{m,n} - C^{l-1}_{m,n}}{\Delta\tau} + \frac{1}{2}\frac{C^{l-1}_{m,n} - C^{l-2}_{m,n}}{\Delta\tau} = 0. \tag{25}$$

Equations (24) and (25) are solved for $m = 1, \ldots, M, n = 1, \ldots, N$ subject to

$$C^1_{m,n}(0) = 0,\ C^1_{m,n}(d^1_{m,n}) = d^1_{m,n} - K,\ \frac{dC^1_{m,n}}{dS}(d^1_{m,n}) = 1.$$

For $m = 0$ and $n = 0$ we have extrapolated values for $C^1_{0,n}$ and $C^1_{m,0}$ and for $m = M$ and $n = N$ selected terms in Eqs. (24) and (25) are dropped so that the boundary pricing equations of Sect. 2.2 are obtained. Finally, to avoid the degeneracy of the equations at $S = 0$, they are regularised with the substitution

$$\frac{v_m S^2}{2} \leftarrow \max\left\{10^{-5}, \frac{v_m S^2}{2}\right\}.$$

(Alternatively, one could solve them on $[S_0, S_{\max}]$ for some $S_0 > 0$.)

The system (24) and (25) is solved with a line Gauss Seidel iteration. Starting with an initial guess we solve in iteration k for each n the equations for $m = 1, \ldots, M$, using the latest available estimates for $C^l_{m+1,n}$, $C^l_{m-1,n}$, $C^l_{m,n+1}$, $C^l_{m,n-1}$, $V^l_{m+1,n}$, $V^l_{m-1,n}$, $V^l_{m,n+1}$ and $V^l_{m,n-1}$. The initial estimates for $C^l_{m,n}$ and $V^l_{m,n}$ are simply $C^{l-1}_{m,n}$ and $V^{l-1}_{m,n}$. Otherwise we use the latest estimates for $C^l_{m,n}$ and $V^l_{m,n}$ found during the current iteration through the variance lines. We iterate until the price profile converges to a desired level of accuracy. We then proceed to the next time step.

The generic first order form for Eqs. (24) and (25) is the system of two scalar equations

$$\frac{dC^l_{m,n}}{dS} = V^l_{m,n}, \tag{26}$$
$$\frac{dV^l_{m,n}}{dS} = A_{m,n}(S)C^l_{m,n} + B_{m,n}(S)V^l_{m,n} + P^l_{m,n}(S), \tag{27}$$

where $P^l_{m,n}(S)$ is a function of $C^l_{m+1,n}$, $C^l_{m-1,n}$,$C^l_{m,n+1}$, $C^l_{m,n-1}$, $V^l_{m+1,n}$, $V^l_{m-1,n}$, $V^l_{m,n+1}$, $V^l_{m,n-1}$, $C^{l-1}_{m,n}$, $C^{l-2}_{m,n}$. We solve Eqs. (26)–(27) using the Riccati transform.[3]

The Riccati transformation is given by

$$C^l_{m,n}(S) = R_{m,n}(S)V^l_{m,n}(S) + W^l_{m,n}(S), \tag{28}$$

where R and W are solutions to the initial value problems

$$\frac{dR_{m,n}}{dS} = 1 - B_{m,n}(S)R_{m,n}(S) - A_{m,n}(S)(R_{m,n}(S))^2, \quad R_{m,n}(0) = 0, \tag{29}$$

$$\frac{dW^l_{m,n}}{dS} = -A_{m,n}(S)R_{m,n}(S)W^l_{m,n} - R_{m,n}(S)P^l_{m,n}(S), \quad W^l_{m,n}(0) = 0, \tag{30}$$

and V is the solution to

$$\frac{dV^l_{m,n}}{dS} = A_{m,n}(S)(R(S)V + W^l_{m,n}(S)) + B_{m,n}(S)V + P^l_{m,n}(S), V^l_{m,n}(b^l_{m,n}) = 1, \tag{31}$$

where we denote the free boundary at grid point (v_m, r_n, τ_l) by $d(v_m, r_n, \tau_l) = d^l_{m,n}$.[4] Since $R_{m,n}$ is independent of τ, we begin by solving Eq. (29) and storing the solution. Next we solve Eq. (30) for increasing values of S, ranging from $0 < S < S_{\max}$, where we select $S_{\max}$ sufficiently large such that $S_{\max} > d^l_{m,n}$ will be guaranteed. We then step backward in S using the generated values of $R_{m,n}$ and $W^l_{m,n}$ until we encounter the value S^* such that[5]

$$S^* - K = R_{m,n}(S^*) + W^l_{m,n}(S^*), \tag{32}$$

and thus S^* is the value of the free boundary at grid point (v_m, r_n, τ_l).[6] Once $d^l_{m,n}$ has been determined we then solve Eq. (31) starting at $S = d^l_{m,n}$ and sweeping backward to $S = 0$. Finally we use the calculated values of $R_{m,n}$, $W^l_{m,n}$ and $V^l_{m,n}$ in Eq. (28) to determine the option price at each grid point along the variance lines at time to maturity τ_l.

An alternative way of solving the second order scalar equation (24) or (25) for fixed (m, n) is with a finite difference method on a discrete grid $\{S_i\}$. The resulting

[3] The Riccati transform basically replaces a given differential system (here Eqs. (26) and (27)) with an equivalent set of uncoupled equations of lower dimension (here Eqs. (29), (30) and (31) below).

[4] All ODEs have been solved by use of the implicit trapezoidal rule, discussed for example by Shampine (1994).

[5] We test Eq. (32) at each grid point and find the grid points at which $S - K - R_{m,n}(S) - W^l_{m,n}(S)$ changes sign. We then use Newton's method to search for the value of S^* by fitting a cubic spline through four points around of this point.

[6] We remind the reader that at S^* the first of the free boundary conditions (7) becomes $V^l_{m,n}(S^*) = 1$.

linear algebraic equation involves a tridiagonal matrix which can quickly be solved with an implementation of Gaussian elimination known as the Thomas algorithm. It is known that the Thomas algorithm converges to the Riccati transformation as $\max_i (S_{i+1} - S_i) \to 0$.

Loosely speaking, the Riccati transformation is the closure of Gaussian elimination as $\Delta S \to 0$ in the finite difference approximation of Eq. (24). Conversely, the numerical solution of the equations of the Riccati transformation corresponds to a finite difference solution for Eq. (24). Solving the free boundary problem for Eq. (24) or (25) with the trapezoidal rule applied to the Riccati transformation and searching for a root of Eq. (32) is roughly equivalent to the Brennan Schwartz method for Eq. (24) (for more details see Meyer (2014)).

4 Sparse Grid Implementation

In order to tackle the computationally demanding task of solving PDEs (4)−(6) with free boundary features, we can also apply the sparse grid approach that turns out to be quite fast and accurate. The sparse grid *combination technique* for solving PDEs was first introduced by Griebel, Schneider and Zenger (1992) after which Reisinger (2004) in his PhD thesis, Reisinger and Wittum (2007), Leentvaar and Oosterlee (2008) and Leentvaar and Oosterlee (2008) discussed the application of this approach to various option pricing problems. The combination technique requires the solution of the original equation only on a set of conventional subspaces defined on Cartesian grids specified in a certain way and a subsequent extrapolation step, but still retains a certain order convergence.

In fact we can identify three desirable properties of the combined solution. First of all, in comparison to the standard full grid approach the number of grid points can be reduced significantly from $O(2^{n \cdot d})$ to $O(2^n \cdot n^{d-1})$ at refinement level n in the d- dimensional case, whereas the point-wise accuracy of the approximation to the solution of the PDE is $O(n^{d-1} \cdot 2^{-n \cdot p})$ which is only slightly worse than $O(2^{-n \cdot p})$. Here, p includes the order of the underlying discretisation scheme, as well as the influence of singularities. Furthermore, each of the Cartesian grids setting up the sparse grid only consists of $O(2^n)$ nodes. Thus, the efficient usage of sparse grids for the computational solution of the PDE greatly reduces storage requirements and computing time at a moderate cost of accuracy.

Secondly, we have to point out the simplicity of the combination concept: we have seen that the sparse grid combined solution represents a linear combination of numerical solutions on Cartesian grids corresponding to the components of a sparse grid at the same refinement level. Thus, the combination technique allows for the integration of existing solvers for partial differential equations on traditional full grids. In contrast to the discretisation on a real sparse grid, which requires hierarchical data structures and thus specially designed solvers, the combined solution is built on simple data structures and can be based on any “black box solver”. Only the final linear combination of these simple solutions has to be newly implemented.

From the combined solution as a linear combination of traditional full gridbreak discretisations we can also deduce a further advantage of the combination technique. Since the $O(n^{d-1})$ problems solved on the Cartesian grids Ω_l that set up the sparse grids are independent of one another, these problems can be solved in parallel on different workstations. Communication has to take place only at the end, where the summation and the extrapolation by linear combination of the different solutions is performed.

4.1 The Sparse Grid Combination Technique

We incorporate the techniques and algorithms used in Chiarella et al. (2010) and the sparse grid approach to solve the linked PDE (4)−(6) with suitable initial, boundary conditions.

In a general $d-$ dimensional unit cube and the family of grids with grid sizes $h_j = 2^{-l_j}$ in direction $j, l_j \in \mathbb{N}_0$, we write the vector of grid sizes as $\mathbf{h} = 2^{-\mathbf{l}}$ with $\mathbf{l} = (l_1, \ldots, l_d) \in \mathbb{N}_0^d$ and denote the solution of the PDE on those grids by $c_{\mathbf{h}}$. The sparse grid solution at level l is then defined as

$$c_l = \sum_{k=l}^{l+d-1} a_{l-k} \sum_{l_1+\cdots+l_d=k} c_{\mathbf{h}}, \tag{33}$$

with

$$a_k = (-1)^{d-1-k} \binom{d-1}{k}, \ 0 \le k \le d-1. \tag{34}$$

In our case $d = 3$ hence we consider a truncated 3−dimensional cube $\Omega := [0, S_{\max}] \times [0, v_{\max}] \times [0, r_{\max}]$ and a Cartesian grid with mesh size $h_j = 2^{-l_j}$ (corresponding to a level $l_j \in \mathbb{N}_0$) in the directions $j = 1, 2, 3$. The indices $j = 1$, $j = 2$ and $j = 3$ represent the directions of the stock price S, the variance v and the interest rate r respectively.

For a vector $\mathbf{h} = (h_1, h_2, h_3)$ we denote by $c_{\mathbf{h}}$ the representation of a function on such a grid with points

$$\mathbf{x}_h = (i_1 \cdot h_1, \ i_2 \cdot h_2, \ i_3 \cdot h_3), \ 1 \le i_j \le N_j, \ N_j = 1/h_j = 2^{l_j}, \ \text{for } j = 1, 2, 3.$$

For a given level l, the above grid consists all possible combinations of (l_1, l_2, l_3) with $0 \le l_1, l_2, l_3 \le l$. Hence, in total, there are $(2^l + 1)^3$ points in the grid. The number of total points in the full grid increases significantly with the increase of the level l. It will be quite expensive to solve the two-pass PDE system on the above full grid. However, with the same level l, the *sparse* grid, will consist of the following points:

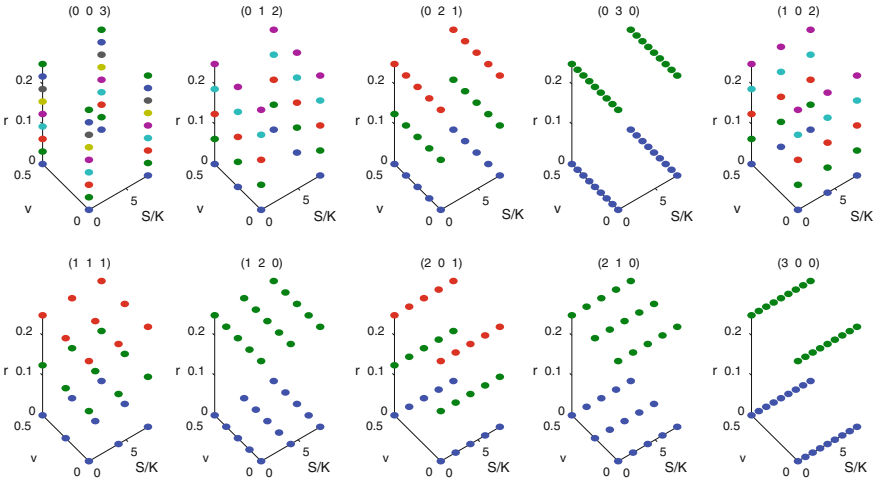

Fig. 1 A sparse grid hierarchy of level 3 with respect to each combination. From left to right and from top to bottom, these are $(0, 0, 3)$, $(0, 1, 2)$, $\ldots$, $(2, 1, 0)$, $(3, 0, 0)$ respectively

$$\mathbf{x}_h = (i_1 \cdot h_1,\ i_2 \cdot h_2,\ i_3 \cdot h_3), \quad 1 \le i_j \le N_j,\ N_j = 1/h_j = c_j 2^{l_j},\ \text{for } j = 1, 2, 3,$$

satisfying $l_1 + l_2 + l_3 = l$ and where c_j are some positive constants with the help of which it is possible to construct a non-equidistant grid.

It is not hard to see that there are $\binom{l+d-1}{d-1}$ choices of such combinations of (l_1, l_2, l_3) such that $l_1 + l_2 + l_3 = l$. Figures 1, 2 and 3 provide an example of a standard sparse grid hierarchy with level $l = 3$, $l = 2$ and $l = 1$ with respect to 10, 6 and 3 different combinations corresponding to each level respectively.

Obviously, the above grids share the common property that they are dense in one direction but sparse in the other directions.

Let $\mathbf{c}_h$ be the discrete vector of function values at the grid points of the standard sparse grid. In general, $\mathbf{c}_h$ is the finite difference solution to the PDE of interest on the corresponding grid $\mathbf{h}$. The solution can be extended to Ω by a suitable multi-linear interpolation operator $\mathscr{I}$[7] in the point wise sense according to

$$c_{\mathbf{h}}(S, v, r, \tau) = \mathscr{I}\mathbf{c}_h, \forall (S, v, r) \in \Omega.$$

Next, we define the family C of solutions corresponding to the different sparse grids (as in Fig. 1 for instance) by $C = (C(\mathbf{i}))_{\mathbf{i} \in \mathbb{N}^3}$ with

$$C(\mathbf{i}) := c_{2-\mathbf{i}},$$

[7] A thorough error analysis of the multi-linear interpolation operator can be found in Reisinger (2008) who gives a generic derivation for linear difference schemes through an error correction technique employing semi-discretisations and obtains error formulae as well.

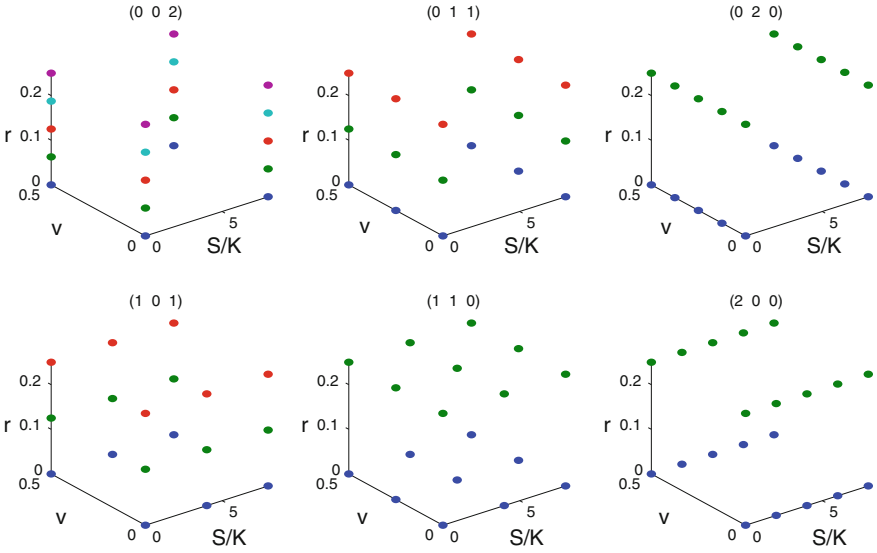

Fig. 2 A sparse grid hierarchy of level 2 with respect to each combination. From left to right and from top to bottom, these are $(0, 0, 2)$, $(0, 1, 1)$, $\ldots$, $(1, 1, 0)$, $(2, 0, 0)$ respectively

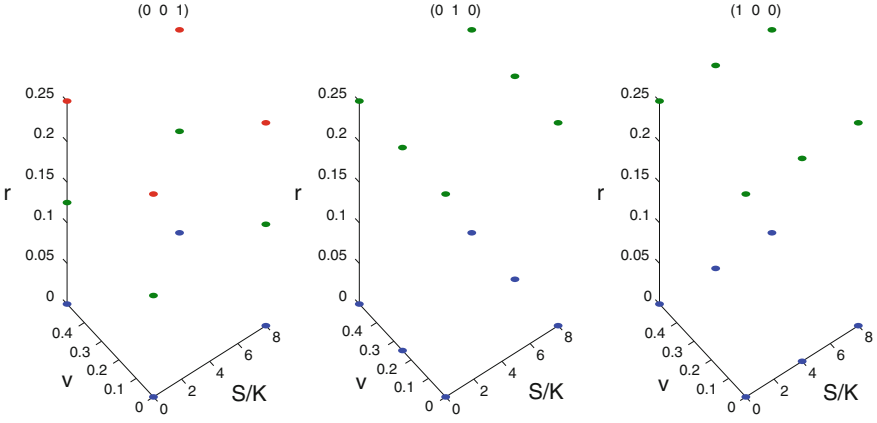

Fig. 3 A sparse grid hierarchy of level 1 with respect to each combination. From left to right, these are $(0, 0, 1)$, $(0, 1, 0)$, $(1, 0, 0)$ respectively

that is the family of numerical approximations (after proper interpolation) $c_{\mathbf{h}}$ on tensor product grids with $h_k = 2^{-i_k}$. For example, the solution on the first grid in Fig. 1 would be $C(0, 0, 3)$ etc. The combination technique in Reisinger and Wittum (2007) tells us that the solution c_l (l is the level of the sparse grid) of the corresponding PDE is

$$c_l = \sum_{l_1+l_2+l_3=l} C(l_1,l_2,l_3) - 2\cdot \sum_{l_1+l_2+l_3=l+1} C(l_1,l_2,l_3) + \sum_{l_1+l_2+l_3=l+2} C(l_1,l_2,l_3). \tag{35}$$

The above is a special case of Eqs. (33) and (34) when $d = 3$.

The procedure involves solving the PDE in parallel on each of the sparse grids of level l, $l+1$ and level $l+2$ respectively. See Figs. 1, 2 and 3 for $l = 1$ and $d = 3$ as an example. Thus we have

$$\sum_{k=l}^{l+2} \binom{k+d-1}{d-1} = \sum_{k=1}^{3} \binom{k+2}{2} = 3+6+10 = 19$$

which means that there will be 19 PDE solvers running simultaneously when $l = 1$. The theory developed by Reisinger and Wittum (2007) shows that Eq. (35) combines all solutions together to yield a more accurate solution to the PDE.

The essential principle of the extrapolation is that all lower order error terms cancel out in the combination formula (35) and only the highest order terms

$$h_1^2 \cdot h_2^2 \cdot h_3^2 = (2^{-l_1} \cdot 2^{-l_2} \cdot 2^{-l_3})^2 = 4^{-l}$$

remain. Taking advantage of this cancelation mechanism, Eq. (35) is able to produce quite accurate results fairly quickly. The details of the error analysis can be found in Reisinger (2004) and Reisinger and Wittum (2007).[8] We implement the above sparse grid combination technique to solve the PDE (4) in order to obtain the desired call option prices. We solve the PDEs (4)−(6) in each of the subspaces on a parallel cluster, which makes the process very efficient.

4.2 Finite Difference Method with PSOR

In the implementation, a standard Crank Nicolson finite difference method with the projected successive over-relaxation (PSOR) method has been applied to each of the sparse grids in Figs. 1, 2 and 3 to work out the solution of PDEs (4)−(6) on the grid points, solutions at other non-grid points are obtained by the same multi-linear interpolation as in Reisinger (2008). The implementation of PSOR is detailed in this section.

It is convenient to consider the time-to-maturity $\tau = T - t$ instead of time t. The three space variables S, v, r and time-to-maturity τ are discretised according to,

[8] The combination method requires, theoretically, smoothness of mixed derivatives of the solution. This is obviously not the case here due to the non-smooth payoff. However, if the payoff is aligned with the grid, which is the case in our problem, then good results have been observed for the combination method (see Leentvaar and Oosterlee (2008)). This is probably due to the rapid smoothing of the payoff in the first few time steps.

$$S_i = (i-1)\cdot\Delta S,\ i = 1,\ldots,N_1+1;\quad v_j = (j-1)\cdot\Delta v,\ j = 1,\ldots,N_2+1; \tag{36}$$

$$r_k = (k-1)\cdot\Delta r,\ k = 1,\ldots,N_3+1;\quad \tau_l = T - l\Delta t,\ l = 1,\ldots,N_\tau, \tag{37}$$

with N_1, N_2, N_3 and N_τ are the number of grid points in the direction S, v, r and τ respectively.

The option prices at the discrete points thus are

$$C^l_{i,j,k} = C(S_i, v_j, r_k, \tau_l).$$

Similar to the discussions in Ekstrom, Lotstedt and Tysk (2009), we use central differences to approximate the first derivative in S direction and the second derivatives in S, v, r directions in the PDE but use the forward and backward finite difference approximations on the boundaries other than the time derivative in Eq. (4).

Since the coefficients of the second order derivative terms go to zero as $v \to 0$ and $r \to 0$, we use an upwinding finite difference scheme for the first order derivative term, such that, at the grid point (S_i, v_j, r_k, τ_l) we have with $\alpha_v = \kappa_v\theta_v$, $\beta_v = \kappa_v + \lambda_v$ and $\alpha_r = \kappa_r\theta_r$, $\beta_r = \kappa_r + \lambda_r$

$$\frac{\partial C}{\partial v} = \begin{cases} \dfrac{C^l_{i,j+1,k} - C^l_{i,j,k}}{\Delta v} \text{ if } v \le \frac{\alpha_v}{\beta_v}, \\ \dfrac{C^l_{i,j,k} - C^l_{i,j-1,k}}{\Delta v} \text{ if } v > \frac{\alpha_v}{\beta_v}. \end{cases} \quad \frac{\partial C}{\partial r} = \begin{cases} \dfrac{C^l_{i,j,k+1} - C^l_{i,j,k}}{\Delta r} \text{ if } r \le \frac{\alpha_r}{\beta_r}, \\ \dfrac{C^l_{i,j,k} - C^l_{i,j,k-1}}{\Delta r} \text{ if } r > \frac{\alpha_r}{\beta_r}. \end{cases} \tag{38}$$

We treat two boundary conditions in the v directions simpler than those implemented in the method of lines. In particular, for large values of v, we set $\partial C/\partial v = 0$ along the variance boundary $v = v_M$ and when v is zero, we fit a quadratic polynomial through the option prices at v_1, v_2 and v_3, and then use this to extrapolate an approximation of the price at v_0. The boundary conditions for the interest rates, when $r = r_M$ and $r = 0$ are handled in a very similar way as those for variance v. This provides us with a satisfactory estimate of the price along v_0 or/and r_0 for the purpose of generating a stable solution for small values of v and r.

We follow Ikonen and Toivanen (2007) to indicate which grid point values we use to approximate the second order mixed-derivative in order to obtain non positive off-diagonal weights in the finite difference stencil, which makes the matrix an M-matrix as much as possible. In fact, to simplify the algorithm and take consideration of the correlations ρ_{ij}, in each two-dimensional space, we use a seven-point stencil and the mixed derivatives when $\rho_{12} \le 0$ are approximated as

$$\frac{\partial C^2}{\partial S \partial v} \approx \frac{1}{2}\Bigg(\frac{C^l_{i+1,j+1,k} - C^l_{i,j+1,k} - (C^l_{i+1,j,k} - C^l_{i,j,k})}{\Delta S \Delta v} + \frac{C^l_{i,j,k} - C^l_{i-1,j,k} - (C^l_{i,j-1,k} - C^l_{i-1,j-1,k})}{\Delta S \Delta v}\Bigg)$$

and when $\rho_{12} > 0$ we have

$$\frac{\partial C^2}{\partial S \partial v} \approx \frac{1}{2}\Bigg(\frac{C^l_{i,j+1,k} - C^l_{i-1,j+1,k} - (C^l_{i,j,k} - C^l_{i-1,j,k})}{\Delta S \Delta v} + \frac{C^l_{i+1,j,k} - C^l_{i,j,k} - (C^l_{i+1,j-1,k} - C^l_{i,j-1,k})}{\Delta S \Delta v}\Bigg).$$

The other mixed derivatives $\frac{\partial^2}{\partial S \partial r}$ and $\frac{\partial^2}{\partial v \partial r}$ are handled in a similar way depending on the sign of the corresponding correlations.

Boundary conditions at the boundaries $S = 0$ and $S = S_{\max}$ for a call option are

$$C^l_{1,j,k} = 0, \;\; C^l_{N_1+1,j,k} = S_{\max} - K, \;\; \forall j = 1, \ldots, N_2 + 1, k = 1, \ldots, N_3 + 1.$$

The spatial discretisation above leads to a semi-discrete equation which has the matrix representation

$$\frac{\partial \boldsymbol{C}}{\partial \tau} + \boldsymbol{A}\boldsymbol{C} = 0 \tag{39}$$

where $\boldsymbol{A}$ is a block tridiagonal $(N_1+1)(N_2+1)(N_3+1) \times (N_1+1)(N_2+1)(N_3+1)$ matrix and $\boldsymbol{C}$ is a vector of length $(N_1 + 1)(N_2 + 1)(N_3 + 1)$.

Next, we implement a more general θ—scheme which includes the implicit ($\theta = 1$), the Crank-Nicolson ($\theta = \frac{1}{2}$) and the explicit ($\theta = 0$) approaches to discretise the semi-discrete problem (39) as

$$(\boldsymbol{I} + \theta \Delta \tau \boldsymbol{A})\, \boldsymbol{C}^{(l+1)} = (\boldsymbol{I} - (1-\theta)\Delta \tau \boldsymbol{A})\, \boldsymbol{C}^{(l)}, \;\; l = 0, \ldots, N_\tau - 1, \tag{40}$$

where N_τ is the number of time steps and $\boldsymbol{I}$ is the identity matrix.

After the discretisation of the underlying PDE with three spatial variables an approximate price of an American option can be obtained by solving a sequence of linear complementarity problems (LCPs)

$$\begin{cases} \boldsymbol{B}\boldsymbol{C}^{(l+1)} \geq \boldsymbol{E}\boldsymbol{C}^{(l)}, \;\; \boldsymbol{C}^{(l+1)} \geq \boldsymbol{g}, \\ \left(\boldsymbol{B}\boldsymbol{C}^{(l+1)} - \boldsymbol{E}\boldsymbol{C}^{(l)}\right)^T \left(\boldsymbol{C}^{(l+1)} - \boldsymbol{g}\right) = 0, \end{cases} \tag{41}$$

for $l = 0, \ldots, N_\tau - 1$. The matrices $\boldsymbol{B}$ and $\boldsymbol{E}$ in Eq. (41) are defined by the left-hand and right-hand sides of Eq. (40) respectively. The initial value $\boldsymbol{C}^{(0)}$ is given by the

Table 1 Values of ω for different sparse grid are used for the American call option when $l = 6$

(l_1, l_2, l_3)	ω	(l_1, l_2, l_3)	ω	(l_1, l_2, l_3)	ω	(l_1, l_2, l_3)	ω	(l_1, l_2, l_3)	ω
(0, 0, 6)	1.6	(0, 1, 5)	1.3	(0, 2, 4)	1.1	(0, 3, 3)	1.1	(0, 4, 2)	1.3
(0, 5, 1)	1.6	(0, 6, 0)	1.8	(1, 0 ,5)	1.3	(1, 1, 4)	1.1	(1, 2, 3)	1.1
(1, 3, 2)	1.1	(1,4,1)	1.3	(1, 5, 0)	1.6	(2, 0, 4)	1.1	(2, 1, 3)	1.1
(2, 2, 2)	1.1	(2, 3, 1)	1.1	(4, 4, 0)	1.3	(3, 0, 3)	1.1	(3, 1, 2)	1.1
(3, 2, 1)	1.1	(3, 3, 0)	1.1	(4, 0, 2)	1.2	(4, 1, 1)	1.2	(4, 2, 0)	1.2
(5, 0, 1)	1.4	(5, 1, 0)	1.4	(6, 0 ,0)	1.7				

discrete form $\boldsymbol{g}$ of the payoff function g of the option, so that the ith element of $\boldsymbol{C}^{(0)}$ is given as

$$\boldsymbol{C}_i^{(0)} = \max(K - S_i, 0). \tag{42}$$

In order to avoid the oscillations that often occur with the CN scheme, we use the implicit Euler scheme ($\theta = 1$) for the first 3 time steps and then switch to the CN scheme ($\theta = \frac{1}{2}$) in the rest of the time steps. We implemented a PSOR finite difference scheme to solve the sequence (41) of LCPs efficiently.

In order to accelerate the convergence of the PSOR, we need to select the over-relaxation parameter ω in the algorithm (see Sect. 6.2.3 in Kwok (2008). We notice that we may not choose the same ω when we apply PSOR to different discretised grids in Figs. 1, 2 and 3. Table 1 in the next section demonstrates how the optimal ω should be chosen for different grids on a specific numerical example.

5 Numerical Examples

To demonstrate the performance of both the method of lines and the sparse grid algorithm outlined in Sects. 3 and 4 we implement the methods for a given set of parameter values. Those parameters are the same as the set of parameters used in Medvedev and Scaillet (2010) for comparison purpose. The parameter values used are listed in Table 2.

When implementing the method of lines we take the following case as an example to show its convergence pattern. We use $L = 50$ time-steps, $M = 50$ volatility lines and $N = 50$ interest rate lines, with maximum volatility $v_M = 50\,\%$ and maximum interest rate $r_N = 25\,\%$. We take a non-uniform grid in S, splitting the domain into four intervals. Given that the strike price is $K = 100$, the maximum value for S is set to 800, with a total of 8,000 grid points (denoted by S_{pts}), distributed between the four intervals such that there are 500 points for $0 \le S \le 50$, 500 points for $50 \le S \le 100$, there are 2000 points for $100 \le S \le 200$, and finally 5,000 grid points for $200 \le S \le 800$.

Table 2 Parameter values used for the American call option

Option parameter	Value	SV parameter	Value	SI parameters	Value
q	0.06	θ_v	0.02	θ_r	0.04
T	0.5	κ_v	1.5	κ_r	0.3
K	100	σ_v	0.15	σ_r	0.1
		λ_v	0.00	λ_r	0.00
		ρ_{12}	-0.50	ρ_{13}, ρ_{23}	0.5

The stochastic volatility (SV) and stochastic interest rate (SI) parameters are those used in Medvedev and Scaillet (2010) to facilitate comparisons

Table 3 American call prices computed using the Method of Lines, the sparse grid (SG) with $(c_1 = 16, c_2 = 8, c_3 = 4)$, PSOR in a fine grid with $N_1 = 1,500$, $N_2 = 100$, $N_3 = 100$ and $N_\tau = 100$. Parameter values are given in Table 2, with $v_0 = 0.04$ and $r_0 = 0.04$

Level	S					Runtime(sec)
l	80	90	100	110	120	
1	0.0808	1.2257	5.8304	11.3672	20.1098	2.67
2	0.1265	1.1134	4.3254	11.3350	20.1497	7.72
3	0.1070	1.1128	4.6958	11.3235	20.1246	32.74
4	0.1071	1.1137	4.7455	11.3224	20.1238	170.73
5	0.1073	1.1157	4.7210	11.3246	20.1239	1509.24
6	0.1074	1.1160	4.6712	11.3248	20.1239	5305.76
MOL	0.1073	1.1160	4.6785	11.3248	20.1238	485.00
PSOR	0.1071	1.1151	4.6781	11.2249	20.1234	220824.00

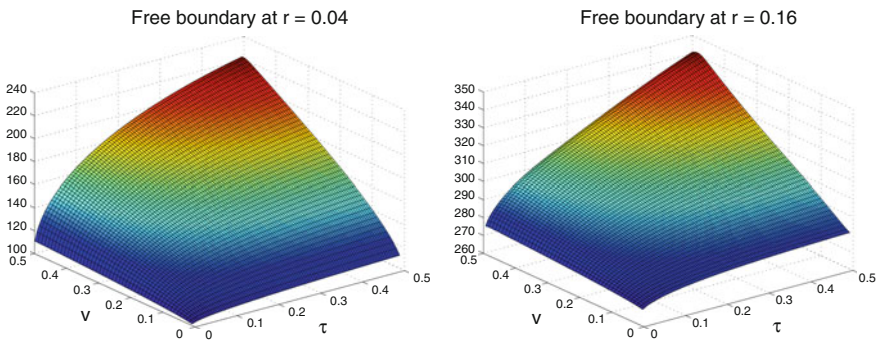

Fig. 4 Free boundary of American Call when $r = 0.04$ and $r = 0.16$

Table 3 shows the American call prices calculated from the Method of Lines, different levels of the Sparse Grid approach and also PSOR with a pretty fine grid. It is apparent that both the MOL and SG approach are efficient and accurate, but the MOL not only produces prices but also early exercise boundaries as shown in Figs. 4, 5 and 6, and hedge ratios.

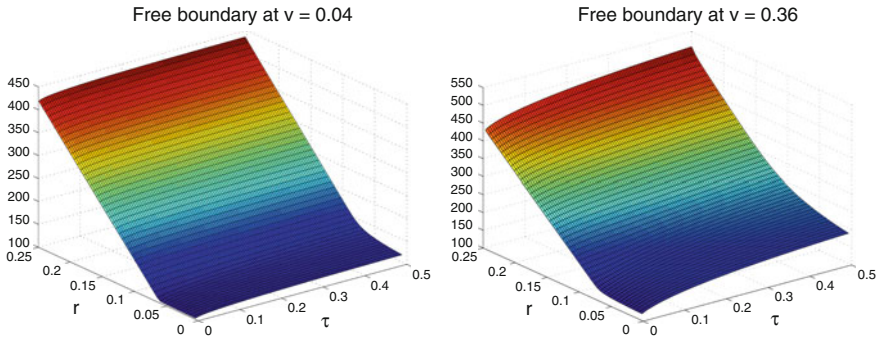

Fig. 5 Free boundary of American Call when $v = 0.04$ and $v = 0.36$

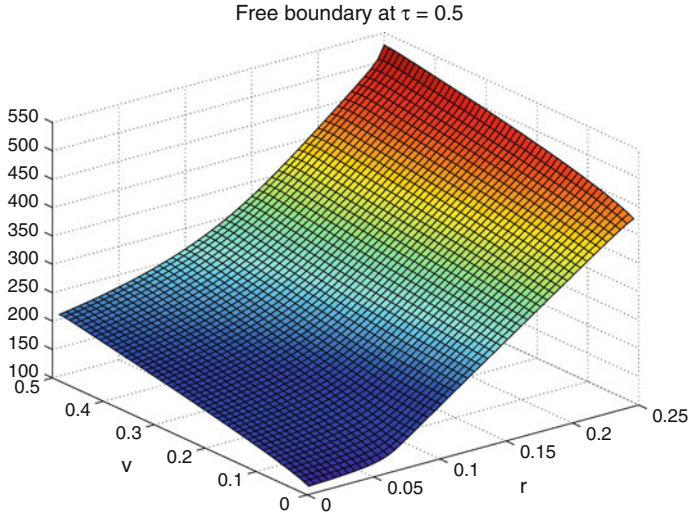

Fig. 6 Free boundary of American Call when $\tau = 0.5$

A number of sample early exercise surfaces are provided in Figs. 4, 5 and 6, generated using the method of lines. The value of the free boundary at expiry is independent of v and r. The free surface, $d(v, r, \tau)$, is an increasing function of v, and along a given value of v we observe an early exercise boundary of the form typically found for American call options. It is worth noting that the free surface generated by the method of lines is smooth, a feature not often displayed in the free boundary estimates generated using finite difference methods, such as Ikonen and Toivanen (2004).

In order to assess how the maximal value of variance $v_{\max}$ and interest rates $r_{\max}$ have influence on the free boundary of the American call option, especially when the time to maturity is short, namely $\tau = 0.1$, we plot in Fig. 7 the free boundaries $s(0.04, r, 0.1)$ over $[0, r_{\max}]$ for $r_{\max} = 0.1, 0.08, 0.05$ and $s(v, 0.04, 0.1)$ over

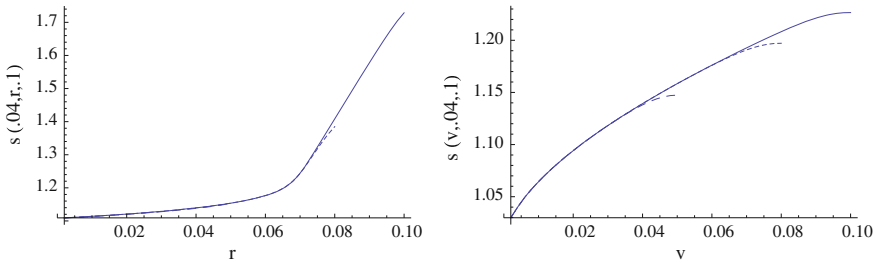

Fig. 7 The panel on the left-hand side shows the free boundary of American Call when $v = 0.04$ and $\tau = 0.1$ with different $r_{\max}$. In particular, $r_{\max} = 0.1$ for the solid line, $r_{\max} = 0.08$ for the short dashes line and $r_{\max} = 0.05$ for the long dashes line. The panel on the left-hand side shows the free boundary of American Call when $r = 0.04$ and $\tau = 0.1$ with different $v_{\max}$. In particular, $v_{\max} = 0.1$ for the solid line, $v_{\max} = 0.08$ for the short dashes line and $v_{\max} = 0.05$ for the long dashes line. The long dash curve coincides to plotting accuracy with the other two for $r < 0.05$

$[0, v_{\max}]$ for $v_{\max} = 0.1, 0.08, 0.05$. The influence of $v_{\max}$ and $r_{\max}$ on the MOL option prices $C(S, v, r, t)$ of Table 3 is quite small but needs to be checked for different v and r. However, the proper $v_{\max}$ and $r_{\max}$ should be chosen if option prices at other grid points of (v_0, r_0) need to be calculated.

In Sect. 4.2, we mentioned that the over-relaxation parameter ω is important to the convergence of the sparse grid approach and it usually depends on the shape of the grids as well. Table 1 illustrates this dependence in the case when $l = 6$. It can be seen from the table that the parameter ω is usually higher for relatively less balanced grids, such as, (0, 0, 6), (0 ,6, 0) etc., in which the calculation will take more time as well.

6 Conclusion

We have studied the pricing of American call options by solving the corresponding PDE using both the method of lines and a sparse grid approach and a benchmark based on a PSOR approach.

It turns out that application of the method of lines approach is able to provide fairly accurate and efficient not only prices but also early exercise boundaries and hedge ratios, such as deltas and gammas for the American call option prices under Heston with stochastic interest rates. The call prices, in particular the early exercise boundaries, are hard to obtain when the interest rate is high relative to the convenience yield q. It is certainly the advantage of the method of lines approach compared with other methods, such as finite differences, to provide accurate, smooth early exercise boundaries in an efficient manner.

Acknowledgments We would like to thank one referee for his/her comments which improved this chapter.

References

Adolfsson, T., Chiarella, C., Ziogas, A., & Ziveyi, J. (2013). *Representation and numerical approximation of American option prices under heston stochastic volatility dynamics*. Quantitative Finance Research Centre, University of Technology Sydney, Working paper No. 327.

Black, F., & Scholes, M. (1973). The pricing of corporate liabilites. *Journal of Political Economy*, *81*, 637–659.

Boyarchenko, S., & Levendorski, S. (2013). American options in the heston model with stochastic interest rate and its generalizations. *Applied Mathematical Finance*, *20*(1), 26–49.

Chiarella, C., & Kang, B. (2013). The evaluation of american compound option prices under stochastic volatility and stochastic interest rates. *Journal of Computational Finance*, *17*(1), 71–92.

Chiarella, C., Kang, B., Meyer, G. H., & Ziogas, A. (2009). The evaluation of American option prices under stochastic volatility and jump-diffusion dynamics using the method of lines. *International Journal of Theoretical and Applied Finance*, *12*(3), 393–425.

Chiarella, C., Ziogas, A., & Ziveyi, J. (2010). Contemporary quantitative finance: Essays in honour of eckhard platen. *Representation of American option prices under heston stochastic volatility dynamics using integral transforms* (pp. 281–315). Berlin: Springer.

Ekstrom, E., Lotstedt, P., & Tysk, J. (2009). Boundary values and finite difference methods for the single factor term structure equation. *Applied Mathematical Finance*, *16*(3), 253–259.

Griebel, M., Schneider, M., & Zenger, C. (1992). A combination technique for the solution of sparse grid problems. *Iterative methods in linear algebra* (pp. 263–281). Amsterdam: Elsevier.

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., & Toivanen, J. (2004). perator splitting methods for american options with stochastic volatility. *Applied Mathematics Letters*, *17*, 809–814.

Ikonen, S., & Toivanen, J. (2007). Ocomponentwise splitting methods for pricing American options under stochastic volatility. *International Journal of Theoretical and Applied Finance*, *10*(2), 331–361.

Kjellin, R., & Lovgren, G. (2006). Option pricing under stochastic volatility, Master's thesis, U. of Gothenburg.

Kwok, Y. -K. (2008). *Mathematical models of financial derivatives*. 2nd edn, Berlin: Springer.

Leentvaar, C., & Oosterlee, C. (2008a). Multi-asset option pricing using a parallel fourier-based technique. *Journal of Computational Finance*, *12*(1), 1–26.

Leentvaar, C., & Oosterlee, C. (2008b). On coordinate transformation and grid stretching for sparse grid pricing of basket options. *Journal of Computational and Applied Mathematics*, *222*, 193–209.

Medvedev, A., & Scaillet, O. (2010). Pricing American options under stochastic volatility and stochastic interest rates. *Journal of Financial Economics*, *98*, 145–159.

Meyer, G. (2014). Pricing Options and Bonds with the Method of Lines, Georgia Institute of Technology, http://people.math.gatech.edu/meyer/MOL-notes/

Meyer, G. H., & van der Hoek, J. (1997). The evaluation of American options with the method of lines. *Advances in Futures and Options Research*, *9*, 265–285.

Oleinik, O., & Radkevic, E. V. (1973). *Second-order equations with non-negative characteristic form*, Berlin: Springer.

Reisinger, C. (2004). Numerische Methoden für hochdimensionale parabolische Gleichungen am Beispiel von Optionspreisaufgaben, PhD thesis, Universität Heidelberg.

Reisinger, C. (2008). *Analysis of linear difference schemes in the sparse grid combination technique. Technical report*. Oxford, UK: Mathematical Institute.

Reisinger, C., & Wittum, G. (2007). Efficient hierarchical approximation of high-dimensional option pricing problems. *SIAM Journal on Scientific Computing*, *29*(1), 440–458.

Shampine, L. F. (1994). *Numerical solution of ordinary differential equations*. New York: Chapman & Hall.

On the Volatility of Commodity Futures Prices

Les Clewlow, Boda Kang and Christina Sklibosios Nikitopoulos

1 Introduction

Commodity futures markets are playing a leading role in the current financial arena. Commercial participants with physical positions in commodities have traditionally been the main traders of commodity futures markets. However, a rapidly increasing number of financially motivated traders such as hedge funds, institutional investors and insurance companies have entered the markets and have been using commodities derivatives for portfolio diversification, and for hedging inflation and the weak U.S. dollar (Tokic (2011)). Commodity markets have experienced noteworthy price swings and significant volatility especially over the past decade. Consequently, the analysis and the management of this volatility is of paramount importance.

In this chapter, a forward price model within the Heath et al. (1992) framework for the entire term structure of futures prices is combined with a multi-factor stochastic volatility model. The proposed three-factor model aims to capture the impact of short-term, medium-term, as well as long-term futures price volatility by using exponential decaying and hump-shaped stochastic volatility factors. For these volatility specifications, the forward price model admits finite-dimensional realizations and is affine in the state space. Consequently, the model is estimated by using an extended dataset of futures prices for six major commodities traded on the CME, spanning 21 years from 1st January 1990 to 31st December 2010. Selecting the most liquid

L. Clewlow
Lacima Group, Sydney, NSW, Australia
e-mail: les@lacimagroup.com

B. Kang
Department of Mathematics, University of York, York, UK
e-mail: boda.kang@york.ac.uk

C. S. Nikitopoulos (✉)
UTS Business School, University of Technology, Sydney, Sydney, NSW, Australia
e-mail: christina.nikitopoulos@uts.edu.au

R. Dieci et al. (eds.), *Nonlinear Economic Dynamics and Financial Modelling*,
DOI: 10.1007/978-3-319-07470-2_18,

commodity futures markets, the following commodity futures prices are included in the study; gold, crude oil, natural gas, soybean, sugar and corn.

The model is well suited to identify the shape, the persistence and the intensity of each volatility factor. An exponential decaying volatility factor typically gauges the impact of short-term variations, subject to the rate of the decay. A hump-shaped volatility factor, in general, captures the impact of medium-term variations, subject to the location of the hump peak. The volatility structure of an infinite maturity forward price gauges the impact of long-term variation. The rate of mean reversion and the volatility for each of the stochastic volatility factors are also assessed and indicate the level of persistence and the intensity of each volatility factor. Furthermore, the model determines the extent to which commodity futures volatility can be spanned by futures contracts (i.e. hedged by futures contracts only), and the nature of the return-volatility relation in commodity futures markets.

Forward price models such as Miltersen and Schwartz (1998), Clewlow and Strickland (2000) and Miltersen (2003), as well as convenience yield models of Gibson and Schwartz (1990) and Korn (2005) have studied commodity futures markets but they are restricted to deterministic volatility. Stochastic volatility models have been proposed by Schwartz and Trolle (2009a) and Chiarella et al. (2013) and have analysed crude oil futures market volatility, under exponential decaying and hump-shaped volatility specifications, respectively. The contribution of this paper rests on using three distinct volatility structures and aiming to analyze their nature and assess their contribution.

The paper is organized as follows. Section 2 presents a Markovian affine forward price model with stochastic volatility. Section 3 describes and analyses the commodity futures price data used in the analysis and outlines the estimation method used to estimate the model. Section 4 presents and discusses the results. Section 5 concludes.

2 A Markovian Commodity Futures Price Model

A filtered probability space $(\Omega, \mathscr{A}_T, \mathscr{F}_0, P)$, $T \in (0, \infty)$ with $\mathscr{F}_0 = (\mathscr{A}_t)_{t\in[0,T]}$ is assumed, satisfying the usual conditions.[1] The uncertainty in the commodity futures market is modelled via a generic stochastic volatility process $\mathbf{V} = \{\mathbf{V_t}, t \in [0, T]\}$. Let us denote as $S(t, \mathbf{V_t})$ the spot commodity price at time $t \geq 0$, and $F(t, T, \mathbf{V_t})$ the commodity futures price at time t, for delivery at time T, (for all maturities $T \geq t$), thus by definition, $S(t, \mathbf{V_t}) = F(t, t, \mathbf{V_t})$, $t \in [0, T]$. Furthermore, no-arbitrage arguments in commodity futures markets imply that the futures price process is equal to the expected future commodity spot price under an equivalent risk-neutral probability measure Q (see Duffie (2001)), namely

$$F(t, T, \mathbf{V_t}) = \mathbf{E}^Q[S(T, \mathbf{V_T})|\mathscr{A}_t].$$

[1] The usual conditions satisfied by a filtered complete probability space are: (a) $\mathscr{F}_0$ contains all the P-null sets of $\mathscr{F}$ and (b) the filtration is right continuous. See Protter (2004) for technical details.

Consequently, the commodity futures price is a martingale under the risk-neutral measure and the commodity futures price process should follow a driftless stochastic differential equation under the risk-neutral measure. Accordingly, a three-factor model is proposed of the form

$$\frac{dF(t,T,\mathbf{V_t})}{F(t,T,\mathbf{V_t})} = \sum_{i=1}^{3} \sigma_i(t,T,\mathbf{V_t}) dW_i(t), \tag{1}$$

where, $W(t) = \{W_1(t), W_2(t), W_3(t)\}$ is a three-dimensional Wiener process. The $\mathscr{F}_0$-adapted futures price volatility processes $\sigma_i(t,T,\mathbf{V_t})$ have the functional forms, for all $T > t$,

$$\begin{aligned}
\sigma_1(t,T,\mathbf{V_t}) &= \kappa_1 \sqrt{\mathbf{V_t^1}}, \\
\sigma_2(t,T,\mathbf{V_t}) &= \kappa_2 e^{-\eta_2(T-t)} \sqrt{\mathbf{V_t^2}}, \\
\sigma_3(t,T,\mathbf{V_t}) &= \kappa_3 (T-t) e^{-\eta_3(T-t)} \sqrt{\mathbf{V_t^3}},
\end{aligned} \tag{2}$$

with κ_i, $(i = 1,2,3)$ and η_i, $(i = 2,3)$ constants. The volatility process $\mathbf{V_t} = \{\mathbf{V_t^1}, \mathbf{V_t^2}, \mathbf{V_t^3}\}$ is a three-dimensional Heston (1993) type process such that

$$d\mathbf{V_t^i} = \mu_i(\nu_i - \mathbf{V_t^i})dt + \varepsilon_i \sqrt{\mathbf{V_t^i}} dW_i^V(t), \tag{3}$$

where μ_i, ν_i, and ε_i are constants for $i = 1,2,3$, and $W^V(t) = \{W_1^V(t), W_2^V(t), W_3^V(t)\}$ is the three-dimensional Wiener process driving the stochastic volatility process $\mathbf{V}_t$, for all $t \in [0,T]$. The first volatility factor can be considered as the factor capturing the volatility of the futures price returns with infinite maturity, thus representing the long-term volatility in commodity futures markets. The second volatility factor predominantly gauges the volatility of the short-term futures price returns as it allows a volatility structure that decays exponentially as the time to maturity increases. The third volatility factor generates humps in the volatility structure, thus reveals principally the impact of the volatility of medium-term futures prices. The proposed volatility structure also allows each of these volatility factors to be driven by a different stochastic volatility processes, consequently the proposed model has the potential to capture the impact and the nature of market shocks to the entire volatility term structure (including short-term, medium-term and long-term) and determine their contribution to the total variance.

In addition, the following correlation structure of innovations between volatility and futures price returns is assumed

$$\mathbf{E}^Q[dW_i(t) \cdot dW_j^V(t)] = \rho_i dt, \text{ for } i = j; \text{ and } 0, \text{ for } i \neq j, \tag{4}$$

where ρ_i are constants for $i = 1,2,3$. The correlation structure of innovations between volatility and futures prices determines the extent to which the volatility risk

can be hedged (spanned) by futures contracts. When the Wiener processes $W_i(t)$ and $W_i^V(t)$ are uncorrelated then the volatility risk is unhedgeable by futures contracts. When the Wiener processes $W_i(t)$ and $W_i^V(t)$ are correlated, then the volatility risk can be partially hedged by futures contracts. Consequently, futures derivatives that are sensitive to volatility, such as options on futures, cannot be completely hedged by using only futures contracts. Furthermore, this modelling framework allows us to assess the dynamic relationship between futures price returns and volatility changes. A negative (positive) correlation implies a negative (positive) relation between price returns and their volatility, a well-known empirical phenomenon termed as asymmetric (inverted asymmetric) volatility. Yet again, the proposed model can capture the volatility reaction of each of the volatility factors, as these factors are modelled as separate identities.

For modelling convenience, we express the system of Eqs. (1) and (3) in terms of independent Wiener process, such that

$$\frac{dF(t,T,\mathbf{V_t})}{F(t,T,\mathbf{V_t})} = \sum_{i=1}^{3} \sigma_i(t,T,\mathbf{V_t}) dW_i^1(t), \tag{5}$$

$$d\mathbf{V_t^i} = \mu_i(\nu_i - \mathbf{V_t^i})dt + \varepsilon_i \sqrt{\mathbf{V_t^i}} \left(\rho_i dW_i^1(t) + \sqrt{1-\rho_i^2} dW_i^2(t) \right), \tag{6}$$

where $W^1(t) = W(t)$ and $W^2(t)$ are three-dimensional independent Wiener processes. Accordingly, the volatility factors $\mathbf{V_t^i}$ with $\rho_i = 0$ carries a risk that cannot be spanned by futures contracts alone and when $\rho_i < 0$ ($\rho_i > 0$) then the volatility factor $\mathbf{V_t^i}$ has an asymmetric (inverted asymmetric) reaction.

It is well known that for general volatility specifications, the forward price model for pricing the commodity futures (5) is Markovian in an infinite dimensional state space. However, the volatility specifications (2) produce finite dimensional realisations of the forward price model, see Chiarella and Kwon (2001) and Björk et al. (2004).

Theorem 1 *Under the volatility specifications of* (2), $\ln F(t,T,\mathbf{V_t})$ *is affine in nine state variables, as described below:*

$$\ln F(t,T,\mathbf{V_t}) = \ln F(0,T,V_0) + \sum_{n=1}^{4} \beta_n(T-t)\phi_n(t) - \frac{1}{2}\sum_{j=1}^{5} \gamma_j(T-t)\boldsymbol{x}_j(t), \tag{7}$$

where

$$\beta_1(T-t) = \kappa_1, \quad \gamma_1(T-t) = \kappa_1^2,$$
$$\beta_2(T-t) = \kappa_2 e^{-\eta_2(T-t)}, \quad \gamma_2(T-t) = \kappa_2^2 e^{-2\eta_2(T-t)}$$

$$\beta_3(T-t) = \kappa_3(T-t)e^{-\eta_3(T-t)}, \quad \beta_4(T-t) = \kappa_3 e^{-\eta_3(T-t)},$$
$$\gamma_3(T-t) = \beta_3(T-t)^2, \quad \gamma_4(T-t) = 2\beta_3(T-t)\beta_4(T-t), \gamma_5 = \beta_4(T-t)^2.$$

The state variables $\phi_n(t)$, $n = 1, \ldots, 4$ *and* $\mathbf{x}_j(t)$, $j = 1, \ldots, 5$ *satisfy the stochastic differential equations*

$$\begin{aligned}
d\phi_1(t) &= \sqrt{\mathbf{V_t^1}}dW_1(t),\\
d\phi_2(t) &= -\eta_2\phi_2(t)dt + \sqrt{\mathbf{V_t^2}}dW_2(t),\\
d\phi_3(t) &= -\eta_3\phi_3(t)dt + \sqrt{\mathbf{V_t^3}}dW_3(t),\\
d\phi_4(t) &= (-\eta_3\phi_4(t) + \phi_3(t))dt,\\
d\mathbf{x}_1(t) &= \mathbf{V_t^1}dt,\\
d\mathbf{x}_2(t) &= \left(-2\eta_2\mathbf{x}_2(t) + \mathbf{V_t^2}\right)dt,\\
d\mathbf{x}_3(t) &= (-2\eta_3\mathbf{x}_3(t) + \mathbf{V_t^3})dt,\\
d\mathbf{x}_4(t) &= (-2\eta_3\mathbf{x}_4(t) + \mathbf{x}_3(t))dt,\\
d\mathbf{x}_5(t) &= (-2\eta_3\mathbf{x}_5(t) + 2\mathbf{x}_4(t))dt,
\end{aligned} \tag{8}$$

subject to $\phi_n(0) = \mathbf{x}_j(0) = 0$*, for all* n *and* j*. The associated stochastic volatility process* $\mathbf{V_t} = \{\mathbf{V_t^1}, \mathbf{V_t^2}, \mathbf{V_t^3}, \}$ *follows the dynamics* (6).

Proof The technical details are summarized in the Appendix. □

The proposed model admits finite dimensional realizations within the affine class of Duffie and Kan (1996) and it is consistent, by construction, with the currently observed futures price curve; consequently, it is time-inhomogeneous. However for estimation purposes, it is necessary to reduce the model to a time-homogeneous one as presented in Sect. 3.2.

In addition, the market price of volatility risk is modelled with "complete" affine specifications (see Doran and Ronn (2008) and Dai and Singleton (2000)) and more specifically as

$$\begin{aligned}
dW_i^{\mathbf{P}}(t) &= dW_i(t) + \lambda_i\sqrt{\mathbf{V_t}^i}dt,\\
dW_i^{\mathbf{P}V}(t) &= dW_i^V(t) + \lambda_i^V\sqrt{\mathbf{V_t}^i}dt,
\end{aligned} \tag{9}$$

for $i = 1, 2, 3$, where $W_i^{\mathbf{P}}(t)$ and $W_i^{\mathbf{P}V}(t)$ are Wiener processes under the physical measure $\mathbf{P}$. Next, the proposed model is estimated by fitting the model to commodity futures prices of six key commodities: gold, crude oil, natural gas, soybean, sugar and corn.

3 Data and the Estimation Method

3.1 Data

An extended dataset of six commodity futures prices provided by CME is used that spans 21 years from 1st January 1990 to 31st December 2010. The selected commodities are the most actively traded commodities with crude oil leading the ladder, following by gold, soybeans, natural gas, sugar and corn.[2] Even though historical data for some of the commodities go back to 1970s, for consistency purposes, our analysis is concentrated in the past 21 years for all commodities. Furthermore, the selected commodities can be regarded as representative commodities within the metal (including gold), energy (including crude oil and natural gas) and agricultural (including soybeans, sugar and corn) commodities products.

Over this period, all commodity markets experienced extreme price movements and volatility due to noteworthy financial market and macroeconomic events such as the oil price crisis in 1990, the financial crisis in 2008, the crude oil bubble in 2008 and the ongoing food crisis initiated in 2008. Throughout the sample period, the number of available futures contracts for all commodities with positive open interest per day has increased significantly as well as the maximum maturity of futures contracts with positive open interest. For example, for crude oil, the open interest per day has increased from 17 on the 1st of January 1990 to 67 on the 31st of December 2010 and the maximum maturity of crude oil futures contracts with positive open interest has increased from 499 (calendar) days to 3,128 days.

As the number of available futures contracts per day is relatively large, for the estimation analysis, the most liquid futures contracts are used, with liquidity measured by the open interest. As contracts close to expiry have very low liquidity, the contracts selected for the study here all have more than 14 days to expiry. Each commodity has different available delivery months and their liquidity is concentrated on different contracts. Thus, the available contract months and the liquidity for each commodity are investigated, and the following selection per commodity has been made. For crude oil, the first seven monthly contracts are used, near to the trade date, followed by the three contracts which have either March, June, September or December expiration months and then we include the next five December contracts. Therefore, the

[2] Information on liquidity was collected by http://www.barchart.com, as well as by computing the average open interest available on the data. For example, on Sept 9, 2013 volumes were; 261,394 for Oct 13 crude oil, 151,589 for Dec 13 gold, 112,208 for Nov 13 soybeans, 100,027 for Oct 13 natural gas, 90,026 for Oct 13 sugar and 89,000 for Oct 13 corn.

number of crude oil futures contracts used on a daily basis ranges between 8 and 15. As natural gas has continuous monthly contracts, a maximum of 15 monthly contracts is selected (near to the trade date). For gold, the first three monthly contracts are included, followed by the four contracts with expiration months of February, April, June, August, October and December and finally four semi-annual contracts with maturities of June or December. Consequently, the number of gold futures contracts used on a daily basis varies between 8 and 11. For soybeans, the available contract months are January, March, May, July, August, September and November, therefore the first 15 contracts are used (near to the trade date), with these maturity months. The available and more liquid contract months for sugar are March, May, July and October, thus the first ten contracts are used, near to the trade date, for these maturities. The corn futures more liquid maturities are March, May, July, September and December thus all available maturities are included, with a maximum number of contracts per day being 15. In accordance with the above selection, the longest maturity, in terms of months, we have chosen for each commodity is 31 for gold, 126 for crude oil, 102 for natural gas, 60 for soybeans, 54 for sugar and 61 for corn.

It is worth noticing that for all commodities of interest the futures price surfaces have changed significantly throughout the sample period with extreme price variation over the last decade, as depicted in Fig. 1. This notable variation is also apparent from the descriptive statistics presented in Tables 1 and 2 that display the statistical features of the selected commodity futures price returns for 1 month and 12 month futures contracts for the commodities of interest.

3.1.1 Number of Stochastic Factors

The number of driving stochastic factors affecting the evolution of the futures curve is investigated by performing a principal component analysis (PCA) of the futures price returns. Table 3 displays varying levels of contributions for different commodities and Fig. 2 depicts the eigenvalues and associated volatility functions. Three factors potentially can explain between 93 % (for natural gas) to 99 % (for crude oil) of the total variation in futures returns. The proposed three-factor volatility structure (2) aims to capture the impact of shocks in the entire term structure (from short-term to long-term), and it could be represented by the three factors revealed by the PCA, even though their nature is distinctively determined by the model.

3.1.2 The Discount Function

The discount function $P(t, T)$ is obtained by fitting a Nelson and Siegel (1987) curve each trading day to LIBOR and swap data consisting of 1, 3, 6, 9 and 12 month LIBOR rates and the 2 year swap rate, similar to Schwartz and Trolle (2009a).

Let $f(t, T)$ denote the time$-t$ instantaneous forward interest rate to time T. Nelson and Siegel (1987) parameterize the forward interest rate curve as

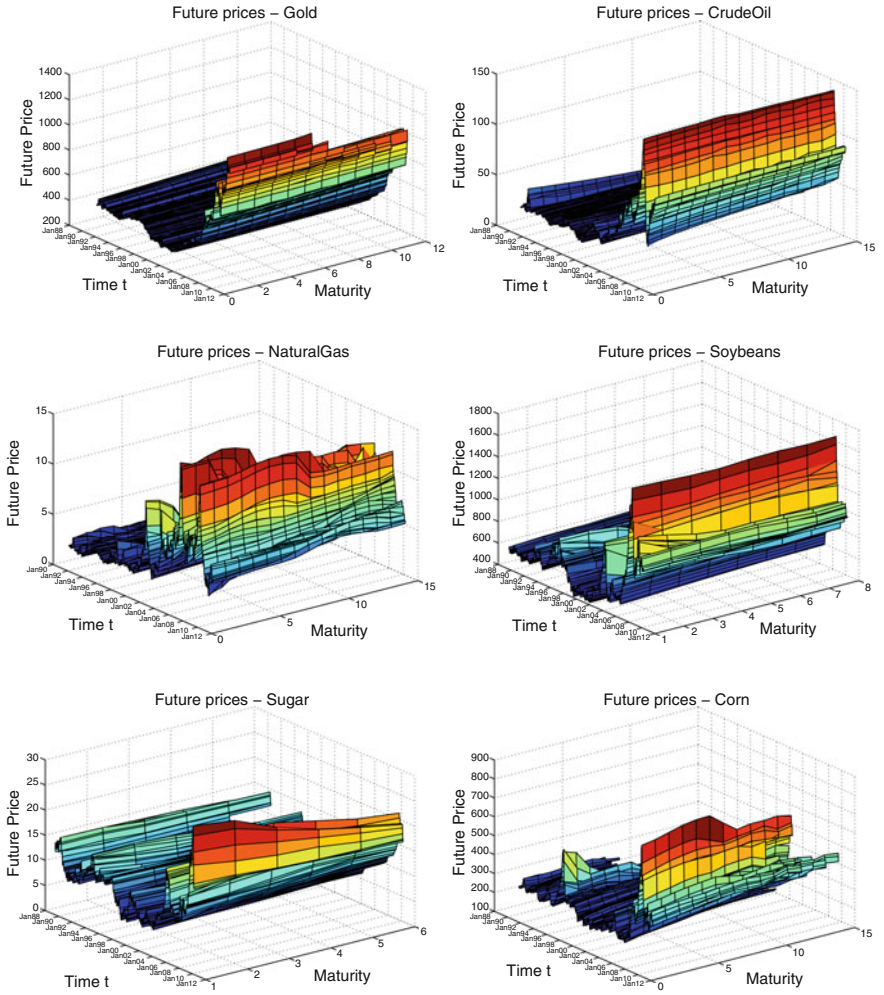

Fig. 1 Commodity futures prices. The figure plots the prices of selected commodity futures contracts from January 2, 1990 to December 31, 2010

$$f(t,T) = \alpha_0 + \alpha_1 e^{-\theta(T-t)} + \alpha_2 \theta (T-t) e^{-\theta(T-t)} \tag{10}$$

from which LIBOR and swap rates can be priced. This also yields for zero-coupon bond prices the expression

$$P(t,T) = \exp\left\{ -\alpha_0 (T-t) - \frac{1}{\theta}(\alpha_1 + \alpha_2)\left(1 - e^{-\theta(T-t)}\right) + \alpha_2 (T-t) \right\}. \tag{11}$$

Table 1 Descriptive statistics—gold, crude oil and natural gas

Maturity	Gold		Crude oil		Natural gas	
	1 M	12 M	1 M	12 M	1 M	12 M
Mean	0.000220	0.000200	0.000291	0.000306	0.000185	0.000179
St. Dev.	0.011972	0.010799	0.026004	0.017870	0.037743	0.022392
Kurtosis	39.837774	28.133742	20.764654	8.156276	16.008901	19.721668
Skewness	−0.124339	−0.162282	−0.893782	−0.298393	0.301666	−0.069086

The table displays the descriptive statistics for daily log returns of futures prices between January 2, 1990 and December 31, 2010 for gold, crude oil and natural gas

Table 2 Descriptive statistics—soybeans, sugar and corn

Maturity	Soybeans		Sugar		Corn	
	1 M	12 M	1 M	12 M	1 M	12 M
Mean	0.000090	0.000084	0.000061	0.000225	0.000077	0.000096
St. Dev.	0.015746	0.013324	0.022504	0.014234	0.016618	0.012512
Kurtosis	12.025013	7.275205	12.437959	7.495627	21.100643	7.647190
Skewness	−0.932488	−0.341833	−0.345857	−0.243873	−0.847868	−0.147756

The table displays the descriptive statistics for daily log returns of futures prices between January 2, 1990 and December 31, 2010 for soybeans, sugar and corn

Table 3 Accumulated percentage of factor contribution

Commodity	One factor	Two factors	Three factors	Four factors
Gold	0.9735	0.9810	0.9881	0.9926
Crude oil	0.9503	0.9825	0.9919	0.9957
Natural gas	0.7636	0.8625	0.9308	0.9666
Soybeans	0.8973	0.9413	0.9639	0.9798
Sugar	0.8788	0.9560	0.9832	0.9956
Corn	0.8764	0.9236	0.9487	0.9712

The table displays the accumulated percentage of PCA factor contribution towards each commodity futures return variation. We found that three factors are able to explain most of the variations of the futures returns for the commodities of interest

On each observation date, the parameters $\alpha_0, \alpha_1, \alpha_2$ and θ are recalibrated, by minimizing the mean squared percentage differences between the model implied forward rates (as described in (10)) and the observed LIBOR and swap curve consisting of the 1, 3, 6, 9 and 12 month LIBOR rates and the 2 year swap rate on that date.

3.2 Estimation Method

The estimation approach is quasi-maximum likelihood in combination with the Kalman filter. The model is cast into a state-space form, which consists of the system equations and the observation equations. For estimation purposes, a time-homogeneous version of the model (7) is considered, by assuming for all T,

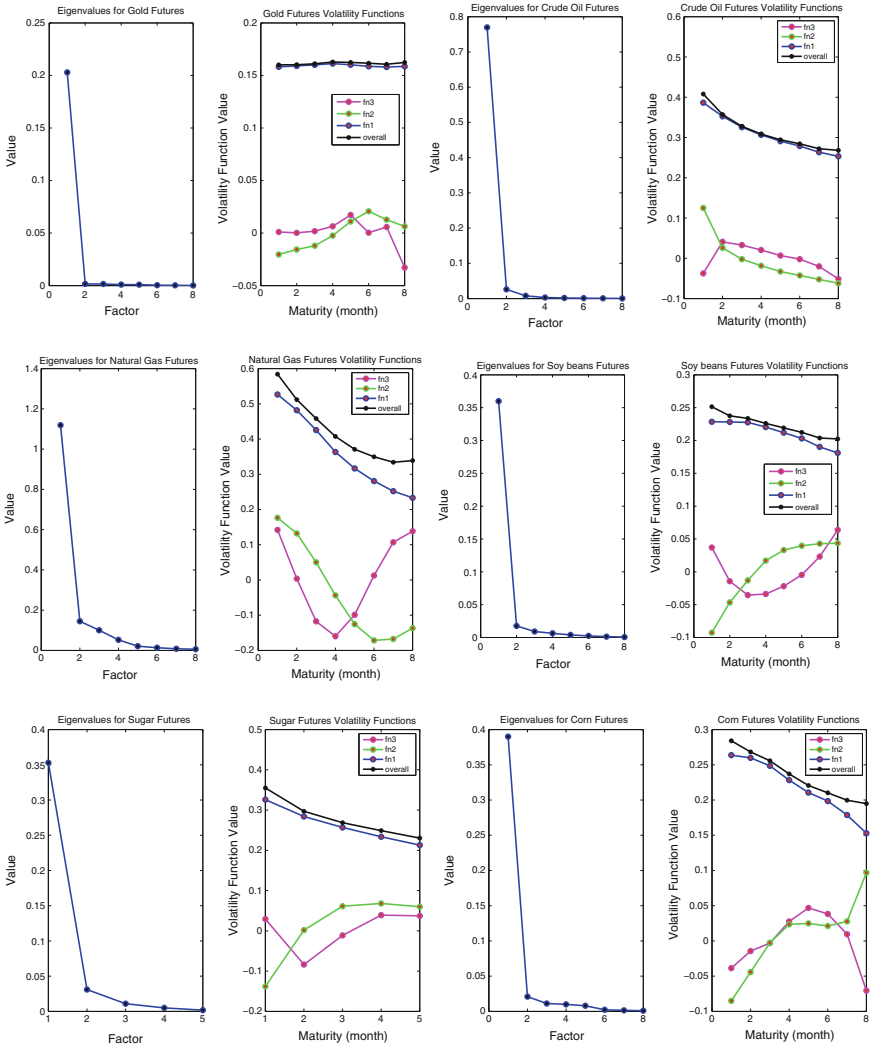

Fig. 2 PCA results (Eigenvalues and volatility functions)

$F(0,T) = f_o$, where f_o is a constant representing the long-term futures price (at infinite maturity). This constant is an additional parameter that is also to be estimated.

The system equations describe the evolution of the underlying state variables. In our case, the state vector is $X_t = \{X_t^m, m = 1, 2, \ldots, 12\}$, where X_t consists of the 12 state variables; $\mathbf{x}_j(t)$, $j = 1, \ldots, 5$, $\phi_n(t)$, $n = 1, \ldots, 4$, and $\mathbf{V_t^i}$, $i = 1, 2, 3$. The continuous time dynamics (under the physical probability measure) of these state variables are defined by Eqs. (6), (8) and (9). The corresponding discrete evolution is

$$X_{t+1} = \Phi_0 + \Phi_X X_t + w_{t+1},\ w_{t+1} \sim iid\ N(0, Q_t), \tag{12}$$

where Φ_0, Φ_X and Q_t can be computed in closed form.

From Eq. (7), log futures prices are linear functions of the state variables, thus the observation equation based on (7) can describe the relationship between observed futures prices and the state variables as

$$z_t = h(X_t) + u_t,\ u_t \sim iid\ N(0, \Omega). \tag{13}$$

The loglikehood function is maximised by using the constrained optimization routine "e04jy" in the NAG library. Several different initial hypothetical parameter values are considered, first on monthly data, then on weekly data and finally on daily data, aiming to obtain global optima.

4 Empirical Results

4.1 Parameter Estimation

Table 4 presents the parameter estimates of a three-factor stochastic volatility model. Figures 3 and 4 depict the estimated deterministic part $\chi_i(t)$ of each volatility factor σ_i ($\sigma_i(t, T, \mathbf{V_t^i}) = \chi_i(t, T)\sqrt{\mathbf{V_t^i}}$) and the estimated time-series of volatility state factor $\mathbf{V_t^i}$.[3] Recall that the first volatility factor portrays the long-term volatility factor, while the second volatility factor is exponential decaying (dies out as the time to maturity increases) and is associated with short-term market uncertainty. The third volatility factor is hump-shaped, (volatility increases with time to maturity to a peak level, then decreases for longer times to maturity) and describes the medium-term volatility.

4.2 Gold Volatility

According to the parameter estimates for the gold futures market, see Table 4, only the first volatility factor is significant in magnitude and in contribution. For the other two factors, κ_i and ν_i^V are very close to zero, thus both the deterministic element and the stochastic element of these volatility factors are very close to zero. Table 5 shows that the long-term volatility factor contributes 99.98 % of the total variance, a result that is also consistent with PCA. The associated volatility state factor $\mathbf{V_t^1}$ is not persistent and reverts relatively quickly to the mean level (evidenced by the high value of μ_i) while its volatility is large, see also the top panels of Fig. 3. Thus,

[3] In absolute terms, $\mathbf{V_t^i}$ is the variance process and $\sqrt{\mathbf{V_t^i}}$ is the volatility process.

Table 4 Parameter Estimates

	f_o	Factor	κ_i	η_i	μ_i	ν_i	ϵ_i	ρ_i	λ_i	λ_i^V
Metals										
Gold										
	6.78	1	−0.4324	0.0000	−1.9822	0.0100	4.0110	0.8996	−3.0000	−2.1974
		2	−0.0109	4.9979	−1.6438	0.0305	1.3018	−0.9127	2.9762	−2.9740
		3	0.0308	4.8705	4.9083	0.0404	1.8998	0.9490	2.8698	−2.5760
Energy										
Crude oil										
	4.50	1	0.0364	0.0000	3.4978	0.1015	0.0100	0.5411	3.0000	−0.6957
		2	0.3735	0.3914	−2.0000	0.2624	5.0000	−0.8337	−3.0000	−1.1797
		3	−0.0804	1.6611	5.0000	5.0000	0.0619	0.9500	−3.0000	−1.6399
Natural gas										
	1.49	1	−0.2904	0.0000	−2.0000	0.9012	5.0000	0.3116	−3.0000	−1.1331
		2	0.6463	4.2571	5.0000	0.1914	5.0000	−0.9500	−3.0000	−2.8280
		3	2.4848	5.0000	−2.0000	5.0000	5.0000	−0.5886	−3.0000	−0.9792
Agricultural										
Soybeans										
	6.30	1	0.5595	0.0000	5.0000	0.0356	0.9964	0.7338	1.9856	1.4988
		2	0.0448	4.9300	4.8135	0.0130	0.9412	−0.0797	0.2685	−2.4983
		3	0.4287	2.8152	0.1185	4.8349	0.0100	−0.2922	0.2036	−0.8500
Sugar										
	1.94	1	0.2228	0.0000	−2.0000	0.5134	5.0000	−0.9155	−3.0000	−3.0000
		2	0.0000	−1.6488	−2.0000	0.1219	1.7671	−0.9221	−2.9525	−2.9494
		3	1.0191	3.3670	0.0828	5.0000	5.0000	−0.9500	0.0642	−0.0199
Corn										
	6.47	1	−0.4188	0.0000	−0.0762	0.1527	5.0000	0.9500	3.0000	−2.5545
		2	−0.0143	3.8723	−1.8526	0.1028	4.6859	−0.8463	0.1100	−0.8652
		3	0.9395	1.3926	1.3400	0.0100	0.4731	0.9500	−3.0000	−2.7719

The table displays the quasi maximum-likelihood estimates for the three-factor model specifications. F is the homogenous futures price at time 0, namely $F(0,t) = f_o, \forall t$

the long-term volatility is the dominant volatility factor in the gold futures market. Furthermore, the innovations of this volatility factor has a correlation of 0.9 with the innovations of the gold futures price returns, implying that gold futures price volatility can be mostly hedged by gold futures contracts. Moreover, the positive sign of the correlation coefficient confirms the well-documented positive gold return-volatility relation that in the spot gold market has predominantly been explained by the safe haven effect, see Baur (2012).

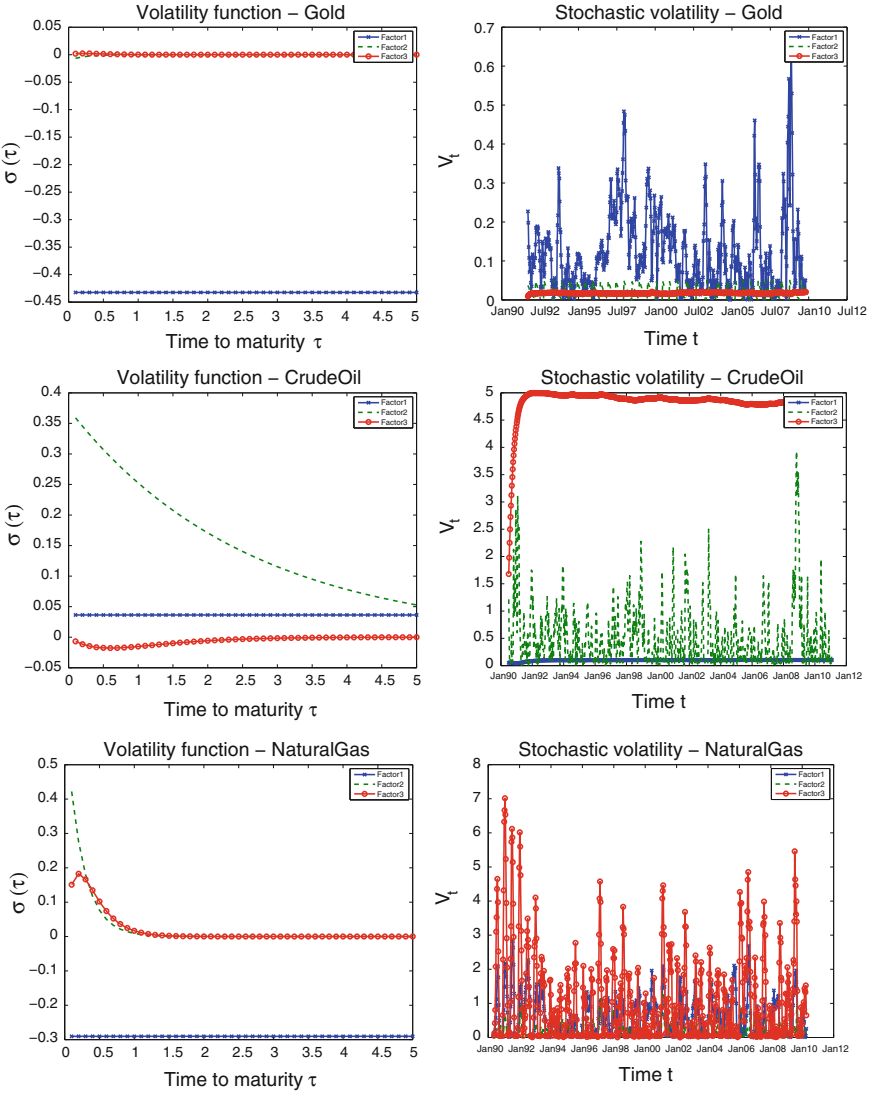

Fig. 3 Estimated volatility factors for metals and energy futures prices

4.3 Crude Oil Volatility

Table 4 demonstrates that all three volatility factors are important in the crude oil futures market. More specifically, from Table 5, the exponential decaying and hump-shaped volatility factors account for the majority of the volatility, with contributions of 65.43 and 27.84 % to the total variance, respectively. The exponential decaying

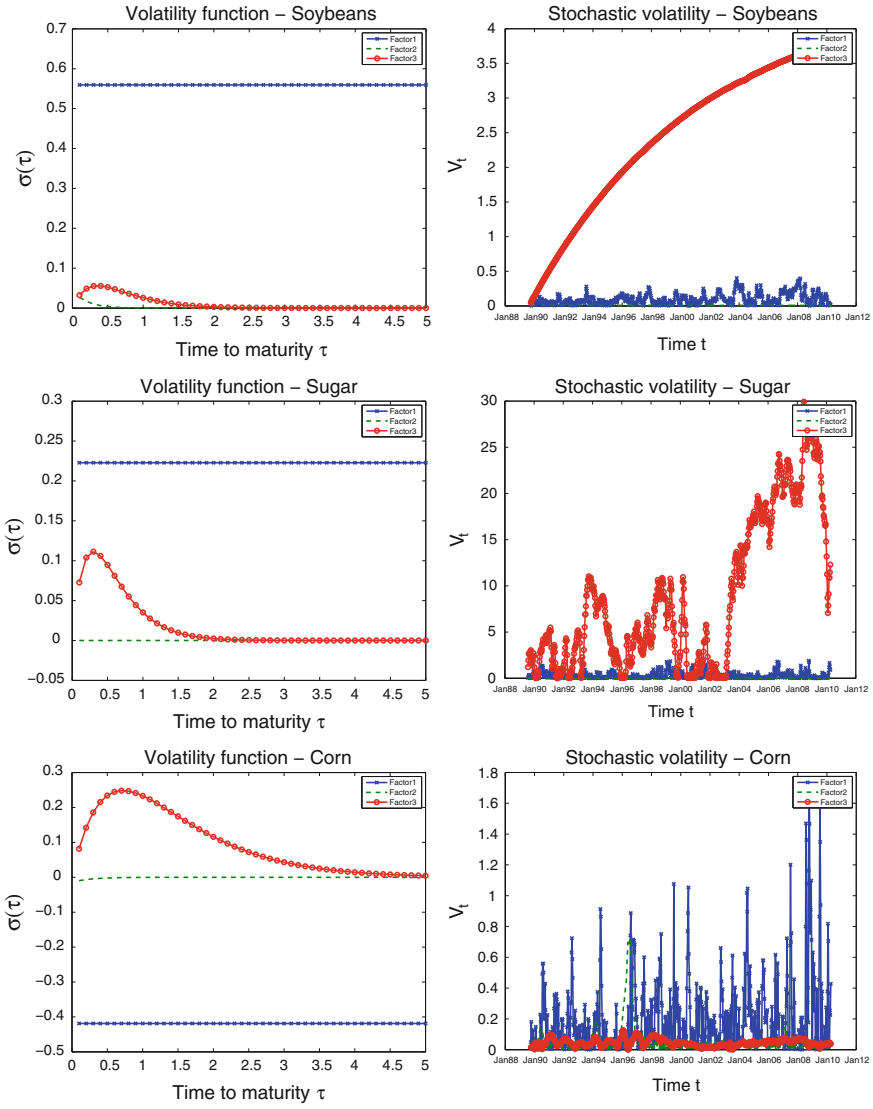

Fig. 4 Estimated volatility factors for agricultural futures prices

volatility state factor $\mathbf{V_t^2}$ is more persistent (reverts slower to the mean) and far more volatile compared to the hump-shaped state factor $\mathbf{V_t^3}$, see also the middle panels of Fig. 3. Therefore, the short-term volatility is the more influential volatility factor, followed by the medium-term and the long-term volatility factor. The correlation coefficient between volatility and futures price returns are −0.8337 and 0.95, respectively. These correlations imply that crude oil futures contracts can hedge more

Table 5 Contribution and shape of volatility factors

	Gold (%)		Crude Oil (%)		Natural Gas (%)	
σ_1	99.98	Long-term	6.73	Long-term	92.65	Long-term
σ_2	0.01	Exp	65.43	Exp	1.36	Exp
σ_3	0.01	Hump	27.84	Hump	5.99	Hump
	Soybean		Sugar		Corn	
σ_1	79.54	Long-term	70.68	Long-term	93.34	Long-term
σ_2	0.00	Exp	0.00	Exp	0.01	Exp
σ_3	20.36	Hump	29.32	Hump	6.65	Hump

The table reports the contribution of each volatility factor to the total variance for the three-factor model. The three volatility structures imposed by the model are long-term volatility (long-term), exponential decaying volatility (exp), and hump-shaped volatility (hump)

efficiently medium-term crude oil futures volatility rather than short-term crude oil futures volatility. The opposite signs of the correlation coefficients in these two dominant volatility factors verify the mixed return-volatility relation, a positive return-volatility relation as explained by the inventory effect, see Ng and Pirrong (1994) and a negative return-volatility relation as explained by the volatility feedback effect, see Salisu and Fasanya (2013).

4.4 Natural Gas Volatility

The parameter estimates for natural gas, see Table 4, demonstrate that the long-term volatility factor dominates in the natural gas futures market. More specifically, from Table 5, the first volatility factor contributes 92.65 % to the total variance. The second most contributing factor is the hump-shaped volatility factor with 6 % to the total variance. The leading volatility state factor $\mathbf{V_t^1}$ is not persistent (reverts quickly to the mean) and very volatile, see also the bottom panels of Fig. 3. Therefore, the long-term volatility is the more influential volatility factor most likely due to the impact of the seasonality associated with this commodity futures market. The correlation coefficient between this volatility factor and futures price returns is 0.3116 implying that natural gas futures contracts alone cannot hedge the natural gas futures volatility. The positive sign of the correlation coefficient of the dominant volatility factor can be explained by the inventory effect, see Ng and Pirrong (1994). Note that the other two factors exhibit a negative return-volatility relation as it can be explained by the volatility feedback effect, see Salisu and Fasanya (2013) but their contribution is marginal.

4.5 Soybean Volatility

From Table 5, two driving volatility factors are present, with the long-term volatility factor contributing 79.54 % of total variance and the hump-shaped volatility factor contributing 20.36 % of the total variance. The hump-shaped state factor $\mathbf{V_t^3}$ is highly persistent (reverts very slowly to the mean) and very volatile compared to the long-term volatility state factor $\mathbf{V_t^1}$ which is not as persistent and less volatile, see the top panels of Fig. 4. The correlation coefficients are 0.7338 for the long-term volatility factor and -0.2922 for the hump-shaped. Thus, soybean futures contracts alone cannot hedge medium-term volatility, yet they can hedge more efficiently the long-term volatility. Moreover, the more contributing long-term volatility factor involves a positive return-volatility relation, while the hump-shaped volatility factor entails a negative return-volatility relation.

4.6 Sugar Volatility

The analysis reveals two main driving volatility factors. The most contributing (with 70.68 % of total variance) is the long-term volatility factor and the second contributing volatility factor (with 29.32 % of the total variance) is the hump-shaped volatility factor, see Table 5. The long-term volatility state factor $\mathbf{V_t^1}$ is not as persistent (reverts quickly to the mean) but equally volatile compared to the hump-shaped state factor $\mathbf{V_t^3}$, see also the middle panels of Fig. 4. The correlation coefficient between the associated volatility factors and futures price returns are both negative and close to -1, thus sugar futures contracts can hedge most of the sugar futures volatility, and a negative return-volatility relation is implied.

4.7 Corn Volatility

The parameter estimates for corn from Table 4 demonstrate that the long-term volatility factor dominates in the corn futures market. More specifically, from Table 5, the long-term volatility factor contributes 93.34 % to the total variance. The second contributing factor is the hump-shaped volatility factor with 6.65 % to the total variance. The leading volatility state factor $\mathbf{V_t^1}$ is highly persistent (reverts slowly to the mean) and very volatile, see also the bottom panels of Fig. 4. Therefore, the long-term volatility is the more influential volatility factor in the corn commodity futures market. The correlation coefficient between the associated volatility factors and futures price returns are positive and high, thus corn futures contracts can hedge the corn futures volatility. The positive sign of the correlation coefficient of the two dominant volatility factors implies a positive return-volatility relation, see Ng and Pirrong (1994) and relates to the inventory effect.

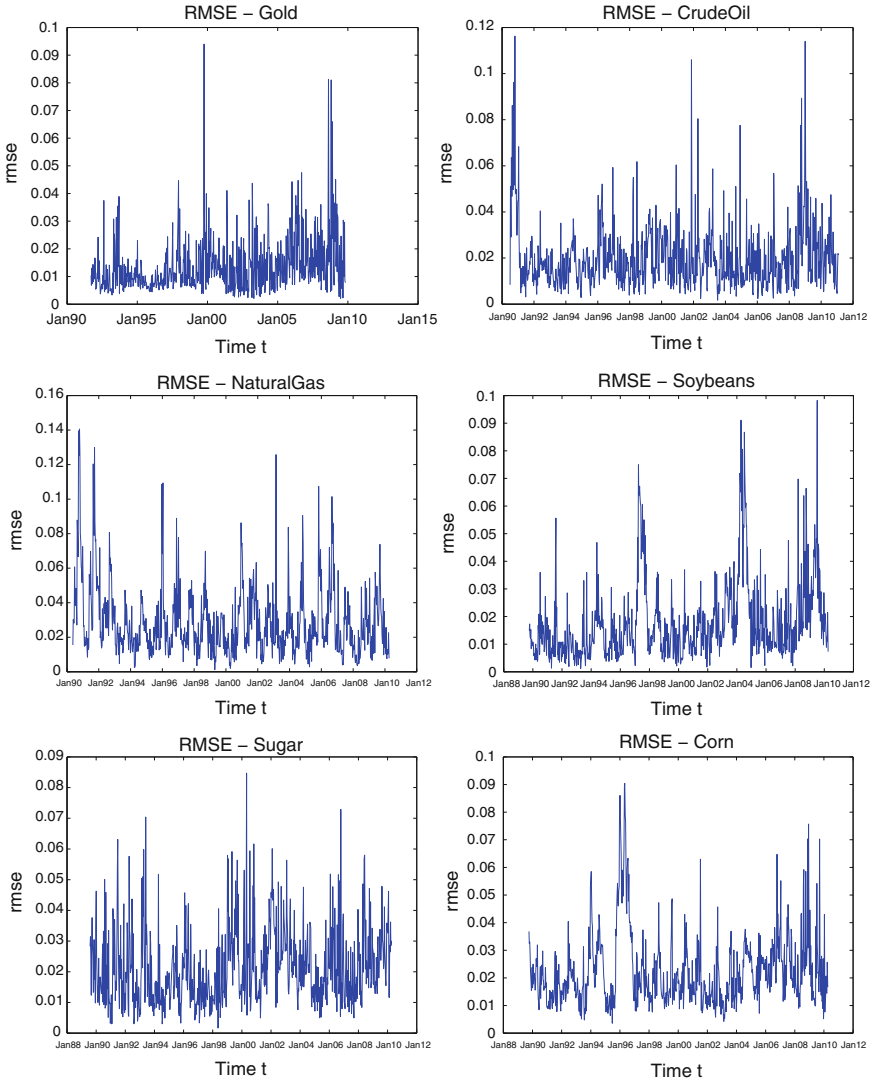

Fig. 5 Model goodness of fit

4.8 Model Fit

For illustrative purposes, the RMSEs of the percentage differences between actual futures prices and fitted futures prices to the proposed three-factor model are depicted in Fig. 5. For all commodities, the overall goodness of fit is satisfactory with most RMSE ranging between 2 and 4 % and occasionally reaching 10 % (only for natural gas the maximum RMSEs reached occasionally 14 %). As natural gas futures prices

are subject to strong seasonal effects, alternate model specifications—e.g. higher dimensions of volatility factors, seasonal dynamics or regime switching volatility schemes—would have the potential to better capture the natural gas futures price dynamics. This also applies for the agricultural commodities, where the RMSEs are relatively low because of persistent seasonality effects. For crude oil and gold, the model fit is reasonable, with exceptional cases being associated with some major socio-economic effects such as the Gulf war and Global Financial Crisis (GFC).

5 Conclusion

In this paper, a tractable forward price model with stochastic volatility is proposed and an empirical study is carried out to analyze the volatility of the most liquid commodity futures markets, including the gold, crude oil, natural gas, soybeans, sugar and corn market. The model allows distinct volatility structures, including exponential decaying, hump-shaped and infinite maturity and can potentially gauge the impact of short-term, medium-term and long-term variation.

The study shows that for most of the commodities futures markets, at least two of the volatility structures are present (with varying levels of persistence and intensity). The long-term volatility is the dominant stochastic volatility factor for most commodities, except crude oil where exponential and hump-shaped volatility factors are contributing more. The extent to which the volatility can be hedged by futures contracts varies across commodities, with futures contracts being least capable to hedge volatility in the crude oil futures market, natural gas futures market and soybeans futures market.

Overall the proposed model provides a relatively good fit for these six commodities, even though the distinctive characteristics of each market is not properly accounted for. A comparative investigation between models with different specifications and their fit performance will better reflect on the quality of the model fit, a study that has been left for further research.

Under the proposed model, option prices can be obtained quasi-analytically, consequently the model can also be estimated by fitting to commodity futures options. Since options are volatility sensitive derivative instruments, this estimation study has the potential to provide constructive and insightful findings about the volatility in commodity futures markets, as well as volatility hedging effectiveness.

Acknowledgments We thank an anonymous referee for helpful suggestions and detailed feedback. The authors also wish to thank the Australian Research Council for financial support in relation to the data purchase (DP 1095177, The Modelling and Estimation of Volatility in Energy Markets).

Appendix: Proof of Theorem 1

We consider the process $X(t,T) = \ln F(t,T,\mathbf{V_t})$, where the forward price dynamics are given by (1) with the volatility specifications (2). Then an application of the Ito's formula derives

$$F(t,T,\mathbf{V_t}) = F(0,T)\exp\left[\sum_{i=1}^{3}\int_0^t \sigma_i(s,T,\mathbf{V_s})dW_i(s) - \frac{1}{2}\sum_{i=1}^{3}\int_0^t \sigma_i^2(s,T,\mathbf{V_s})ds\right]. \tag{14}$$

By introducing the state variables

$$\phi_1(t) = \int_0^t \sqrt{\mathbf{V_s^1}}dW_1(s), \qquad \mathsf{x}_1(t) = \int_0^t \mathbf{V_s^1}ds$$

$$\phi_2(t) = \int_0^t e^{-\eta_2(t-s)}\sqrt{\mathbf{V_s^2}}dW_2(s), \qquad \mathsf{x}_2(t) = \int_0^t e^{-2\eta_2(t-s)}\mathbf{V_s^2}ds$$

$$\phi_3(t) = \int_0^t e^{-\eta_3(t-s)}\sqrt{\mathbf{V_s^3}}dW_3(s), \qquad \phi_4(t) = \int_0^t (t-s)e^{-\eta_3(t-s)}\sqrt{\mathbf{V_s^3}}dW_3(s)$$

$$\mathsf{x}_3(t) = \int_0^t e^{-2\eta_3(t-s)}\mathbf{V_s^3}ds, \qquad \mathsf{x}_4(t) = \int_0^t (t-s)e^{-2\eta_3(t-s)}\mathbf{V_s^3}ds,$$

$$\mathsf{x}_5(t) = \int_0^t (t-s)^2 e^{-2\eta_3(t-s)}\mathbf{V_s^3}ds.$$

and performing some basic manipulations, Eq. (14) can be expressed as (7).

References

Baur, D. G. (2012). Asymmetric volatility in the gold market. *The Journal of Alternative Investments*, *14*(4), 26–38.

Björk, T., Landén, C., & Svensson, L. (2004). Finite dimensional realisations for stochastic volatility forward rate models. *The Royal Society—Proceedings: Mathematical, Physical and Engineering Sciences*, *460*(2041), 53–83.

Chiarella, C., & Kwon, O. (2001). Forward rate dependent Markovian transformations of the Heath-Jarrow-Morton term structure model. *Finance and Stochastics*, *5*(2), 237–257.

Chiarella, C., Kang, B., Nikitopoulos, C. S., & Tô, T. (2013). Humps in the volatility structure of the crude oil futures market: New evidence. *Energy Economics*, *40*, 989–1000.
Clewlow, L., & Strickland, C. (2000). *Energy Derivatives: Pricing and Risk Management* (1st ed.). London: Lacima Publications.
Dai, Q., & Singleton, K. (2000). Specification analysis of affine term structure models. *Journal of Finance*, *55*, 1943–1978.
Doran, J. S., & Ronn, E. I. (2008). Computing the market price of volatility risk in the energy commodity market. *Journal of Banking and Finance*, *32*, 2541–2552.
Duffie, D. (2001). *Dynamic Asset Pricing Theory*. Princeton, NJ: Princeton University Press.
Duffie, D., & Kan, R. (1996). A yield-factor model of interest rates. *Mathematical Finance*, *6*(4), 379–406.
Gibson, R., & Schwartz, E. (1990). Stochastic convenience yield and the pricing of oil contingent claims. *Journal of Finance*, *45*, 959–976.
Heath, D., Jarrow, R., & Morton, A. (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claim valuation. *Econometrica*, *60*(1), 77–105.
Heston, S. L. (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.
Korn, O. (2005). Drift matters: An analysis of commodity derivatives. *Journal of Futures Markets*, *25*, 211–241.
Miltersen, K. (2003). Commodity price modelling that matches current observables: A new approach. *Quantitative Finance*, *3*, 51–58.
Miltersen, K. R., & Schwartz, E. (1998). Pricing of options on commodity futures with stochastic term structure of convenience yields and interest rates. *Journal of Financial and Quantitative Analysis*, *33*(1), 33–59.
Nelson, C., & Siegel, A. (1987). Parsimonious modelling of yield curves. *Journal of Business*, *60*, 473–489.
Ng, V. K., & Pirrong, S. C. (1994). Fundamentals and volatility: Storage, spreads, and the dynamics of metals prices. *Journal of Business*, *67*(2), 203–230.
Protter, P. (2004). *Stochastic Integration and Differential Equations*. New York: Springer.
Salisu, A. A., & Fasanya, I. O. (2013). Modelling oil price volatility with structural breaks. *Energy Policy*, *52*(C), 554–562.
Schwartz, E. S & Trolle, A. B. (2009a). Unspanned stochastic volatility and the pricing of commodity derivatives. *Review of Financial Studies, 22*(11), 4423–4461.
Tokic, D. (2011). Rational destabilizing speculation, positive feedback trading and the oil bubble on 2008. *Energy Policy*, *39*(4), 2051–2061.

A Multi-factor Structural Model for Australian Electricity Market Risk

John Breslin, Les Clewlow and Chris Strickland

1 Introduction

In this paper, we develop a general framework for the modelling of Australian electricity market risk based on the structural relationships in the market. The model framework is designed to be consistent with temperature and load mean forecasts, market forward price quotes, the dependence of load on temperature, and the dependence of price on load. The primary use of the model is for the accurate evaluation of the market risk of an electricity generation and retail company but it can also be used for the valuation of electricity market derivatives and assets. We demonstrate the application of our framework to the Australian National Electricity Market (NEM) by estimating the model using recent historical data from the NEM and then simulating the market using the estimated model.

Historically, the majority of published work on modelling electricity prices has taken the traditional finance approach of applying stochastic processes directly to the spot price (see for example Clewlow and Strickland (2000); Weron et al. (2004); Cartea and Figueroa (2005); Geman and Roncoroni (2006); Benth et al. (2007), (2008); Barndorff-Nielsen et al. (2010); Klüppelberg et al. (2010); Kholodnyi (2011); Veraart and Veraart (2013)). However, this approach has some fundamental disadvantages. Observation suggests that the spot price of electricity is linked to other key market variables, such as the temperature and electricity demand, and with a non-linear relationship which cannot be accurately captured by simple correlations. Furthermore, the dynamics of electricity spot prices are difficult to capture with a

J. Breslin · L. Clewlow · C. Strickland
Lacima, Level 32, 1 Market Street, Sydney, NSW 2000, Australia
e-mail: John.Breslin@lacimagroup.com

L. Clewlow (✉)
e-mail: les@lacimagroup.com

C. Strickland
e-mail: chris@lacimagroup.com

R. Dieci et al. (eds.), *Nonlinear Economic Dynamics and Financial Modelling*,
DOI: 10.1007/978-3-319-07470-2_19,

model which considers the spot price dynamics without any reference to the market structure.

An alternative "structural" or "hybrid" approach is based on jointly modelling the electricity price as a function of supply and demand based market variables and the structural links between them. Examples of models in this category are reviewed in Carmona and Coulon (2013). One of the first papers on this approach was Barlow (2002) which considered demand as the only driving variable with the spot price being a power function of an Ornstein-Uhlenbeck diffusion process. Later this approach was extended to consider alternative drivers: fuel prices (Carmona et al. 2012; Pirrong and Jermakyan 2008), capacity (Burger et al. 2004; Cartea and Villaplana 2008), fuel prices and capacity (Aïd et al. 2012; Coulon and Howison 2009). The advantage of this approach is it makes use of the key information which is available in electricity markets such as the demand factors, fuel prices, and generation capacity. In this paper, we propose a model which uses temperature as a key demand factor and demand as a key driver of the electricity price.

The paper is organised as follows: In Sect. 2, we introduce the model and discuss its key features. Section 3 presents the results of applying the model to the Australian NEM. Finally, Sect. 4 contains our conclusions.

2 The Model

The Australian NEM consists of five regions, or nodes, which correspond to the South-Eastern Australian states: Queensland (QLD), New South Wales (NSW), Victoria (VIC), South Australia (SA), and Tasmania (TAS). The Australian Energy Market Operator (AEMO) has responsibility for power system operations (i.e. security of supply and system reliability) as well as market operations. Ignoring the interconnection between regions, the electricity spot price at each node is set every 5 min(a Dispatch Interval), with the six prices for each half-hour averaged to provide market prices every 30 min (a Trading Interval). Generator offers for supply at specific prices are matched against demand in each dispatch interval, and the generators are dispatched by AEMO according to a "least cost" optimisation algorithm, taking in to account the physical constraints on the generators and the transmission network.[1]

Under normal market conditions, the spot price will typically be in the range \$30/MWh–\$100/MWh. The price reflects the marginal cost of generation, so the price will generally rise along with the regional demand, as more expensive generators are dispatched to meet the demand. We refer to this as the underlying spot price. But when transmission network constraints bind or another event occurs (such as a forced outage of a generator) the price can rise rapidly well above these levels, sometimes to the market price cap of \$13,100/MWh and returning, usually within a

[1] In reality the dispatch algorithm co-optimises the dispatch for the energy as well as ancilliary service markets, and across all regions, taking into account interconnector flows and constraints.

few hours, to the "normal" range. We refer to these sudden and rapid price changes as price spikes.

In terms of modelling the NEM spot price for risk measurement and analysis of simple derivatives such as futures, swaps and swaptions, a stochastic model of spot prices such as a mean reverting jump diffusion process can capture the key characteristics of the market. Such a model can be sufficiently rich to represent the key risks in their market exposure, but also simple enough to allow for analytical pricing of derivatives for efficient valuation and Value-at-Risk calculations.

However, in general, NEM market participants have diverse portfolios consisting of a combination of derivative contracts, generation assets, retail customers and environmental/carbon liabilities. There is a well-defined relationship between these different portfolio components. For example, on a day when temperatures are high we expect an above average electricity demand (due to additional air conditioning load), which in turn will lead to additional (more expensive) generation being dispatched and hence higher than average spot price outcomes. From a risk measurement perspective, it is important to capture these relationships between different components of the portfolio. In the case of a participant with both generation assets and retail customers, in this example the cost to supply the additional electricity demand from their customers should be (partially) offset by additional generation revenue, so capturing this in the model would avoid overstating the risk that may occur if the model does not properly model that relationship. These structural relationships can be hard to capture using purely stochastic models, so a more practical approach is to use a hybrid model, so called because models of this type explicitly recognise the structural relationships between key variables (like temperature, demand and price) as well as a stochastic components to capture the variability in these variables and the impact of variables which can't be easily modelled.

The modelling framework that we propose in this paper is illustrated in Fig. 1.

The temperature is modeled as an independent source of risk around a mean temperature forecast. The temperature feeds into the load model via a time-dependent functional relationship between the temperature and load. The load also has an additional non-temperature dependent source of risk added to the temperature-dependent load. The underlying spot price model has a functional dependence on the load and an additional non-load dependent source of risk. Finally, there is the price spike model driven by a Poisson process which is dependent on the level of the load. In the following sections, we analyse the NEM data and propose specific models for each of the components in the model framework. The Brownian motions which provide the sources of risk may all be correlated.

2.1 The Temperature Model

In order to determine an appropriate model for the regional temperature, we analyse the seasonal variation and intraday variation of the half-hourly historical temperature data from 1 January 2006 to 31 December 2012 for the Bankstown

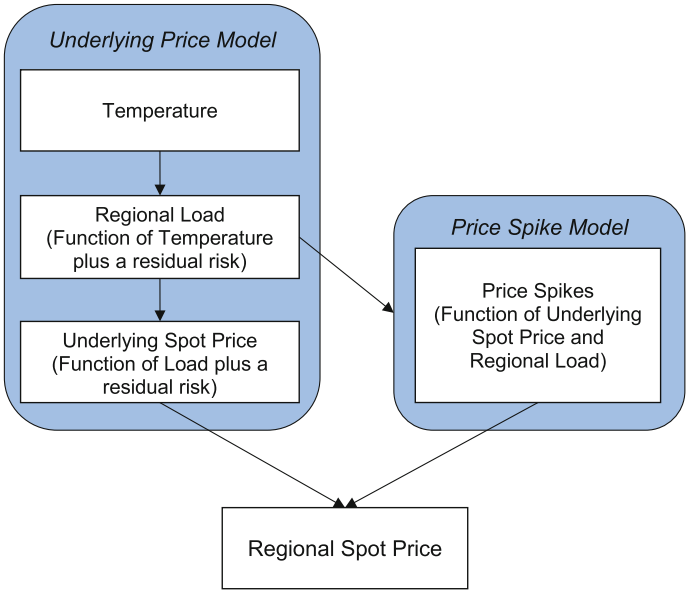

Fig. 1 The structural relationships in the model

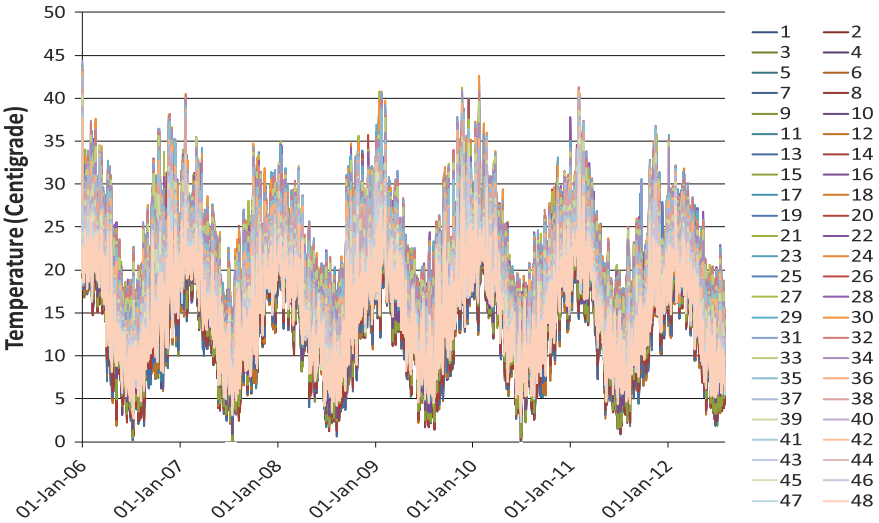

Fig. 2 Half-hourly temperature at Bankstown Airport weather station

Airport weather station. This weather station is typically chosen by market participants as being representative of the NSW regional temperature. Figure 2 shows a time series plot of the data.

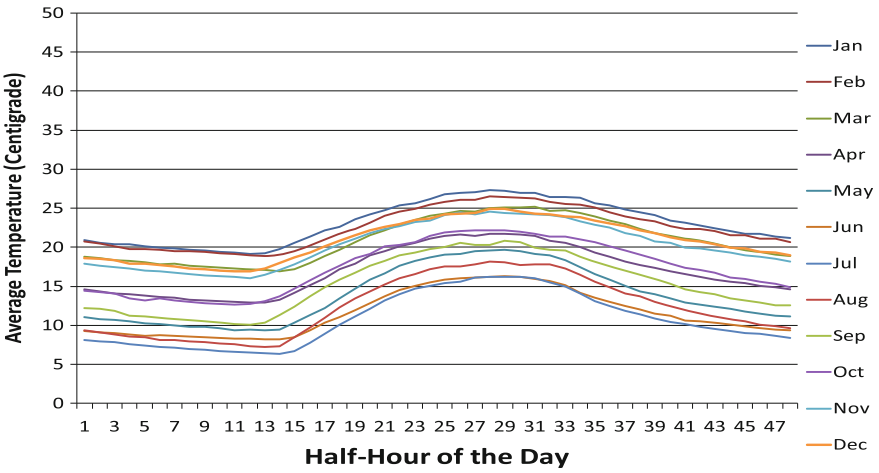

Fig. 3 Half-hourly average temperature for each month at Bankstown Airport weather station

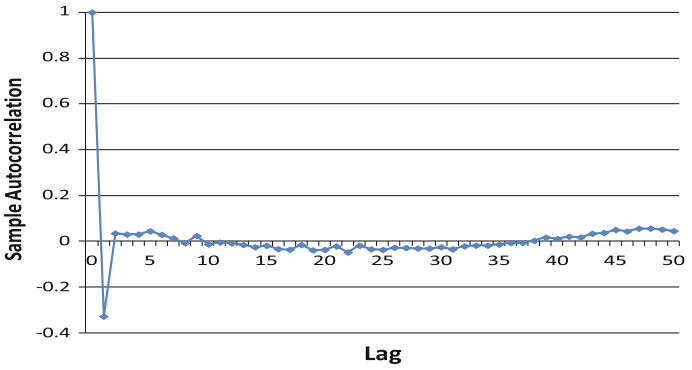

Fig. 4 Auto-correlation of the half-hourly temperature for Bankstown Airport weather station

Each half-hour is plotted as a separate line to show the range of intra-day temperature by the colour range as well as the seasonal variation. Figure 3 shows the average intra-day temperature profile for each month. The maximum standard error over all the mean estimates for each period and each month is 0.5 so the estimated profiles accurately reflect the intraday temperature profile.

The auto-correlation structure of the temperature changes is shown in Fig. 4. This indicates that a simple mean reverting model with normally distributed shocks as previously proposed by Alaton et al. (2002), Benth and Saltyte-Benth (2007) is appropriate.

Given the above analysis of the temperature data and the previous work discussed in Sect. 1, we propose the following model for the evolution of the half-hourly regional temperature ($H(t)$):

$$\mathrm{d}H(t) = \Big[\frac{\partial \bar{H}(t)}{\partial t} - \int_{t_0}^{t} \sigma_H(u,t)\frac{\partial \sigma_H(u,t)}{\partial t}\mathrm{d}u + \int_{t_0}^{t} \frac{\partial \sigma_H(u,t)}{\partial t}\mathrm{d}z_H(u)\Big]\mathrm{d}t + \sigma_H(t)\mathrm{d}z_H(t) \tag{1}$$

$$\sigma_H(u,t) = \sigma_H(u)\exp(-\alpha_H(t-u)) \tag{2}$$

where $\bar{H}(t)$ is the time-dependent deterministic mean forecast of the daily average temperature, $\sigma_H(t)$ is the time dependent deterministic volatility of the temperature, α_H is the mean reversion rate of the temperature, $z_H(t)$ is a Brownian motion. This model is a generalisation of the model introduced by Alaton et al. (2002). The generalisation uses the results in Sect. 8.5 of Clewlow and Strickland (2000) to allow the specification of the temperature forecast which is exactly equal to the mean temperature. The form of the volatility function in Eq. (2) is equivalent to mean reverting dynamics in the temperature as shown in Clewlow and Strickland (1999). This model also allows for generalisations of the volatility specification as shown in Clewlow and Strickland (2000). We use this same general model specification for the residual in all the components of the model. This has the advantage of limiting the effective complexity of the model whilst still maintaining its flexibility.

2.2 The Regional Load Model

The electricity demand in any region of the NEM is partly dependent on retail and commercial customer demand for heating and cooling. The demand profile over the calendar year and intra-day therefore has a strong relationship with the month of year and time of day as well as the prevailing weather conditions. Figure 5 shows a time series plot of the NSW regional electricity demand for the period from 1 July 2006 to 31 December 2012. This shows that the seasonal profile has higher average demand and lower volatility in the cooler months of April to September and lower average demand but higher volatility in the warmer months October through to March.

Each half-hour is plotted as a separate line to show the range of intra-day demand by the colour range as well as the seasonal variation. Figure 6 shows the average intra-day demand profile for each month.

The maximum standard error is 103 so the smoothness of the profiles accurately reflects the intraday profile. The intra-day demand profile has a relatively simple shape for the warmer months October through to March with the demand rising smoothly from around 5–10 am where it levels off before smoothly declining from around 4 pm through to 5 am. For the cooler months April to September the profile is similar but there are pronounced peaks at 10 am and 7 pm reflecting the high demand of electricity based heating in the morning and evening. Figures 7, 8, 9 and 10 show the relationship between the half-hourly NSW regional demand and temperature for January, April, July and October business days 8–10 am, business days 4–6 pm, non-

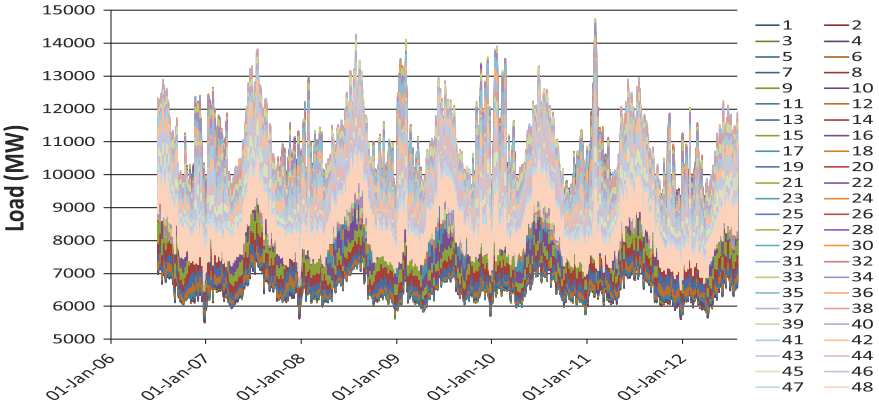

Fig. 5 Half-hourly NSW regional demand

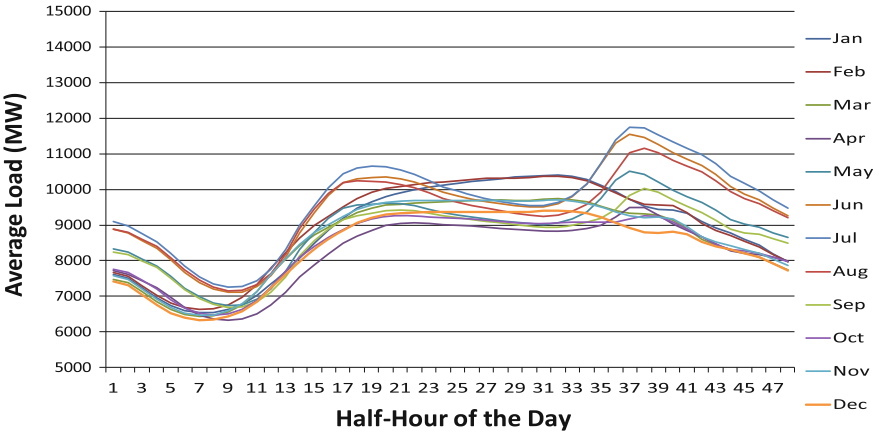

Fig. 6 Half-hourly average NSW regional demand for each month

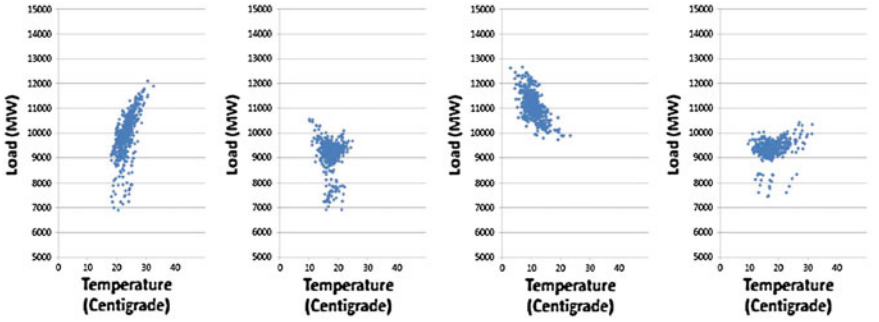

Fig. 7 NSW regional demand as a function of temperature in January, April, July and October business days 8–10 am

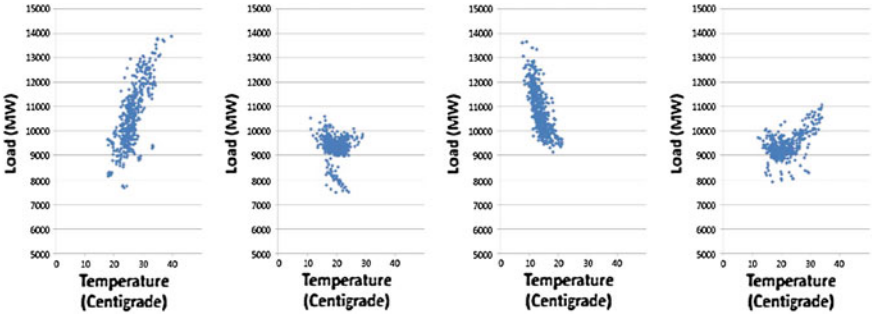

Fig. 8 NSW regional demand as a function of temperature in January, April, July and October business days 4–6 pm

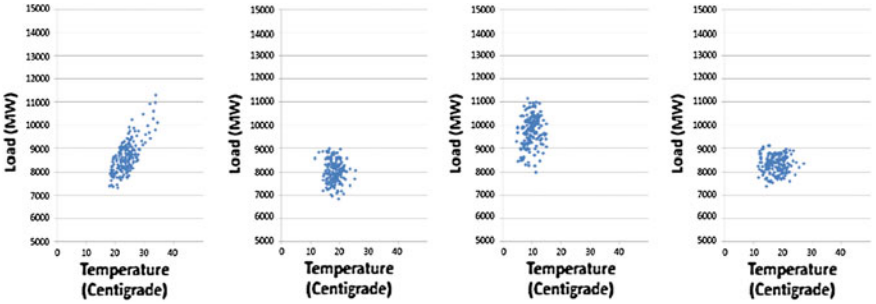

Fig. 9 NSW regional demand as a function of temperature in January, April, July and October non-business days 8–10 am

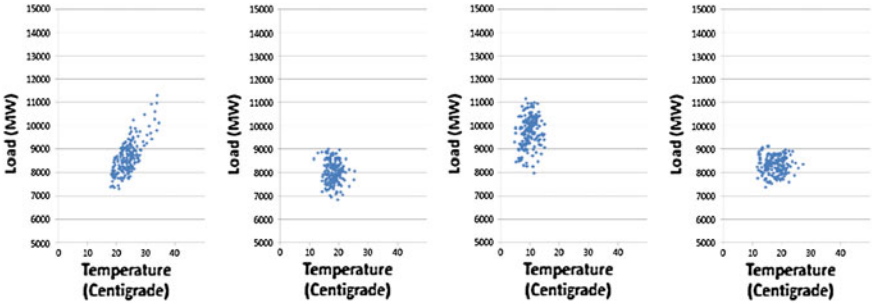

Fig. 10 NSW regional demand as a function of temperature in January, April, July and October non-business days 4–6 pm

business days 8–10 am, and non-business days 4–6 pm. Each point on the charts represents an observed temperature and load pair in a particular half hour.

The form of the variation of regional demand on the ambient temperature depends on the seasonal period, day of the week and time of day. Where the ambient temperature is above or below 18.5 °C, the relationship is quite linear and where the

temperature variation is centred on 18.5 °C the relationship shows a strong U shape. This suggests that the relationship can be modelled using a U-shaped functional form with a minimum at 18.5 °C.

Based on the above observations, we propose that the regional load ($L(t)$) can be modeled as the sum of a load which is a quadratic function of the regional temperature ($H(t)$) and a non-temperature dependent load component ($\tilde{L}(t)$):

$$L(t) = a(t) + c(t)(H(t) - H_{\min})^2 + \tilde{L}(t) \tag{3}$$

where $a(t)$ represents the time varying base load when the temperature is at $H_{\min}$, $c(t)$ is the responsiveness of the temperature-dependent load to the squared difference in the temperature from $H_{\min}$ and $H_{\min}$ is the temperature at which there is there is no demand for heating or cooling. We further propose that the non-temperature dependent load component can be modeled as a mean reverting stochastic process:

$$\mathrm{d}\tilde{L}(t) = \Big[\frac{\partial \ln \bar{\tilde{L}}(t)}{\partial t} - \int_{t_0}^{t} \sigma_{\tilde{L}}(u,t)\frac{\partial \sigma_{\tilde{L}}(u,t)}{\partial t}\mathrm{d}u + \int_{t_0}^{t} \frac{\partial \sigma_{\tilde{L}}(u,t)}{\partial t}\mathrm{d}z_{\tilde{L}}(u)\Big]\mathrm{d}t + \sigma_{\tilde{L}}(t)\mathrm{d}z_{\tilde{L}}(t) \tag{4}$$

$$\sigma_{\tilde{L}}(u,t) = \sigma_{\tilde{L}}(u)\exp(-\alpha_{\tilde{L}}(t-u)) \tag{5}$$

where $\bar{\tilde{L}}(t)$ is the time-dependent deterministic mean forecast, $\sigma_{\tilde{L}}(t)$ is the time-dependent deterministic volatility, $\alpha_{\tilde{L}}$ is the mean reversion rate and $z_{\tilde{L}}(t)$ is a Brownian motion. The regional load ($L(t)$) can be made consistent with a time-dependent mean load forecast ($\bar{L}(t)$) by adjusting the time varying base load ($a(t)$) such that the average of the simulated loads at each time step are equal to the load forecast:

$$\Delta a(t) = \bar{L}(t) - \frac{1}{N}\sum_{i=1}^{N} L_i(t) \tag{6}$$

where $\Delta a(t)$ is the adjustment to the time varying base load ($a(t)$) and $i = 1, \ldots, N$ are the simulation indices.

2.3 The Regional Reference Price Model

In Sect. 2, we described how the regional reference price depends on the lowest price which allows the demand to be met by the generation offers to supply. This means that as demand increases such that more expensive generation is required to meet that demand the regional reference price will increase. Figure 11 shows the historical relationship between the regional reference price and the regional demand for the months of January, April, July and October between 2008 and 2012. Based

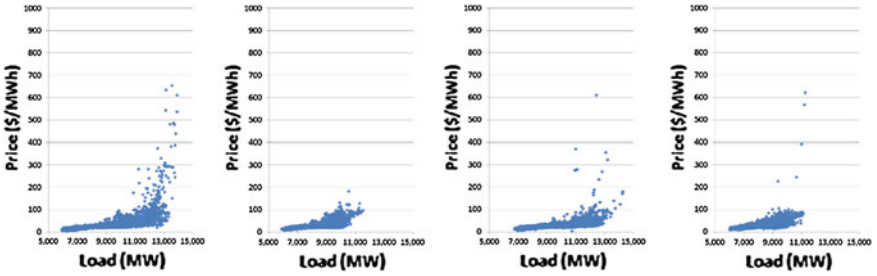

Fig. 11 NSW regional price as a function of demand in January, April, July and October 2008–2012

on the observed relationship, we propose that the price can be decomposed into a piecewise linear function of load plus a residual random component for prices below \$100/MWh and an additional price spike component for loads greater than around 10,000 MW.

We therefore propose the non-spiking or normal price ($S_N(t)$) can be modeled as the sum of a piecewise linear price function ($PLP(t, L(t))$) of the regional load ($L(t)$) and a non-load dependent component ($\tilde{S}(t)$):

$$S_N(t) = PLP(t, L(t)) + \tilde{S}(t) \tag{7}$$

The non-load dependent price component can be modeled as a mean reverting stochastic process:

$$\mathrm{d}\tilde{S}(t) = \left[\frac{\partial \ln \bar{\tilde{S}}(t)}{\partial t} - \int_{t_0}^{t} \sigma_{\tilde{S}}(u,t)\frac{\partial \sigma_{\tilde{S}}(u,t)}{\partial t}\mathrm{d}u + \int_{t_0}^{t} \frac{\partial \sigma_{\tilde{S}}(u,t)}{\partial t}\mathrm{d}z_{\tilde{S}}(u)\right]\mathrm{d}t + \sigma_{\tilde{S}}(t)\mathrm{d}z_{\tilde{S}}(t) \tag{8}$$

$$\sigma_{\tilde{S}}(u,t) = \sigma_{\tilde{S}}(u)\exp(-\alpha_{\tilde{S}}(t-u)) \tag{9}$$

where $\bar{\tilde{S}}(t)$ is the time-dependent deterministic mean forecast, $\sigma_{\tilde{S}}(t)$ is the time-dependent deterministic volatility, $\alpha_{\tilde{S}}$ is the mean reversion rate and $z_{\tilde{S}}(t)$ is a Brownian motion.

For the price spikes ($S_{PS}(t)$), we use a mean reverting Poisson process conditional on the regional load level:

$$\mathrm{d}S_{PS(t)} = \begin{cases} 0 & L(t) < L_{PS}(t) \\ -\alpha_{PS}S_{PS}(t)\mathrm{d}t + \kappa_{PS}(t)^{+}\mathrm{d}p(t;\phi_{PS}) & L(t) \geq L_{PS}(t) \end{cases} \tag{10}$$

where α_{PS} is the mean reversion rate of the price spikes, $\kappa_{PS}(t)$ is the random price spike size which is normally distributed with mean zero and time-dependent standard deviation $\gamma_{PS}(t)$, $\phi_{PS}(t)$ is the time-dependent deterministic annualised arrival rate

of the Poisson process $\mathrm{d}p(t)$, the $+$ superscript on $\kappa_{PS}(t)$ indicates only the positive jumps are taken and $L_{PS}(t)$ is the time-dependent deterministic price spike trigger level below which no price spikes occur.

The total regional reference price ($S(t)$) is given by:

$$S(t) = S_N(t) + S_{PS}(t) \tag{11}$$

The total regional reference price ($S(t)$) can be made consistent with an initial (time 0) forward price curve ($F(0, t)$) by proportionally adjusting the simulated total regional reference price ($S(t)$) such that the average at each time step is equal to the forward price:

$$S_{A,i}(t) = S_i(t)\Delta S(t) \tag{12}$$

$$\Delta S(t) = \frac{F(0, t)}{\frac{1}{N}\sum_{i=1}^{N} S_i(t)} \tag{13}$$

where $S_{A,i}(t)$ is the adjusted total regional reference price for simulation $i = 1, \ldots, N$, $S_i(t)$ is the unadjusted total regional reference price for simulation i, $\Delta S(t)$ is the proportional adjustment and $i = 1, \ldots, N$ are the simulation indices.

3 Numerical Results

In this section, we estimate the model and then compare the statistical properties of the observable simulated model variables with those of the data.

3.1 The Data

We use half-hourly data from the NSW region of the NEM. The temperature data is for the Bankstown Airport weather station, which is typically chosen by market participants as being the most representative for modeling the NSW regional temperature. The regional demand and price data is from the NEM website.

3.2 Estimation

3.2.1 Temperature Model

We use the half-hourly historical temperature data from 1 January 2006 to 31 December 2012 for the Bankstown Airport weather station. We first estimate a piecewise

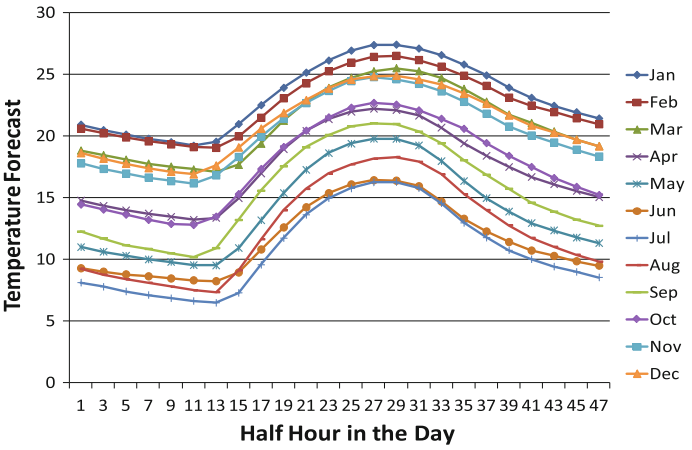

Fig. 12 Half-hourly temperature forecast for Bankstown Airport weather station

linear hourly forecast curve function $\bar{H}(t)$ from the average temperature for each month and hour as shown in Fig. 12. In order to estimate the mean reversion rate and volatility functions, we first detrend the data by subtracting the estimated mean from the historical temperature. To estimate the mean reversion rate we note that, using Eq. (1), the discrete time model for the detrended temperature is[2]:

$$\Delta H(t) = c_0 + c_1 H(t) + \sigma_H(t) dz(t) \tag{14}$$

where $c_1 = -\alpha_H \Delta t$. Regressing the detrended temperature changes against the detrended temperature levels gives an estimate of the reversion rate:

$$\alpha_H = 306.057$$

We estimate the temperature volatility function from the detrended temperature data as the annualised sample standard deviation of the temperature changes in "4 hourly" buckets for each month. The results are shown in tabular form in Table 1.

3.2.2 The Regional Load Model

We use the half-hourly NSW regional electricity demand for the period from 1 July 2006 to 31 December 2012 together with the corresponding temperature from Sect. 3.2.1. We estimate the parameters $a(t)$ and $c(t)$ by assuming a value for $H_{\min} = 18.5$ and regressing the demand $(L(t))$ against the square of the differ-

[2] See Sect. 2.9 in Clewlow and Strickland (1999) for details.

Table 1 NSW regional temperature volatility function

	Hr1–4	Hr5–8	Hr9–12	Hr13–16	Hr17–20	Hr21–24
January	51.8397	79.41113	88.26753	92.49929	72.64073	55.78484
February	51.53123	66.56551	88.57734	103.5264	71.29501	56.12818
March	51.43421	69.99356	81.35919	86.48897	64.26129	64.55958
April	57.65264	92.71098	86.49076	83.37752	67.30978	68.2138
May	70.38631	100.7211	81.88056	70.56995	91.16441	77.76141
June	69.76199	85.14521	80.43355	73.7549	82.84931	74.7811
July	84.78846	96.64674	89.42756	74.8942	87.49551	86.64252
August	74.65451	119.9571	95.05989	82.45058	82.36089	82.51347
September	76.94269	119.409	97.13468	83.4198	74.65842	93.57401
October	61.62445	100.8084	91.79875	96.28022	69.67073	73.72769
November	57.17481	86.44402	95.42086	102.0853	72.58685	61.2784
December	54.68051	82.40801	83.04329	99.62553	69.26126	59.72474

Table 2 NSW regional load model base load function $a(t)$

	Hr1–4	Hr5–8	Hr9–12	Hr13–16	Hr17–20	Hr21–24
January	6735.91	7618.79	9291.89	9381.29	9035.72	8158.80
February	6817.91	8034.46	9560.15	9473.18	9212.90	8231.10
March	6792.71	7931.97	9301.56	9102.28	8973.21	8027.74
April	6912.86	7192.69	8894.96	8833.10	9099.53	8127.20
May	7478.78	7871.04	9384.53	9081.67	9763.80	8933.73
June	7859.26	8152.13	9937.03	9363.26	10388.66	9667.22
July	8193.61	8486.54	10116.03	9367.24	10355.77	10216.50
August	7903.55	8400.09	9727.63	9271.92	9888.80	9462.26
September	7402.11	7708.02	9327.91	9018.26	9423.44	8616.41
October	7006.37	7667.81	9070.50	8849.37	8980.30	8243.97
November	6844.06	7971.42	9388.96	9209.22	9051.07	8164.09
December	6701.19	7602.08	8983.95	8748.04	8574.18	7995.82

ences $(H(t) - H_{\min})$ for "4 hourly" blocks and for each month. The results are shown in tabular form in Tables 2 and 3.

To estimate the mean reversion rate and volatility function of the residual non-temperature dependent load, we subtract the model values of the temperature-dependent load from the historical load data and then use the same methodology as in Sect. 3.2.1, which gives the following mean reversion estimate:

$$\alpha_{\tilde{L}} = 968.108$$

The residual load volatility function estimates are given in tabular form in Table 4.

Table 3 NSW regional load model quadratic function $c(t)$

	Hr1–4	Hr5–8	Hr9–12	Hr13–16	Hr17–20	Hr21–24
January	18.36	17.93	11.80	9.65	12.24	18.99
February	17.21	17.32	13.45	11.42	12.22	15.46
March	0.00	0.00	13.13	10.92	12.45	18.38
April	2.67	1.33	5.81	3.46	8.77	11.28
May	0.74	0.00	8.93	3.24	15.09	3.40
June	1.34	1.62	7.79	27.08	16.96	3.49
July	0.00	0.00	7.12	26.07	18.42	0.00
August	1.29	0.00	12.54	10.28	25.18	4.87
September	1.38	0.00	0.00	0.54	7.91	6.98
October	0.00	0.00	5.56	5.94	6.52	2.66
November	0.90	0.00	8.04	7.64	8.28	10.24
December	0.00	0.00	11.48	11.31	12.55	16.54

Table 4 NSW regional residual load model volatility function

	Hr1–4	Hr5–8	Hr9–12	Hr13–16	Hr17–20	Hr21–24
January	16326.16	11296.01	14862.76	13746.73	20065.02	13981.04
February	17253.32	6089.364	15157.53	14520.53	19295.87	12464.99
March	16605.26	2800.248	13171.52	11854.15	18978.77	12409.56
April	5043.313	7666.157	13728.24	6782.552	32092.11	15769.11
May	4247.429	14392.97	13017.31	8582.868	46095.06	14658.43
June	4042.978	18356.67	13944.13	11926.71	49868.67	10790.51
July	2894.532	20121.26	18962.07	13120.82	53588.92	10902.32
August	1889.344	17207.58	17890.93	10342.04	53564.29	13617.13
September	2175.913	13434.13	12692.27	7058.399	34081.65	16488.39
October	12882.66	14521.21	11639.84	8417.06	20158.98	12822.62
November	17120.56	8321.531	12295.69	10736.86	16220.48	11962.53
December	16598.01	9458.615	12619.3	12127.87	17347.81	14304.59

3.2.3 The Regional Reference Price Model

We use the half-hourly NSW regional electricity price and demand for the period from 1 January 2008 to 31 December 2011.[3] We set the knot points of the piecewise linear function (Eq. 7) as 10,000, 11,000 and 12,000 MW as this is the load levels at which the transition from base load to higher cost generation occurs. Prices above \$300/MWh are removed for the piecewise linear fitting as these will be modeled by the price spike process. The estimation is performed separately for quarterly, business/non-business and peak/off-peak data. The results are shown in Table 5.

To estimate the mean reversion rate and volatility function of the residual non-load-dependent price we subtract the model values of the load-dependent price from

[3] We omit 2012 from the price model estimation as carbon pricing was introduced into the market in July 2012.

Table 5 NSW piecewise linear regional price model

		Constant		Slope 1		Slope 2		Slope 3		Slope 4	
		Peak	Off peak	Peak	Off peak	Peak	Off peak	Peak	Off Peak	Peak	Off peak
Q1	Business	−14.62	−17.13	0.005	0.005	0.014	0.051	0.023	0.000	0.070	0.000
Q1	Non business	−25.02	−17.16	0.006	0.005	0.026	0.076	0.078	0.000	0.056	0.000
Q2	Business	−15.40	−10.01	0.005	0.005	0.008	−0.004	0.008	0.038	0.047	0.000
Q2	Non business	−26.80	−6.99	0.007	0.004	0.012	−0.041	0.045	0.000	0.000	0.000
Q3	Business	−11.52	−5.96	0.004	0.004	0.005	0.012	0.015	0.000	0.037	0.000
Q3	Non business	−7.61	−6.87	0.004	0.004	0.017	0.032	0.029	0.000	0.106	0.000
Q4	Business	−21.08	−10.49	0.006	0.005	0.011	−0.028	0.063	0.000	0.081	0.000
Q4	Non business	−8.27	−6.01	0.004	0.004	0.052	0.000	0.151	0.000	0.000	0.000

Table 6 NSW regional price model residual volatility function

		Peak	Off peak
Q1	Business day	253.95	237.65
Q1	Non business day	162.40	211.95
Q2	Business day	258.93	246.26
Q2	Non business day	207.07	220.66
Q3	Business day	242.78	244.86
Q3	Non business day	184.98	197.55
Q4	Business day	202.46	224.90
Q4	Non business day	139.14	191.70

the historical price data and then use the same methodology as in Sect. 3.2.1, which gives the following estimate for the reversion rate:

$$\alpha_{\tilde{L}} = 557.102$$

The regional price model residual volatility function estimates are given in tabular form in Table 6.

The mean reversion rate of the price spikes component (α_{PS}) is estimated by using the same approach as Sect. 3.2.1. The price spike standard deviation ($\gamma_{PS}(t)$) and annualised arrival rate ($\phi_{PS}(t)$) are estimated using the methodology of Clewlow and Strickland (1999) Sect. 2.10. We obtain the following reversion rate estimate:

$$\alpha_{PS} = 5857.204$$

The price spikes model arrival rate and standard deviation function estimates are given in tabular from in Table 7.

The off-peak data categories contain zero or only a few samples so we set all the estimates to zero.

3.3 Simulation Results

In this section, we use the model parameters estimated in Sect. 3.2 to simulate the market variables for the period January 2009–December 2012. Figure 13 shows the temperature simulations, Fig. 14 shows the simulated load as a function of the simulated temperature for January, April, July and October months and Fig. 15 shows the historical load against temperature for comparison. The simulations capture the seasonal structural relationship and distributions of the temperature and load. There are some noticeable discrepancies for the extreme values of the temperature. The historical temperature distribution has a positive skew which is not present in the simulations. This leads to a slight under representation of the higher load levels. This can be seen in the January chart. The July chart shows that the temperature simu-

Table 7 NSW regional price spike model parameter estimates

		$\phi_{PS}(t)$		$\gamma_{PS}(t)$	
		Peak	Off peak	Peak	Off peak
Q1	Business day	206.3825	0	1188.83	0
Q1	Non business day	130.122	0	129.708	0
Q2	Business day	156.1464	0	917.1514	0
Q2	Non business day	102.2387	0	116.0713	0
Q3	Business day	155.6113	0	1297.957	0
Q3	Non business day	46.47215	0	123.8917	0
Q4	Business day	101.0725	0	1671.039	0
Q4	Non business day	23.01478	0	135.1728	0

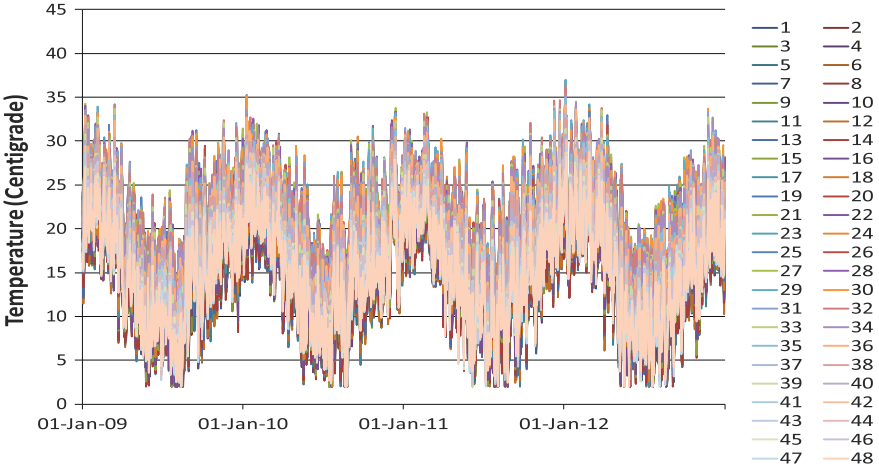

Fig. 13 NSW regional temperature simulation for 2009–2012

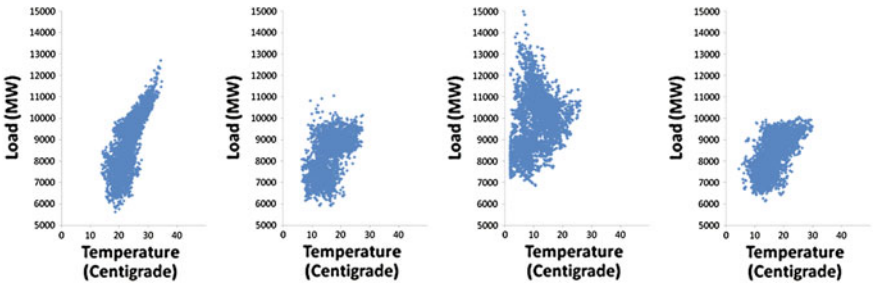

Fig. 14 NSW regional load versus temperature simulation for 2009–2012

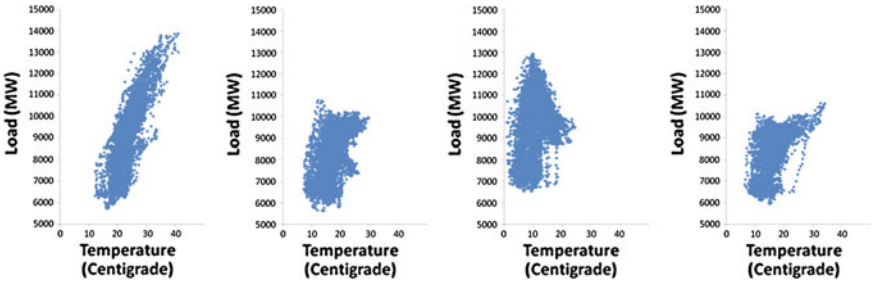

Fig. 15 NSW regional load versus temperature history for 2009–2012

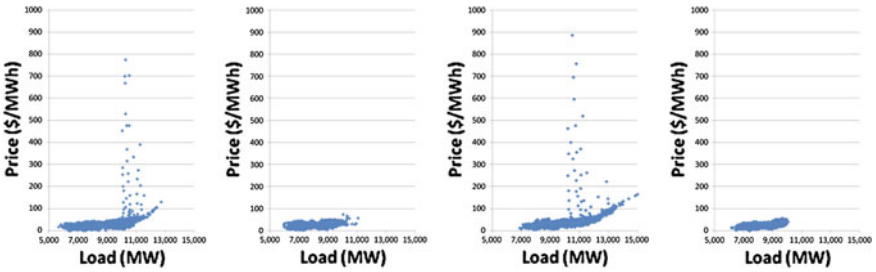

Fig. 16 NSW regional price versus load simulation for 2009–2011

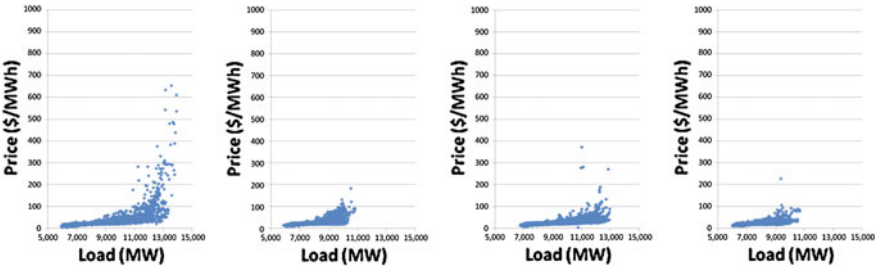

Fig. 17 NSW regional price versus load history for 2009–2011

lations have too much of a spread on the low side for the peak hours compared to the historical data which leads to loads which are too high (too much heating based load). This implies that the temperature volatility is too high for the winter peak.

The comparison of simulated price against load and historical price against load for 2009–2011 is shown in Figs. 16 and 17.

The simulations capture the seasonal structural relationship and distributions of the price and load. There are some noticeable discrepancies for the extreme values of the load in the January and July charts. This is again due to the discrepancies in the temperature model discussed above feeding through the load into the price simulations. The piecewise linear fit could also be refined to better capture the peak load price dependence, which can be seen in the January and July charts. It is also

clear from the direct comparison of the January simulations and history that the price spike volatility has a dependence on the load level. The model could be easily extended to take this effect into account although estimation would be problematic given the scarcity of the price spike data.

4 Summary and Conclusions

In this paper, we developed a general framework for the modeling of Australian electricity market risk based on the structural relationships in the market. The model framework is designed to be consistent with temperature and load mean forecasts and market forward price quotes, the dependence of load on temperature, and the dependence of price on load. The primary use of the model is for the accurate evaluation of the market risk of an electricity generation and retail company but it can also be used for the valuation of electricity market derivatives and assets. We demonstrated the application of our framework by estimating the model using historical data for the Australian NEM and then simulating the NEM market variables and comparing the simulations to the historical data. The model is able to accurately capture the structure and dynamics of the NEM.

References

Aïd, R., Campi, L., & Langrene, N. (2012). A structural risk-neutral model for pricing and hedging power derivatives. *Mathematical Finance*, *13* Published online 13 Feb.

Alaton, P., Djehiche, M., & Stillberger, D. (2002). On modelling and pricing weather derivatives. *Applied Mathematical Finance*, *9*(1), 1–20.

Barlow, M. (2002). A diffusion model for electricity prices. *Mathematical Finance*, *12*(4), 287–298.

Barndorff-Nielsen, O. E., Benth, F. E., & Veraart, A. E. D. (2010). Modelling electricity forward markets by ambit fields, CREATES Research Paper 2010–41, Aarhus University.

Benth, F. E., Benth, J. S., & Koekebakker, S. (2008). *Stochastic modeling of electricity and related markets*. Singapore: World Scientific.

Benth, F. E., & Saltyte-Benth, J. (2007). The volatility of temperature and pricing of weather derivatives. *Quantitative Finance*, *7*(5), 553–561.

Benth, F. E., Kallsen, J., & Meyer-Brandis, T. (2007). A non-Gaussian Ornstein-Uhlenbeck process for electricity spot price modeling and derivatives pricing. *Applied Mathematical Finance*, *14*, 153–169.

Burger, M., Klar, B., Müller, A., & Schindlmayr, G. (2004). A spot market model for pricing derivatives in electricity markets. *Quantitative Finance*, *4*, 109–122.

Carmona, R., & Coulon, M. (2013). A survey of commodity markets and structural models for electricity prices. In F. E. Benth, V. Kholodnyi & P. Laurence (Eds.), *Quantitative energy finance: Modeling, pricing and Hedging in energy and commodity markets*. New York: Springer.

Carmona, R., Coulon, M., & Schwarz, D. (2012). Electricity price modeling and asset valuation: a multi-fuel structural approach. *Mathematics and Financial Economics*, *7*(2), 167–202.

Cartea, A., & Figueroa, M. (2005). Pricing in electricity markets: A mean reverting jump diffusion model with seasonality. *Applied Mathematical Finance*, *12*(4), 313–335.

Cartea, A., & Villaplana, P. (2008). Spot price modeling and the valuation of electricity forward contracts: The role of demand and capacity. *Journal of Banking and Finance*, *32*, 2501–2519.

Clewlow, L., & Strickland, C. (1999). Valuing Energy Options in a One Factor Model Fitted to Forward Prices, SSRN paper 160608.

Clewlow, L., & Strickland, C. (2000). *Energy derivatives: Pricing and risk management*. London: Lacima Publications.

Coulon, M., & Howison, S. (2009). Stochastic behaviour of the electricity bid stack: from fundamental drivers to power prices. *Journal of Energy Markets*, *2*, 29–69.

Geman, H., & Roncoroni, A. (2006). Understanding the fine structure of electricity prices. *Journal of Business*, *79*, 1225–1261.

Kholodnyi, V. A. (2011). Modeling power forward prices for power spot prices with upward and downward spikes in the framework of the non-Markovian approach. *Journal of Mathematics in Engineering, Science and Aerospace*, *2*(2), 105–120.

Klüppelberg, C., Meyer-Brandis, T., & Schmidt, A. (2010). Electricity spot price modelling with a view towards extreme spike risk. *Quantitative Finance*, *10*, 963–974.

Pirrong, C., & Jermakyan, M. (2008). The price of power: The valuation of power and weather derivatives. *Journal of Banking and Finance*, *32*, 2520–2529.

Veraart, A. E. D., & Veraart, L. A. M. (2013). Modelling electricity day-ahead prices by multivariate Lévy semistationary processes. In F. E. Benth, V. Kholodnyi, & P. Laurence (Eds.), *Quantitative Energy Finance*. Vienna: Springer.

Weron, R., Bierbrauer, M., & Truck, S. (2004). Modeling electricity prices: Jump diffusion and regime switching. *Physica A: Statistical Methods and its Applications*, *336*, 39–48.

On an Integral Arising in Mathematical Finance

Mark Craddock

1 Introduction

The purpose of this paper is to present a tractable form for an integral that arises in a number of problems in analysis, financial mathematics and in other areas. For example, it is connected to Asian option pricing, the Hartman-Watson law in stochastic calculus as well as the problem of pricing zero coupon bonds in the Dothan model (See below for the details). The integral in question was first derived by Yakubovich as a fundamental solution to the parabolic PDE

$$u_t = x^2 u_{xx} + x u_x - x^2 u, \tag{1}$$

see Yakubovich (2011). The solution of this PDE can be obtained by using an index transform. This leads us to the so-called Yakubovich heat kernel (alternatively the Yakubovich integral), which is defined by

$$h_t(x, w) = \frac{2}{\pi^2} \int_0^\infty k e^{-k^2 t} \sinh(\pi k) K_{ik}(x) K_{ik}(w) \mathrm{d}k, \tag{2}$$

where the function K_ν is McDonald's modified Bessel function, Abramowitz and Stegun (1972). The Yakubovich integral does not seem to have been evaluated in closed form, and the fact that we need to integrate over the index of two Bessel functions presents difficulties.

The outline of the paper is as follows. First, we provide motivation by showing how the PDE (1) arises when we seek to price an Asian option. Then we show that

[1] The integral derived by Yakubovich is slightly different, but equivalent to our formulation.

M. Craddock (✉)
School of Mathematical Sciences, University of Technology Sydney, PO Box 123, Broadway, NSW 2007, Australia
e-mail: Mark.Craddock@uts.edu.au

R. Dieci et al. (eds.), *Nonlinear Economic Dynamics and Financial Modelling*,
DOI: 10.1007/978-3-319-07470-2_20,

h_t is actually the heat kernel for (1). Using the Fourier sine transform, we are then able to show how the integral may be reduced to a simpler form, which may be explicitly evaluated as a series of error functions. Finally, we discuss some financial applications.

2 The Asian Option Pricing PDE

The Yakubovich integral has a number of applications. To motivate our investigation, we consider Asian option pricing. Asian options are among the most popular of path-dependent options currently traded. There is a very considerable literature on the subject, and there are several alternative methods available for pricing. The first is to solve a pricing PDE. Much attention has been focused on the PDE introduced by Vecer (2001). Solving the PDE must be done numerically as no analytical solution satisfying the necessary boundary and terminal conditions is known. Second, one may attempt to numerically invert the Laplace transform of the price derived by Geman and Yor (1992), Craddock et al. (2000), Fu et al. (1995), Eydeland and Geman (1995). A fast and reasonably efficient third approach is to use the well-known approximation for the Asian price derived by moment matching. Hull (1997) describes this method. We can also use Monte Carlo simulation. The literature is extensive, but see Fu et al. (1995) again. Finally, we mention that a fast and accurate approximation using Taylor expansions was obtained by Ju (2002).

Before Vecer introduced a simpler PDE, attention was focused on the PDE found in, for example, the introductory book by Dewynne et al. (1995). This equation has two-state variables, but is second order in only one of them. The problem of solving this PDE motivates the current work and leads directly to the Yakubovich heat kernel.

Suppose then that we are interested in pricing an Asian option with arithmetic average on the underlying $S = \{S(t) : t \geq 0\}$, which follows geometric Brownian motion. That is

$$\mathrm{d}S(t) = rS(t)\mathrm{d}t + \sigma S(t)\mathrm{d}B(t), \qquad S(0) = S_0,$$

with $B = \{B(t) : t \geq 0\}$ a standard Brownian motion. Here r denotes the short rate of interest and σ is the volatility of the stock. Introduce $\xi(t) = \int_0^t S(\tau)\mathrm{d}\tau$ and let $t \to T - t$. It is well known that if the price of the option is $V(S, \xi, t) = \bar{V}(\ln S, \xi, t)$ then

$$\bar{V}_t = \frac{1}{2}\sigma^2 \bar{V}_{yy} + \left(r - \frac{1}{2}\sigma^2\right)\bar{V}_y + e^y \bar{V}_\xi - r\bar{V}, \tag{3}$$

with $\bar{V}(\ln S, \xi, 0) = f(S, \xi)$ for some payoff f. Boundary conditions and the derivation of the PDE are in Dewynne et al. (1995). We now take the Laplace transform in ξ and define

$$v(y, p, t) = \int_0^\infty \bar{V}(y, \xi, t) e^{-p\xi} \mathrm{d}\xi. \tag{4}$$

Integrating by parts gives

$$\int_0^\infty e^y \bar{V}_\xi(y, \xi, t) e^{-p\xi} \mathrm{d}\xi = p e^y v(y, p, t) - e^y \bar{V}(y, 0, t).$$

For the term $\bar{V}(y, 0, t)$ we observe that $\xi(t') = 0$ implies that $S(t') = 0$ for all $t < t'$. As an option on an underlying whose value is always zero would usually be worthless, it is reasonable to set $\bar{V}(y, 0, t) = \bar{V}(0, 0, t) = 0$. However if we do desire a non-zero value for $\bar{V}(y, 0, t)$, then this can be incorporated later using standard variation of parameter techniques. Our problem now is to solve the PDE

$$v_t = \frac{1}{2}\sigma^2 v_{yy} + \left(r - \frac{1}{2}\sigma^2\right) v_y + (pe^y - r)v, \tag{5}$$

with $v(z, p, 0) = F(e^z, p)$ where F denotes the Laplace transform of f in the second variable. Equation (5) has time-independent solutions in terms of Bessel functions. This motivates the change of variables

$$v(z, p, t) = e^{(\frac{1}{2} - r/\sigma^2)z} e^{\frac{1}{8}\sigma^2\alpha^2 t} U\left(\frac{2\sqrt{2e^z p}}{\sigma}, p, \frac{1}{8}\sigma^2 t\right), \tag{6}$$

where $\alpha = (2r/\sigma^2 + 1)$ and U satisfies

$$U_\tau = y^2 U_{yy} + y U_y + y^2 U.$$

This is very close to (1). Finally, letting $y = ix$, $u(y, p, \tau) = U(-iy, p, \tau)$, we have

$$u_\tau = x^2 u_{xx} + x u_x - x^2 u$$

subject to

$$u(x, p, 0) = \left(\frac{i\sigma x}{\sqrt{8p}}\right)^{2r/\sigma^2 - 1} F\left(-\frac{\sigma^2 x^2}{8p}, p\right).$$

Thus, the PDE (5) can be reduced to (1). A solution of this initial value problem is given by

$$u(x, p, \tau) = \int_0^\infty u(w, p, 0) h_\tau(x, w) \mathrm{d}w,$$

where $h_\tau(x, w)$ is the heat kernel. So our next task is to compute this heat kernel.

3 The Yakubovich Heat Kernel

We consider the PDE

$$u_t = x^2 u_{xx} + x u_x - x^2 u, \tag{7}$$

with $u(x, 0) = \phi(x)$ for some bounded function ϕ. Yakubovich has studied this problem in detail, see Yakubovich (2011). In Yakubovich (2012) he relates the heat kernel for this equation to the Yor integral for the density arising from the Hartman-Watson law. In fact the PDE may be solved by means of the Kontorovich-Lebedev transform, which was introduced by M.I. Kontorovich and N.N. Lebedev in 1938, see Gutiérrez-Tovar and Méndez-Pérez (2007) and Yakubovich and Luchko (1994). We use the transform pair

$$(\mathscr{K} g)(k) = \int_0^\infty \frac{1}{x} g(x) K_{ik}(x) \mathrm{d}x, \tag{8}$$

where $K_{ik}(x)$ is the modified Bessel function of the second kind, see Abramowitz and Stegun (1972). The inversion integral is

$$g(x) = \frac{2}{\pi^2} \int_0^\infty k \sinh(\pi k) (\mathscr{K} g)(k) K_{ik}(x) \mathrm{d}k. \tag{9}$$

Recall that the Bessel function $K_{ik}(x)$ satisfies the equation

$$x^2 K''_{ik}(x) + x K'_{ik}(x) - (x^2 - k^2) K_{ik}(x) = 0.$$

We suppose that the solution of the PDE (7) is $u(x, t)$ and let

$$\widehat{u}(k, t) = \int_0^\infty \frac{1}{x} u(x, t) K_{ik}(x) \mathrm{d}x.$$

Then assuming suitable behaviour at zero for u we obtain

$$\begin{aligned}
\widehat{u_t}(k,t) &= \int_0^\infty \frac{1}{x} u_t(x,t) K_{ik}(x) \mathrm{d}x \\
&= \int_0^\infty \frac{1}{x} (x^2 u_{xx}(x,t) + x u_x(x,t) - x^2 u(x,t)) K_{ik}(x) \mathrm{d}x \\
&= \int_0^\infty (x u_{xx}(x,t) + u_x(x,t) - x u(x,t))) K_{ik}(x) \mathrm{d}x \\
&= \int_0^\infty u(x,t) \left((x K_{ik}(x))'' - K'_{ik}(x) - x K_{ik}(x) \right) \mathrm{d}x \\
&= \int_0^\infty u(x,t) (x K''_{ik}(x) + K'_{ik}(x) - x K_{ik}(x)) \mathrm{d}x \\
&= \int_0^\infty \frac{1}{x} u(x,t) (x^2 K''_{ik}(x) + x K'_{ik}(x) - x^2 K_{ik}(x)) \mathrm{d}x \\
&= -k^2 \int_0^\infty \frac{1}{x} u(x,t) K_{ik}(x) \mathrm{d}x.
\end{aligned}$$

So the transformed solution $\widehat{u}$ satisfies $\widehat{u_t}(k,t) = -k^2 \widehat{u}(k,t)$. Solving this gives $\widehat{u}(k,t) = \widehat{u}(k,0) e^{-k^2 t}$, where of course $\widehat{u}(k,0) = (\mathscr{K}\phi)(k)$. Inverting $\widehat{u}(k,t)$ we have a solution to the initial value problem given by

$$u(x,t) = \frac{2}{\pi^2} \int_0^\infty (\mathscr{K}\phi)(k) k \sinh(\pi k) K_{ik}(x) \mathrm{d}k. \tag{10}$$

This can be written as

$$u(x,t) = \frac{2}{\pi^2} \int_0^\infty \int_0^\infty \frac{\phi(w)}{w} k e^{-k^2 t} \sinh(\pi k) K_{ik}(x) K_{ik}(w) \mathrm{d}k \mathrm{d}w = \int_0^\infty \frac{\phi(w)}{w} h_t(x,w) \mathrm{d}w,$$

where

$$h_t(x,w) = \frac{2}{\pi^2} \int_0^\infty k e^{-k^2 t} \sinh(\pi k) K_{ik}(x) K_{ik}(w) \mathrm{d}k \tag{11}$$

is the heat kernel for (7). We will refer to this as the Yakubovich heat kernel. We will prove the following result.

Theorem 1 *The Yakubovich heat kernel is given by*

$$h_t(x,w) = \frac{1}{4\sqrt{\pi}t^{3/2}} \int\limits_{\cosh^{-1}\left(\frac{x^2+w^2}{2xw}\right)}^{\infty} \xi e^{-\frac{\xi^2}{4t}} J_0\left(2xw\cosh\xi - x^2 - w^2\right) d\xi. \quad (12)$$

Proof The proof uses the Fourier sine transform define by

$$(\mathscr{F}_s f)(y) = g(y) = \int\limits_0^{\infty} f(k)\sin(ky)\mathrm{d}k. \quad (13)$$

The inverse sine transform is

$$(\mathcal{F}_s^{-1} g)(k) = f(k) = \frac{2}{\pi}\int\limits_0^{\infty} (\mathcal{F}_s f)(y)\sin(ky)\mathrm{d}k.$$

On page 189 of Oberhettinger (1990), we find the Fourier sine transform

$$\int\limits_0^{\infty} \sinh(\pi x) K_{ix}(a) K_{ix}(b)\sin(xy)\mathrm{d}x = \mathcal{J}(u(a,b,y))$$

$$= \begin{cases} \frac{\pi^2}{4} J_0(u(a,b,y)), & u > 0 \\ 0, & u < 0, \end{cases}$$

where $u(a,b,y) = 2ab\cosh(y) - a^2 - b^2$. Here J_0 is the zeroth order Bessel function of the first kind Abramowitz and Stegun (1972). Now we have by Fubini's Theorem

$$\begin{aligned} h_t(x,w) &= \frac{2}{\pi^2}\int\limits_0^{\infty} k e^{-k^2 t} \mathcal{F}_s^{-1}\left(\mathscr{J}(u(a,b,\xi))\right)(k)\mathrm{d}k \\ &= \frac{4}{\pi^3}\int\limits_0^{\infty}\int\limits_0^{\infty} k e^{-k^2 t}\sin(ky)\mathcal{J}(u(a,b,\xi)\mathrm{d}k\mathrm{d}\xi \\ &= \frac{2}{\pi^2}\int\limits_0^{\infty} \frac{\xi e^{-\frac{\xi^2}{4t}}}{2\sqrt{\pi}t^{3/2}}\mathcal{J}(u(a,b,\xi))\mathrm{d}\xi \end{aligned}$$

$$= \frac{1}{4\sqrt{\pi}t^{3/2}} \int_{\cosh^{-1}\left(\frac{x^2+w^2}{2xw}\right)}^{\infty} \xi e^{-\frac{\xi^2}{4t}} J_0\left(2xw\cosh\xi - x^2 - w^2\right) \mathrm{d}\xi.$$

We finally note that for all positive x and w, the quantity $\frac{x^2+w^2}{2xw} \geq 1$, so that the lower bound of the integral is always real. □

This is a form which is much easier to evaluate numerically than the original form, as we do not have to integrate with respect to the index of the Bessel function. The Integral can also be further reduced to the following result.

Corollary 1 *The Yakubovich heat kernel $h_t(x, w)$ may be written*

$$h_t(x,w) = \frac{1}{\sqrt{4\pi t}} \exp\left(-\frac{\left(\cosh^{-1}\left(\frac{x^2+w^2}{2xw}\right)\right)^2}{4t}\right) + \frac{2wx}{\sqrt{4\pi t}} \int_{\cosh^{-1}\left(\frac{x^2+w^2}{2xw}\right)}^{\infty} \sinh\xi \exp\left(-\frac{\xi^2}{4t}\right) J_1\left(x^2+w^2-2xw\cosh\xi\right) d\xi.$$

Equivalently

$$h_t(x,w) = \frac{1}{\sqrt{4\pi t}} \exp\left(-\frac{\left(\cosh^{-1}\left(\frac{x^2+w^2}{2xw}\right)\right)^2}{4t}\right) + \frac{1}{\sqrt{4\pi t}} \int_0^{\infty} \exp\left(-\frac{\left(\cosh^{-1}\left(\frac{x^2+w^2+u}{2xw}\right)\right)^2}{4t}\right) J_1(u) du.$$

Proof We simply integrate by parts and use the change of variable $u = w^2 + x^2 - 2wx\cosh\xi$, and the relation $J_0'(x) = -J_1(x)$, then replace u with $-u$. □

It is straightforward to expand the heat kernel in a series. From the definition of $J_1(x)$ (Abramowitz and Stegun (1972), p 360) one easily sees that

$$J_1(u) = \frac{\left(x^2+w^2-2xw\cosh\xi\right)}{2} \sum_{j=0}^{\infty} \frac{\left(x^2+w^2-2xw\cosh\xi\right)^{2j}}{2^{2j}\, j!(j+1)!},$$

with u as in the preceding proof. The important point is that the integrals

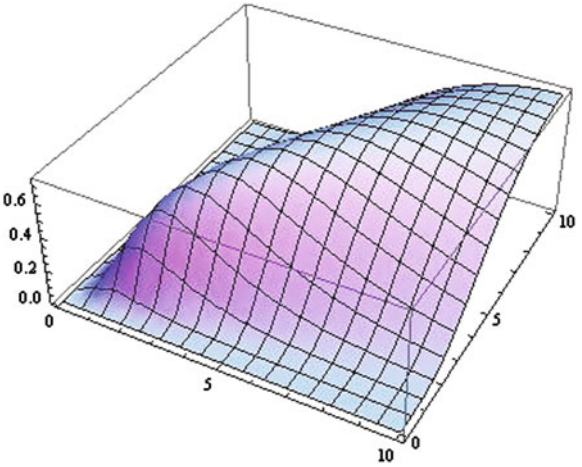

Fig. 1 The Yakubovich Heat Kernel at $t = 0.2$

$$\int\limits_{\cosh^{-1}\left(\frac{x^2+w^2}{2xw}\right)}^{\infty} \sinh\xi \exp\left(-\frac{\xi^2}{4t}\right)\cosh^l\xi \mathrm{d}\xi, \quad l = 0, 1, 2, 3, \ldots$$

can all be evaluated in terms of the error function. Consequently, we can expand the heat kernel as a series of error functions. For example,

$$\int\limits_{\cosh^{-1}\left(\frac{x^2+w^2}{2xw}\right)}^{\infty} \sinh\xi e^{-\frac{\xi^2}{4t}} \mathrm{d}\xi = \frac{1}{2}\sqrt{\pi}e^t\sqrt{t}\left(\mathrm{erf}\left(\frac{a+2t}{2\sqrt{t}}\right) - \mathrm{erf}\left(\frac{a-2t}{2\sqrt{t}}\right)\right),$$

where $a = \cosh^{-1}\left(\frac{x^2+w^2}{2xw}\right)$. So we have an expansion

$$\begin{aligned} h_t(x,w) = \frac{1}{\sqrt{4\pi t}} \exp\left(-\frac{\left(\cosh^{-1}\left(\frac{x^2+w^2}{2xw}\right)\right)^2}{4t}\right) \\ + \frac{wx}{4}(x^2+w^2)e^t\left(\mathrm{erf}\left(\frac{a+2t}{2\sqrt{t}}\right) - \mathrm{erf}\left(\frac{a-2t}{2\sqrt{t}}\right)\right) \\ - \frac{1}{4}e^{4t}w^2x^2\left(\mathrm{erfc}\left(\frac{a-4t}{2\sqrt{t}}\right) - 4\mathrm{erfc}\left(\frac{a+4t}{2\sqrt{t}}\right)\right) + \cdots . \end{aligned} \tag{14}$$

Finally, since $|J_1(u)| \leq 3/5$ we have the easy estimate

$$|h_t(x,w)| \le \frac{1}{\sqrt{4\pi t}} \exp\left(-\frac{\left(\cosh^{-1}\left(\frac{x^2+w^2}{2xw}\right)\right)^2}{4t}\right) + \frac{3}{10\sqrt{\pi t}} \int_0^\infty \exp\left(-\frac{\left(\cosh^{-1}\left(\frac{x^2+w^2+u}{2xw}\right)\right)^2}{4t}\right) \mathrm{d}u.$$

We plot the kernel for $t = 0.2$ in 1.

4 Alternative Forms for $h_t(x, w)$

We may establish other equivalent forms for the heat kernel. One such is as follows.

Proposition 1 *Let*

$$I = I(x,w,t) = \int_{-\infty}^{\infty} k e^{-k^2 t} \sinh(\pi k) K_{ik}(x) K_{ik}(w) \mathrm{d}k.$$

Then

$$I = \sqrt{\frac{\pi}{t}} e^{\frac{\pi^2}{4t}} \int_{-\infty}^{\infty} e^{\frac{-\xi^2}{4t}} \sin\left(\frac{\pi\xi}{2t}\right) \sinh\xi \frac{K_1(\sqrt{x^2+w^2+2xw\cosh\xi})}{\sqrt{x^2+w^2+2xw\cosh\xi}} \mathrm{d}\xi,$$

and $h_t(x,w) = \frac{1}{2} I$.

Proof The second part follows trivially from the fact that the integrand is even in k. Now we use the identities

$$K_{ik}(u) = \int_0^\infty e^{-u\cosh\xi + ik\xi} \mathrm{d}\xi$$

and

$$K_{ik}(x) K_{ik}(w) = \frac{1}{2} \int_{-\infty}^{\infty} \exp\left(-\frac{uw}{2x} - \frac{x(u^2+w^2)}{2uw}\right) K_{ik}(u) \frac{\mathrm{d}u}{u},$$

see Yakubovich and Luchko (1994). Using Fubini's Theorem we have

$$I(x,w,t) = \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{\infty} k e^{-k^2 t}\sinh(\pi k) \times \exp\left(-\frac{uw}{2x} - \frac{x(u^2+w^2)}{2uw}\right) e^{-u\cosh\xi + ik\xi}\frac{\mathrm{d}u}{u} dk d\xi.$$

Now from Gradshteyn and Ryzhik (2000) we have the integral (which can also be done in Mathematica 9),

$$\int_0^{\infty} e^{-\frac{uw}{2x}-\frac{x(u^2+w^2)}{2uw}} e^{-u\cosh\xi}\frac{\mathrm{d}u}{u} = 2K_0(\sqrt{x^2+w^2+2xw\cosh\xi}).$$

So that

$$\begin{aligned} I &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} k e^{-k^2 t}\sinh(\pi k) e^{ik\xi} K_0(\sqrt{x^2+w^2+2xw\cosh\xi})\mathrm{d}k\mathrm{d}\xi \\ &= -i\sqrt{\frac{\pi}{4t}}\int_{-\infty}^{\infty}\frac{\mathrm{d}}{\mathrm{d}\xi}\left[e^{\frac{(\pi+i\xi)^2}{4t}} - e^{\frac{(\pi-i\xi)^2}{4t}}\right] K_0(\sqrt{x^2+w^2+2xw\cosh\xi})\mathrm{d}\xi \\ &= \sqrt{\frac{\pi}{t}} e^{\frac{\pi^2}{4t}}\int_{-\infty}^{\infty} e^{-\frac{\xi^2}{4t}}\sin\left(\frac{\pi\xi}{2t}\right)\sinh(\xi)\frac{K_1(\sqrt{x^2+w^2+2xw\cosh\xi})}{\sqrt{x^2+w^2+2xw\cosh\xi}}\mathrm{d}\xi, \end{aligned}$$

upon integrating by parts and using Euler's formula to reduce the exponential terms. We made use of the identity $K_0'(x) = -K_1(x)$. □

Another approach to the integral is to observe that

$$I(x,w,t) = \frac{\partial}{\partial\alpha}\int_{-\infty}^{\infty} e^{-k^2 t}\cosh(\alpha k) K_{ik}(x) K_{ik}(w)\mathrm{d}k\Big|_{\alpha=\pi}. \tag{15}$$

By using the same arguments as in the previous proposition, we have the following result.

Proposition 2 *Let*

$$J(x,w,t) = \int_{-\infty}^{\infty} e^{-k^2 t}\cosh(\alpha k) K_{ik}(x) K_{ik}(w)\mathrm{d}k$$

Then

$$J(x,w,t) = 2\sqrt{\frac{\pi}{t}}e^{\frac{\alpha^2}{4t}}\int_{-\infty}^{\infty} e^{-\frac{\xi^2}{4t}}\cos\left(\frac{\alpha\xi}{2t}\right)K_0(\sqrt{x^2+w^2+2xw\cosh\xi})\mathrm{d}\xi.$$

As $K_0(x)$ decays exponentially, it should be possible to numerically evaluate these integrals quite efficiently.

5 Some Applications

For brevity, we will only sketch the applications. The forthcoming paper Craddock and Roberts (2014) will address these in considerable detail.

5.1 Asian Options

The initial motivation for this investigation was the problem of pricing an Asian option. Our results on the Yakubovich heat kernel immediately give us the solution of the reduced form of the pricing equation for an Asian option derived previously. We have

$$u(x,p,\tau) = \frac{1}{\sqrt{4\pi\tau}}\int_0^\infty u(w,p,0)\exp\left(-\frac{\left(\cosh^{-1}\left(\frac{x^2+w^2}{2xw}\right)\right)^2}{4\tau}\right)\mathrm{d}w$$
$$+\frac{1}{\sqrt{4\pi\tau}}\int_0^\infty\int_0^\infty u(w,p,0)\exp\left(-\frac{\left(\cosh^{-1}\left(\frac{x^2+w^2+u}{2xw}\right)\right)^2}{4\tau}\right)J_1(u)\mathrm{d}u\mathrm{d}w,$$

where $u(w,p,0) = \left(\frac{i\sigma w}{\sqrt{8p}}\right)^{2r/\sigma^2-1}F\left(-\frac{\sigma^2w^2}{8p},p\right)$. Back substitution will lead us to the Laplace transform in the ξ variable for the Asian option and the result depends on the payoff f that we pick. One interesting question is how this relates to the Geman-Yor formula? One can relate the Geman-Yor formula to our result, but the details are involved, so we defer them to Craddock and Roberts (2014). We note that there are already efficient methods for pricing Asian options, but this method does have some attractive features. Most notably, the double integral term can often be ignored. This requires justification, but similar comments hold for the Dothan bond pricing model below, and as this requires less prepatory material, we will present some preliminary numerical examples below for this model.

5.2 Yor's Integral

We note that there is another connection between Asian option pricing and the Yakubovich integral. In 1980, Yor expressed a density which is related to the Hartman-Watson law as an integral involving only elementary functions, Yor (1980). Specifically, the density defined by the Laplace transform

$$\int_0^\infty e^{-\lambda t} \mathrm{d}\eta_r(t) = \frac{I_{\sqrt{2\lambda}}(r)}{I_0(r)}, \tag{16}$$

has the density

$$p(t,r) = \frac{1}{\sqrt{2\pi t^3} I_0(r)} H_t(r), \tag{17}$$

where

$$H_t(r) = \frac{rt}{\pi} \int_0^\infty \exp\left(\frac{1}{2t}(\pi^2 - x^2) - r\cosh x\right) \sinh x \sin\left(\frac{\pi x}{t}\right) \mathrm{d}x. \tag{18}$$

The Hartman-Watson law is fundamental in deriving the Geman-Yor Laplace transform for the price of an Asian option, Geman and Yor (1992). There is a direct relationship between the Yakubovich heat kernel and (18). In Yakubovich (2012), Yakubovich showed that

$$h_{t/2}(x,w) = \int_0^\infty x \exp\left(-\frac{1}{2}\left(r\frac{x^2+w^2}{xw} + \frac{xw}{r}\right)\right) H_t(r)\mathrm{d}r.$$

As a consequence, we have the following easy result.

Corollary 2 *Let $r = \{r(t) : t \geq 0\}$ be the process with density $p(t,r)$ defined by (16), with $r(t) = y$. Then*

$$\begin{aligned} &E_x\left[I_0\left((r(t))\right) \exp\left(-\frac{1}{2}\left(r(t)\frac{x^2+w^2}{xw} + \frac{xw}{r(t)}\right)\right)\right] \\ &= \int_0^\infty I_0(r) \exp\left(-\frac{1}{2}\left(r\frac{x^2+w^2}{xw} + \frac{xw}{r}\right)\right) p(t,r)dr \\ &= \frac{1}{\sqrt{2\pi t^3}x}\left(\frac{1}{\sqrt{4\pi t}} \exp\left(-\frac{\left(\cosh^{-1}\left(\frac{x^2+w^2}{2xw}\right)\right)^2}{4t}\right)\right) \end{aligned}$$

$$+\frac{1}{\sqrt{4\pi t}}\int_0^\infty \exp\left(-\frac{\left(\cosh^{-1}\left(\frac{x^2+w^2+u}{2xw}\right)\right)^2}{4t}\right)J_1(u)du\Bigg).$$

Here $E_x[f(r(t))] = E[f(r_t)|r(0) = x]$.

5.3 Bond Pricing in the Dothan Model

The Dothan model, Dothan (1978), for the short rate of interest $r = \{r(t) : t \geq 0\}$ is

$$\mathrm{d}r(t) = \frac{1}{2}(1-p)\sigma^2 r(t)\mathrm{d}t + \sigma r(t)\mathrm{d}B(t),$$

where $p\sigma/2$ is the market price of risk. As usual, σ denotes volatility and $B = \{B(t) : t \geq 0\}$ is a standard Brownian motion. In Pintoux and Privault (2011), Pintoux and Privault prove that in the Dothan model, the price of a zero coupon bond $P(t, T)$ is given by $P(t, T) = F(T - t, r_t)$, where

$$F(\tau, r) = \frac{2p(2r)^{p/2}}{\sigma^p}\exp\left(-\frac{\sigma^2 p^2\tau}{8}\right)\int_0^\infty e^{-py}h_{\sigma^2\tau/8}\left(\frac{\sqrt{8r}}{\sigma}, e^y\right)\mathrm{d}y. \tag{19}$$

Using our expression for the Yakubovich integral, we can express the bond price as

$$F(\tau, r) = \alpha_{p,r}\Bigg[\int_0^\infty \frac{e^{-py}}{\sqrt{\pi\sigma^2\tau/2}}\exp\left(-\frac{\cosh^{-1}\left(\frac{\sigma(e^{2y}+8r/\sigma^2)}{2e^y\sqrt{8r}}\right)^2}{\sigma^2\tau/2}\right)\mathrm{d}y$$
$$+\int_0^\infty\int_0^\infty \frac{e^{-py}}{\sqrt{\pi\sigma^2\tau/2}}\exp\left(-\frac{\cosh^{-1}\left(\frac{\sigma(e^{2y}+8r/\sigma^2+u)}{2e^y\sqrt{8r}}\right)^2}{\sigma^2\tau/2}\right)J_1(u)\mathrm{d}u\mathrm{d}y\Bigg], \tag{20}$$

where $\alpha_{p,r} = \dfrac{2p(2r)^{p/2}}{\sigma^p}e^{-\frac{1}{8}\sigma^2 p^2\tau}$.

5.4 Numerical Implementation

We will not attempt to establish the most efficient method of approximation, numerical evaluation of the various quantities derived above. Nor will we derive error estimates here. In Craddock and Roberts (2014), we will investigate these issues in

detail. However, we will make a few preliminary observations about implementation and the consequences for pricing.

On the interval $[0, 2n]$ the Bessel function $J_k(u)$ is approximated by

$$J_k(x) \approx \sum_{m=k}^{n} d_{knm} x^{2m-k}, \tag{21}$$

where

$$d_{knm} = \frac{(-1)^{m+k} 2^{k-2m} n^{1-2m} (n+m-1)!}{(m-k)!(n-m)!(m!)}, \tag{22}$$

see Millane and Eads (2003). Thus,

$$\begin{aligned} h_t(x,w) &= \frac{1}{4\sqrt{\pi} t^{3/2}} \int_{\cosh^{-1}\left(\frac{x^2+w^2}{2xw}\right)}^{\infty} \xi e^{-\frac{\xi^2}{4t}} J_0\left(2xw\cosh\xi - x^2 - w^2\right) \mathrm{d}\xi \\ &\approx \frac{1}{4\sqrt{\pi} t^{3/2}} \sum_{m=0}^{2n} d_{0nm} \int_{\cosh^{-1}\left(\frac{x^2+w^2}{2xw}\right)}^{2n} \xi e^{-\frac{\xi^2}{4t}} \left(2xw\cosh\xi - x^2 - w^2\right)^{2m} \mathrm{d}\xi. \end{aligned} \tag{23}$$

The order of the polynomial approximation required will depend on the lower limit of integration. Each term in (23) can be evaluated in terms of exponential and error functions. Using this we may readily compute the value of the heat kernel and establish the prices in the models we are interested in. In practice, we often need few terms of the series and in fact it is frequently the case that the first term will suffice, particularly for $t < 1$. The in-built numerical routines in software such as Mathematica and Matlab can also be used to evaluate the integral with relative ease and high accuracy.

Actually, for the Dothan model, a good approximation can often be obtained simply by using the first term in (20), since the double integral is often small enough that we may neglect it. For example, consider the case when $r = 0.02$, $\sigma = 0.4$ and $p = 2$, with time to maturity of 2 years. Evaluation of (19) using (20) gives a bond price of 0.291. The contribution of the second integral is 0.006. In fact for bonds with maturities less than 2 years, one can show numerically that the double integral makes essentially no contribution to the price. For example, with the same parameters as before and time to maturity of half a year, the bond price is 0.389 and the double integral only contributes 0.0004. For longer maturities, we have the same effect. A 10-year bond with the same parameters has a price of 0.182, with the double integral contributing 0.0002 to the value.

A detailed investigation will be given in Craddock and Roberts (2014), but in practice

$$F(\tau, r) \approx \alpha_{p,r} \int_0^\infty \frac{e^{-py}}{\sqrt{\pi\sigma^2\tau/2}} \exp\left(-\frac{\cosh^{-1}\left(\frac{\sigma(e^{2y}+8r/\sigma^2)}{2e^y\sqrt{8r}}\right)^2}{\sigma^2\tau/2} \right) \mathrm{d}y$$

will usually give an excellent approximation to the bond price. Similar comments hold for Asian options, but the discussion is rather more involved and so we defer it to Craddock and Roberts (2014).

References

Abramowitz, M., & Stegun, I. (1972). *Handbook of mathematical functions, with formulas, graphs and mathematical tables* (10th ed.). New York: Dover.

Craddock, M., Heath, D., & Platen, E. (2000). Numerical inversion of laplace transforms: A survey of techniques with applications to derivative pricing. *Journal of Computational Finance*, *4*, 57–81.

Craddock, M., & Roberts, D. (2014). Yakubovichs integral and the pricing of derivative securities (In Preparation).

Dewynne, J., Howison, S., & Wilmott, P. (1995). *The mathematics of financial derivatives. A student introduction*. Cambridge: Cambridge University Press.

Dothan, L. (1978). On the term structure of interest rates. *Journal of Financial Econometrics*, *6*, 59–69.

Eydeland, A., & Geman, H. (1995). Asian options revisited: Inverting the Laplace transform. *Risk*, *8*, 65–67.

Fu, M., Madan, D., & Wang, T. (1995). Pricing Asian options: A comparison of analytical and Monte Carlo methods. College Park:University of Maryland, College of Business and Management (Working Paper).

Geman, H., & Yor, M. (1992). Bessel processe. Asian options and perpetuities. *Mathematical Finance*, *3*, 471–474.

Gradshteyn, I., & Ryzhik, I. (2002) Table of integrals, series and products (6th edn). New York: Academic Press.

Gutiérrez-Tovar, Y., & Méndez-Pérez, J. (2007). The Kontorovich-Lebedev integral transformation-with a Hankel function kernel in a space of generalized functions of doubly exponential descent. *Journal of Mathematical Analysis and Applications*, *328*, 359–369.

Hull, J.C. (1997). Options, futures and other derivatives (3rd edn). New Jersey: Prentice Hall International Inc.

Ju, N. (2002). Pricing Asian and basket options via Taylor expansion. *Jouranl of Computational Finance*, *5*, 79–103.

Millane, R. P., & Eads, J. L. (2003). Polynomial approximations to Bessel functions. *IEEE Transactions on Antennas and Propagation*, *51*, 1398–1400.

Oberhettinger, F. (1990). *Tables of Fourier transforms and Fourier transforms of distributions*. Berlin: Springer.

Pintoux, C., & Privault, N. (2011). The Dothan pricing model revisited. *Mathematical Finance*, *21*, 355–363.

Vecer, J. (2001). A new PDE approach for pricing arithmetic average Asian options. *Journal of Computation Finance*, *4*, 105–113.

Yakubovich, S. (2011). The heat kernel and Heisenberg inequalities related to the Kontorovich-Lebedev transform. *Communications on Pure and Applied Analysis*, *10*, 745–760.

Yakubovich, S. (2012). On the Yor integral and a system of polynomials related to the Kontorovich-Lebedev transform. arXiv:1210.7309v1.

Yakubovich, S., & Luchko, Y. (1994). The hypergeometric approach to integral transforms and convolutions. Mathematics and its applications 287. Dordrecht: Kluwer.
Yor, M. (1980). Loi de lindice du lacet Brownien et distribution de Hartman-Watson. *Z. Wahrscheinlichkeitsth*, *53*, 71–95.

Change of Numéraire and a Jump-Diffusion Option Pricing Formula

Gerald H. L. Cheang and Gim-Aik Teh

1 Introduction

Black and Scholes (1972) give an option pricing formula in terms of the stock price and the price of a deterministic zero-coupon bond in their seminal work. Merton (1973) provides an alternative formulation of the Black-Scholes formula in terms of the forward price of the stock. Merton (1976) also gives an option pricing formula where the underlying stock price process contains a compound Poisson component, in addition to a continuous log-normally distributed component. Both Black and Scholes (1972) and Merton (1976) assume a constant risk-free rate in their models. However, under a stochastic interest rate term structure regime, the price of a zero-coupon bond is not deterministic. Geman et al. (1995), through an appropriate change in numéraire technique, give an option pricing formula where the bond price is assumed to take on HJM-type (Heath et al. 1992) dynamics, that has a single Wiener noise factor in common with the stock price dynamics. Their result is an extension of the Black-Scholes-Merton option pricing formula for the case where both the stock price and bond price share a common source of Wiener noise.

However, it has been observed that the behaviour of both stock prices and interest rates (thus bond prices) exhibit discontinuous behaviour. For example, Ball and Torus (1985) have documented empirical evidence of jumps in common stock prices, and Das (2002) and Dungey et al. (2007), in interest rates and bond prices. Björk et al. (1997a, b) extended the HJM framework to the case where the interest rates are driven by a general marked point process as well as by a Wiener process. In this

G. H. L. Cheang (✉)
School of Information Technology and Mathematical Sciences,
Centre for Industrial and Applied Mathematics,
University of South Australia, Adelaide, SA, Australia
e-mail: Gerald.Cheang@unisa.edu.au

G.-A. Teh
BNP Paribas, Hong Kong SAR, China
e-mail: gimaik@gmail.com

R. Dieci et al. (eds.), *Nonlinear Economic Dynamics and Financial Modelling*,
DOI: 10.1007/978-3-319-07470-2_21,

paper, we adopt the Geman et al. (1995) change of numéraire technique, and provide a European call option pricing formula where the stock price dynamics follow the Merton (1976) jump-diffusion model, under a stochastic interest rate term structure that is an HJM type model with jumps. The stock price and the bond price processes have different sources of Wiener noise but share a common source of jump noise in our model. Duffie et al. (2000) have priced options under affine jump-diffusion models with stochastic interest rates. However, their general results require the use of Fourier transform techniques. In this paper, the derivation of our pricing formula for a more specific model does not require Fourier transforms.

This paper is organized as follows. We introduce our extended model in Sect. 2. In Sect. 3, we introduce a Radon-Nikodým derivative process that facilitates the transformation of measures when we change the numérarie in our analysis. The forward measure and the reciprocal forward measures are defined in Sects. 4 and 5 respectively. Finally before we conclude, we present our option pricing formula in Sect. 6.

2 The Extended Model of Black-Scholes-Merton

Let $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}, \mathbb{Q})$ be a probability measure space. We are only interested in the filtration $\{\mathscr{F}_t\}_{t\leq T}$ over $0 \leq t \leq T$ for some fixed T, the expiry time of the option. We assume that we are already working in an equivalent risk-neutral measure $\mathbb{Q}$ that corresponds to the money market account $M_t = \exp(\int_0^t r_u \mathrm{d}u)$ as the numéraire, where r_t is the risk-free rate (possibly stochastic, and adapted to the filtration).

The market which contains the stock and bond with jump components, is inherently incomplete in the Harrison and Pliska (1981) sense. There will be many equivalent martingale measures that correspond to the use of the money market account as the numéraire. It is not the aim of this paper to discuss the selection of a particular suitable martingale measure from the original market measure $\mathbb{P}$. In order to select a suitable risk-neutral measure from the original market measure, one could employ the minimal entropy martingale measure found in Miyahara (2001). Alternatively, one could seek to minimize the divergence $\mathbb{E}_{\mathbb{P}}\left[\left(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right)^q\right]$ in the manner of Jeanblanc et al. (2007). Yet another approach would be to calibrating the model to market data, with the aim of minimizing the relative entropy of the calibrated risk-neutral measure relative to the original measure, as done in Cont and Tankov (2004b). Hence we assume that an equivalent risk-neutral measure $\mathbb{Q}$ is already specified and we define the stock price and bond price dynamics with respect to this measure $\mathbb{Q}$.

We assume that both the stock price and bond price dynamics are driven by different Wiener components. In our model, $W_{1,t}$ and $W_{2,t}$ are correlated standard Brownian motion under $\mathbb{Q}$ with correlation ρ, that is, $\mathrm{d}W_{1,t}.\mathrm{d}W_{2,t} = \rho \mathrm{d}t$, adapted to the filtration $\{\mathscr{F}_t\}$. Also adapted to the filtration is a homogeneous Poisson counting measure $N(\mathrm{d}y, \mathrm{d}t) \equiv q(\mathrm{d}y, \mathrm{d}t) \equiv q(\mathrm{d}y_1, \mathrm{d}y_2, \mathrm{d}t)$ defined over $\mathbb{R}^2 \times [0, T]$ which is associated with a marked point process $(\{Y_{T_n}\}, N_t)$. The intensity of the Poisson measure

$N(\mathrm{d}y, \mathrm{d}t)$ is $\lambda m_{\mathbb{Q},t}(\mathrm{d}y)\mathrm{d}t$, where $\lambda > 0$ is the constant arrival rate of the jumps of the Poisson process N_t under $\mathbb{Q}$, and $m_{\mathbb{Q},t}(\mathrm{d}y)$ is the joint probability density of the marks Y_{T_n} under $\mathbb{Q}$ when $T_n = t$. The marks Y_{T_n} at different times are independent of each other, and they are also independent of N_t and $W_t = (W_{1,t}, W_{2,t})^\top$. The independency assumption between the marked point process and the Wiener processes is partly for mathematical convenience, in particular when we apply appropriate forms of the Radon-Nikodým derivative in Theorem 1 to transform the risk-neutral measure that we are working in to the forward measure in Sect. 4 and to the reciprocal forward measure in Sect. 5. This is also consistent with Merton's (1976) financial economic arguments that the systematic risk in the model is suitably modelled with the Wiener processes, and the unsystematic risk with an independent marked point process.

We also assume that $m_{\mathbb{Q},t}(\mathrm{d}y)$ is non-atomic. For short, we will denote its compensated measure under $\mathbb{Q}$ as

$$\hat{q}(\mathrm{d}y, \mathrm{d}t) = N(\mathrm{d}y, \mathrm{d}t) - \lambda m_{\mathbb{Q},t}(\mathrm{d}y)\mathrm{d}t. \tag{1}$$

Since we have two assets in our model, the jump-size $Y_t = (Y_{1,t}, Y_{2,t})^\top$ is a bivariate process taking values $y_t = (y_{1,t}, y_{2,t})^\top$ in $\mathbb{R}^2$.

Throughout this paper, we assume that S_t is a traded stock with continuously paying deterministic dividend yield ζ_t, with return dynamics under the chosen risk-neutral measure $\mathbb{Q}$ given as

$$\frac{\mathrm{d}S_t}{S_{t-}} = (r_t - \zeta_t)\mathrm{d}t + \sigma_{1,t}\mathrm{d}W_{1,t} + \int_{\mathbb{R}^2} [e^{y_{1,t}} - 1]\hat{q}(\mathrm{d}y, \mathrm{d}t) \tag{2}$$

where r_t is the risk-free rate and $\sigma_{1,t}$ is the (deterministic) instantaneous stock volatility per unit time. It must be noted that although the integral $\int_{\mathbb{R}^2}[e^{y_{1,t}} - 1]\hat{q}(\mathrm{d}y, \mathrm{d}t)$ in (2) is a double integral with respect to $\mathrm{d}y$, we are effectively integrating the marginal density $m_{\mathbb{Q},t}(\mathrm{d}y_2)$ out.

There is also a traded zero-coupon bond with return dynamics under the risk-neutral measure $\mathbb{Q}$ given as

$$\frac{\mathrm{d}P_{t,T}}{P_{t-,T}} = r_t\mathrm{d}t + \sigma_{2,t}\mathrm{d}W_{2,t} + \int_{\mathbb{R}^2} [e^{y_{2,t}} - 1]\hat{q}(\mathrm{d}y, \mathrm{d}t), \tag{3}$$

where $\sigma_{2,t}$ is the (deterministic) instantaneous bond volatility per unit time, and the value of the bond at maturity is $P_{T,T} = 1$. Analogous to (2), the integral $\int_{\mathbb{R}^2}[e^{y_{2,t}} - 1]\hat{q}(\mathrm{d}y, \mathrm{d}t)$ in (3) is a double integral with respect to $\mathrm{d}y$ and we are effectively integrating the marginal density $m_{\mathbb{Q},t}(\mathrm{d}y_1)$ out. We note that (3) models the bond price dynamics obtained by introduction of jump components into the HJM model (Heath et al. 1992), in the manner of Björk et al. (1997a, b). If the jump-sizes are fixed instead, then we get the special case of the HJM model with fixed jumps as in Chiarella and Nikitipoulos-Sklibosios (2003).

The $Y_{i,t}$ $(i = 1, 2)$ are random jump-sizes assumed to be correlated pairwise with correlation $\text{Corr}[Y_{1,t}, Y_{2,t}] = \rho_{Y_t}$, and we define the expected proportional jump-size

$$\kappa_{i,t} \equiv \mathbb{E}_{\mathbb{Q}}\left[e^{Y_{i,t}} - 1\right] = \int_{\mathbb{R}^2} \left[e^{y_{i,t}} - 1\right] m_{\mathbb{Q},t}(\text{dy}). \tag{4}$$

The moment generating function of all the jump-sizes $Y_t = (Y_{1,t}, Y_{2,t})^\top$ is given as

$$M_{\mathbb{Q},Y_t}(u) = \mathbb{E}_{\mathbb{Q}}\left[e^{u^\top Y_t}\right]. \tag{5}$$

We state the first assumption.

Assumption 1 The moment generating function (5) of the jump-sizes Y_t exists and is finite for all values $u \in \mathbb{R}^2$.

For each asset, the nth jumps Y_{1,T_n} and Y_{2,T_n} occur together, driven by the same Poisson arrival process N_t. The correlation between the nth return jump-sizes of each asset is determined by the moment generating function of the joint jump-size distribution (5). The jumps can be related to macroeconomic shocks to the system. Should the outcome of one jump-size be zero, and the other non-zero, for example, if $Y_{1,T_n} = 0$ and $Y_{2,T_n} \neq 0$, then it could be attributed to idiosyncratic shocks.

In addition to the assumption on existence of the moment generating function of the jump-sizes Y_{T_n}, we also require some technical assumptions on the coefficients of the diffusion components in the dynamics of the stock and bond prices.

Assumption 2 The interest rate $(r_t \geq 0)$ satisfies

$$0 < \mathbb{E}_{\mathbb{Q}} \exp\left(\int_0^T r_u \text{d}u\right) < \infty \text{ a.s.} \tag{6}$$

and the deterministic volatilities satisfy

$$\int_0^T \sigma_{1,u}^2 \text{d}u < \infty \text{ and } \int_0^T \sigma_{2,u}^2 \text{d}u < \infty. \text{ a.s.} \tag{7}$$

where $\lim_{t \to T} \sigma_{2,t} = 0$.

Assumption 3 For $0 \leq t \leq T$, the mean and the variance of the relative jump-sizes of bond price satisfy

$$\lim_{t \to T} \mathbb{E}_{\mathbb{Q}}[Y_{2,t}] = 0 \text{ and } \lim_{t \to T} \text{Var}[Y_{2,t}] = 0. \tag{8}$$

Denote the doléans-dade exponential martingale under $\mathbb{Q}$ by $\mathscr{E}_{\mathbb{Q}}[\cdot]$. Then the solutions to the stochastic differential equations for the stock (2) and the bond (3) are

$$S_t = S_0 \exp\left[\int_0^t (r_u - \zeta_u)\,\mathrm{d}u\right] \mathscr{E}_{\mathbb{Q}}\left[\int_0^t \sigma_{1,u}\mathrm{d}W_{1,u} + \sum_{n=1}^{N_t} Y_{1,T_n}\right], \tag{9}$$

and

$$P_{t,T} = P_{0,T} \exp\left[\int_0^t r_u \mathrm{d}u\right] \mathscr{E}_{\mathbb{Q}}\left[\int_0^t \sigma_{2,u}\mathrm{d}W_{2,u} + \sum_{n=1}^{N_t} Y_{2,T_n}\right]. \tag{10}$$

We reiterate that the only correlations in the model are those between the Wiener components $W_{1,t}$ and $W_{2,t}$, and those between the jump-size components Y_{1,T_n} and Y_{2,T_n}. The Wiener processes and the jump-sizes are independent of each other. We also assume that the correlations ρ and $\rho_{Y_{T_n}}$ are neither ± 1. For now, we make no assumptions about the distribution of Y_t. We only require its joint moment generating function (5) to exist and that its joint density $m_{\mathbb{Q},t}(\mathrm{d}y)$ is non-atomic.

The model (2) when the interest rate is deterministic is essentially the jump-diffusion model of Merton (1976). We have extended Merton's (1976) jump-diffusion framework by introducing stochastic interest rate with jumps, resulting in the bond price dynamics given by (3). We have also extended the option pricing example given in Geman et al. (1995) by introducing another correlated Wiener process component as well as jumps into the stock and bond prices. In our model given by (2) and (3), randomness in the model is driven by the Wiener processes

$$W_u = \begin{pmatrix} W_{1,u} \\ W_{2,u} \end{pmatrix} \tag{11}$$

with correlation matrix

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \tag{12}$$

and the bivariate compound Poisson process

$$\sum_{n=1}^{N_t} Y_{T_n} = \begin{pmatrix} \sum_{n=1}^{N_t} Y_{1,T_n} \\ \sum_{n=1}^{N_t} Y_{2,T_n} \end{pmatrix} \tag{13}$$

where Y_{T_n} is distributed in the measure $\mathbb{Q}$ as determined by the moment generating function (5) with joint density $m_{\mathbb{Q},t}(\mathrm{d}y)$ when $T_n = t$. The bivariate compound process (13) has independently but non-identically distributed jumps.

3 Transformation of Measures

In option pricing problems, the main goal is a suitable risk-neutral evaluation of the final payoff conditional on information about the underlying asset prices up to the current time. However, as shown in Geman et al. (1995) for the pure-diffusion case, it is often convenient to transform the risk-neutral measure to another measure associated with a more appropriate numéraire, in order to facilitate the derivation of the option pricing formula. A suitable form for the Radon-Nikodým derivative that induces a measure transformation from the chosen risk-neutral measure $\mathbb{Q}$ to some equivalent measure $\widetilde{\mathbb{Q}}$ corresponding to some other numéraire asset, has to be a function of W_t as well as the compound Poisson process (13).

Following Runggaldier (2003), let

$$L_t = \exp\left[-\frac{1}{2}\int_0^t \theta_u^\top \Sigma^{-1}\theta_u \mathrm{d}u - \int_0^t (\Sigma^{-1}\theta_u)^\top \mathrm{d}W_u\right] \times \exp\left[-\lambda\int_0^t \kappa_u' \mathrm{d}u + \sum_{n=1}^{N_t}\left(\gamma^\top Y_{T_n} + \nu\right)\right], \tag{14}$$

where

$$\theta_u = \begin{pmatrix}\theta_{1,u}\\ \theta_{2,u}\end{pmatrix} \tag{15}$$

and

$$\kappa_u' = e^\nu M_{\mathbb{Q},Y_u}(\gamma) - 1 = e^\nu \mathbb{E}_{\mathbb{Q}}\left[e^{\gamma^\top Y_u}\right] - 1. \tag{16}$$

The parameters, if stochastic, must be adapted to the filtration. The process L_t is a Radon-Nikodým derivative parameterized by θ_u, γ and ν if $\mathbb{E}_{\mathbb{Q}}[L_t] = 1$ for $0 \le t \le T$. The following theorem which we use is standard. More general representations expressed in terms of Lévy measures, jump measures or compensated jump measures can be found in Colwell and Elliott (1993), Cont and Tankov (2004a) and Runggaldier (2003).

Theorem 1 *Consider the probability measure space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{Q})$ over the time interval $[0, T]$ such that the Wiener process W_t and the compound Poisson process $\sum_{n=0}^{N_t} Y_{T_n}$ are adapted to the filtration $\mathcal{F}_t$. Suppose L_t is given by (14) with non-stochastic parameters θ_u, γ and ν, and that $\mathbb{E}_{\mathbb{Q}}[L_T] = 1$. Then L_t is a Radon-Nikodým derivative of some equivalent measure $\widetilde{\mathbb{Q}}$ with respect to $\mathbb{Q}$, that is*

$$L_t = \left.\frac{\mathrm{d}\widetilde{\mathbb{Q}}}{\mathrm{d}\mathbb{Q}}\right|_t = \mathscr{E}_{\mathbb{Q}}\left[-\int_0^t (\Sigma^{-1}\theta_u)^\top \mathrm{d}W_u\right] \times \mathscr{E}_{\mathbb{Q}}\left[\sum_{n=1}^{N_t}\left(\gamma^\top Y_{T_n} + \nu\right)\right]. \tag{17}$$

Then the Wiener processes $W_{i,t}$ have drift $\theta_{i,t}$, for $i = 1, 2$, under the measure $\widetilde{\mathbb{Q}}$, and the compound Poisson process $\sum_{n=0}^{N_t} Y_{T_n}$ has a new arrival intensity rate

$$\tilde{\lambda}_t = \lambda(1 + \kappa_t') \tag{18}$$

with a new distribution for the jump-sizes determined by the moment generating function

$$M_{\widetilde{\mathbb{Q}},Y_t}(u) = \frac{M_{\mathbb{Q},Y_t}(u+\gamma)}{M_{\mathbb{Q},Y_t}(\gamma)} \tag{19}$$

over the interval $[t, t + \mathrm{d}t)$.

Proof Because of the independency assumption of the Wiener and the jump-components in the model and in the Radon-Nikodým derivative (17), we can obtain the new distributions of the Wiener processes and the jump-components separately using Wiener part and the jumps part of the Radon-Nikodým derivative respectively.

The Wiener part of the Radon-Nikodým derivative, given by the first term on the right-hand side of (17), is the usual Radon-Nikodým derivative for Wiener processes and via the usual Girsanov's theorem, there exists $\widetilde{W}_{i,t}$, a standard Brownian motion process under the transformed measure $\widetilde{\mathbb{Q}}$ such that

$$\mathrm{d}\widetilde{W}_{i,t} = \theta_{i,t}\mathrm{d}t + \mathrm{d}W_{i,t},\ i = 1, 2. \tag{20}$$

In order to determine the new distribution of the jump part, we derive its moment generating function under the new measure $\widetilde{\mathbb{Q}}$. It is given as

$$\begin{aligned}
\mathbb{E}_{\widetilde{\mathbb{Q}}}\left[e^{u^\top \sum_{n=1}^{N_t} Y_{T_n}}\right] &= \mathbb{E}_{\mathbb{Q}}\left[\mathscr{E}_{\mathbb{Q}}\left[\sum_{n=1}^{N_t}\left(\gamma^\top Y_{T_n} + \nu\right)\right] e^{u^\top \sum_{n=1}^{N_t} Y_{T_n}}\right] \\
&= \exp\left(\lambda \int_0^t \left(e^{\nu} M_{\mathbb{Q},Y_s}(u+\gamma) - 1\right) ds - \lambda \int_0^t \kappa_s' ds\right) \\
&= \exp\left(\lambda \int_0^t \left(e^{\nu} M_{\mathbb{Q},Y_s}(u+\gamma) - 1\right) ds - \lambda \int_0^t \left(e^{\nu} M_{\mathbb{Q},Y_s}(\gamma) - 1\right) ds\right)
\end{aligned}$$

$$= \exp\left(\int_0^t \lambda(1+\kappa'_s)\left[\frac{M_{\mathbb{Q},Y_s}(u+\gamma)}{M_{\mathbb{Q},Y_s}(\gamma)} - 1\right] ds\right). \tag{21}$$

Thus under the transformed measure $\widetilde{\mathbb{Q}}$, the compound Poisson process $\sum_{n=0}^{N_t} Y_{T_n}$ has a new arrival intensity rate given by (18) and the distribution of the jump-sizes are determined by the moment generating function (19) over the interval $[t, t+\mathrm{d}t)$. □

Remark 1 The jump part of the Radon-Nikodým derivative (17) can be written in a more general form involving a compensated Poisson measure (see for example Runggaldier (2003)) if the arrival intensity of the Poisson process N_t modelling the jump-arrivals under $\mathbb{Q}$ is non-homogeneous. In our model (2) and (3), even though the initial intensity of the jump-arrivals is homogeneous, we see that it will be non-homogeneous after the appropriate measure transformations.

4 The Forward Measure

In financial economics, the stock yield, expressed in terms of units of the bond, is known as the forward price. Thus the forward price is given as

$$\begin{aligned} F_{t,T} &= \frac{S_t e^{\int_0^t \zeta_u \mathrm{d}u}}{P_{t,T}} \\ &= F_{0,T} \exp\left[\int_0^t \left(-\frac{\sigma_{1,u}^2}{2} + \frac{\sigma_{2,u}^2}{2}\right)\mathrm{d}u + \int_0^t \sigma_{1,u}\mathrm{d}W_{1,u} - \int_0^t \sigma_{2,u}\mathrm{d}W_{2,u}\right] \\ &\quad \times \exp\left[-\lambda \int_0^t (\kappa_{1,u} - \kappa_{2,u})\mathrm{d}u + \sum_{n=1}^{N_t}(Y_{1,T_n} - Y_{2,T_n})\right]. \end{aligned} \tag{22}$$

In the same manner, the value of the money market account can be expressed in terms of units of the bond

$$\frac{M_t}{P_{t,T}} = \frac{1}{P_{0,T}} \exp\left[\int_0^t \left(\frac{\sigma_{2,u}^2}{2} + \lambda\kappa_{2,u}\right)\mathrm{d}u - \int_0^t \sigma_{2,u}\mathrm{d}W_{2,u} - \sum_{n=1}^{N_t} Y_{2,T_n}\right]. \tag{23}$$

Under the forward measure $\mathbb{Q}_P$, the bond $P_{t,T}$ is numéraire, and the forward price of the stock yield $\left\{\frac{S_t e^{\int_0^t \zeta_u \mathrm{d}u}}{P_{t,T}}\right\}$, and the money market account in terms of units of the bond $\left\{\frac{M_t}{P_{t,T}}\right\}$ must be martingales.

Set $\ell_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \ell_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $\ell = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. By the application of Itô's Lemma for jump-diffusion processes (see Protter (1990) or Applebaum (2004)), the dynamics for the forward price are

$$\begin{aligned} \frac{\mathrm{d}F_{t,T}}{F_{t-,T}} &= (\sigma_{2,t}^2 - \rho\sigma_{1,t}\sigma_{2,t})\mathrm{d}t + \sigma_{1,t}\mathrm{d}W_{1,t} - \sigma_{2,t}\mathrm{d}W_{2,t} \\ &\quad - \lambda(\kappa_{1,t} - \kappa_{2,t})\mathrm{d}t + \int_{\mathbb{R}^2} \left[e^{\ell^\top y_t} - 1\right] N(\mathrm{d}y, \mathrm{d}t). \end{aligned} \tag{24}$$

Similarly, the dynamics for the discounted money market account are

$$\mathrm{d}\left(\frac{M_t}{P_{t,T}}\right) = \frac{M_t}{P_{t-,T}} \left[\sigma_{2,t}^2 \mathrm{d}t - \sigma_{2,t}\mathrm{d}W_{2,t} + \lambda\kappa_{2,t}\mathrm{d}t + \int_{\mathbb{R}^2} \left[e^{-\ell_2^\top y_t} - 1\right] N(\mathrm{d}y, \mathrm{d}t) \right]. \tag{25}$$

From (24) and (25), we see that the choice of $\theta_{P,t} = (\,-\rho\sigma_{2,t} \;\; -\sigma_{2,t}\,)^\top$, $\gamma = \ell_2$ and $\nu = 0$ in the (17) in Theorem 1 yields the Radon-Nikodým derivative

$$\begin{aligned} \left.\frac{\mathrm{d}\mathbb{Q}_P}{\mathrm{d}\mathbb{Q}}\right|_t &= \mathscr{E}_{\mathbb{Q}}\left[-\int_0^t (\Sigma^{-1}\theta_{P,u})^\top \mathrm{d}W_u \right] \times \mathscr{E}_{\mathbb{Q}}\left[\sum_{n=1}^{N_t} \ell_2^\top Y_{T_n} \right] \\ &= \mathscr{E}_{\mathbb{Q}}\left[\int_0^t \sigma_{2,u}\mathrm{d}W_{2,u} + \sum_{n=1}^{N_t} Y_{2,T_n} \right]. \end{aligned} \tag{26}$$

Thus from Girsanov's Theorem for Wiener processes, there exist standard Brownian motion process $\widetilde{W}_{1,t}$ and $\widetilde{W}_{2,t}$ under the forward measure $\mathbb{Q}_P$ such that

$$\mathrm{d}\widetilde{W}_{1,t} = -\rho\sigma_{2,t}\mathrm{d}t + \mathrm{d}W_{1,t}, \tag{27}$$

and

$$\mathrm{d}\widetilde{W}_{2,t} = -\sigma_{2,t}\mathrm{d}t + \mathrm{d}W_{2,t}. \tag{28}$$

Furthermore from Theorem 1, it follows that the Poisson process N_t has a new non-homogeneous arrival intensity

$$\tilde{\lambda}_t = \lambda\mathbb{E}_{\mathbb{Q}}[e^{\ell_2^\top Y_t}] = \lambda\mathbb{E}_{\mathbb{Q}}[e^{Y_{2,t}}] \tag{29}$$

and the moment generating function of the jump-size distributions under the forward measure $\mathbb{Q}_P$ is

$$M_{\mathbb{Q}_P,Y_t}(u) = \frac{M_{\mathbb{Q},Y_t}(u+\ell_2)}{M_{\mathbb{Q},Y_t}(\ell_2)}. \tag{30}$$

We also define the compensated Poisson counting measure under the forward measure $\mathbb{Q}_P$ as

$$\hat{q}_P(\mathrm{d}y,\mathrm{d}t) = N(\mathrm{d}y,\mathrm{d}t) - \tilde{\lambda}_t m_{\mathbb{Q}_P,t}(\mathrm{d}y)\mathrm{d}t. \tag{31}$$

Thus the dynamics for the forward price as a martingale in the forward measure $\mathbb{Q}_P$ are

$$\frac{\mathrm{d}F_{t,T}}{F_{t-,T}} = \sigma_{1,t}\mathrm{d}\widetilde{W}_{1,t} - \sigma_{2,t}\mathrm{d}\widetilde{W}_{2,t} + \int_{\mathbb{R}^2}[e^{\ell^\top y_t} - 1]\hat{q}_P(\mathrm{d}y,\mathrm{d}t). \tag{32}$$

The dynamics of the discounted money market account in the forward measure are

$$\mathrm{d}\left(\frac{M_t}{P_{t,T}}\right) = \frac{M_t}{P_{t-,T}}\left[-\sigma_{2,t}\mathrm{d}\widetilde{W}_{2,t} + \int_{\mathbb{R}^2}[e^{-y_{2,t}} - 1]\hat{q}_P(\mathrm{d}y,\mathrm{d}t)\right]. \tag{33}$$

From (32), the forward price can be expressed as a doléans-dade exponential martingale under the forward measure $\mathbb{Q}_P$ as

$$F_{t,T} = F_{0,T}\mathscr{E}_{\mathbb{Q}_P}\left(\int_0^t \sigma_{1,u}\mathrm{d}\widetilde{W}_{1,u} - \int_0^t \sigma_{2,u}\mathrm{d}\widetilde{W}_{2,u}\right)\mathscr{E}_{\mathbb{Q}_P}\left(\sum_{n=1}^{N_t}(Y_{1,T_n} - Y_{2,T_n})\right). \tag{34}$$

5 The Reciprocal Forward Measure

Now we turn our attention to using the stock yield $S_t e^{\int_0^t \zeta_u \mathrm{d}u}$ as the numéraire. Since $\frac{P_{t,T}}{S_t e^{\int_0^t \zeta_u \mathrm{d}u}} = \frac{1}{F_{t,T}}$, we term the measure corresponding to the stock price the reciprocal forward measure. The reciprocal forward price is given as

$$\begin{aligned}\frac{1}{F_{t,T}} &= \frac{P_{t,T}}{S_t e^{\int_0^t \zeta_u \mathrm{d}u}} \\ &= \frac{1}{F_{0,T}}\exp\left[\int_0^t\left(-\frac{\sigma_{2,u}^2}{2} + \frac{\sigma_{1,u}^2}{2}\right)\mathrm{d}t + \int_0^t \sigma_{2,u}\mathrm{d}W_{2,u} - \int_0^t \sigma_{1,u}\mathrm{d}W_{1,u}\right]\end{aligned}$$

$$\times \exp\left[-\lambda \int_0^t (\kappa_{2,u} - \kappa_{1,u})\mathrm{d}u + \sum_{n=1}^{N_t}(Y_{2,T_n} - Y_{1,T_n})\right]. \tag{35}$$

Similarly, we express the value of the money market in terms of units of the stock yield

$$\frac{M_t}{S_t e^{\int_0^t \zeta_u \mathrm{d}u}} = \frac{1}{S_0}\exp\left[\int_0^t \left(\frac{\sigma_{1,u}^2}{2} + \lambda\kappa_{1,u}\right)\mathrm{d}u - \int_0^t \sigma_{1,u}\mathrm{d}W_{1,u} - \sum_{n=1}^{N_t} Y_{1,T_n}\right]. \tag{36}$$

Under the reciprocal forward measure $\mathbb{Q}_S$, the stock yield $S_t e^{\int_0^t \zeta_u \mathrm{d}u}$ is numéraire, and the reciprocal forward price $\left\{\frac{P_{t,T}}{S_t e^{\int_0^t \zeta_u \mathrm{d}u}}\right\}$, and the money market account in terms of units of stock $\left\{\frac{M_t}{S_t e^{\int_0^t \zeta_u \mathrm{d}u}}\right\}$ must be martingales.

The dynamics of the reciprocal forward price are

$$\begin{aligned} d\left(\frac{1}{F_{t,T}}\right) = \frac{1}{F_{t-,T}}\Big[&(\sigma_{1,t}^2 - \rho\sigma_{1,t}\sigma_{2,t})\mathrm{d}t + \sigma_{2,t}\mathrm{d}W_{2,t} - \sigma_{1,t}\mathrm{d}W_{1,t} \\ &- \lambda(\kappa_{2,t} - \kappa_{1,t})\mathrm{d}t + \int_{\mathbb{R}^2}\left[e^{-\ell^\top y_t} - 1\right]N(\mathrm{d}y, \mathrm{d}t)\Big]. \end{aligned} \tag{37}$$

Similarly, the dynamics for the discounted money market account are

$$d\left(\frac{M_t}{S_t e^{\int_0^t \zeta_u \mathrm{d}u}}\right) = \frac{M_t}{S_{t-} e^{\int_0^t \zeta_u \mathrm{d}u}}\left[\sigma_{1,t}^2 \mathrm{d}t - \sigma_{1,t}\mathrm{d}W_{1,t} + \lambda\kappa_{1,t}\mathrm{d}t + \int_{\mathbb{R}^2}\left[e^{-\ell_1^\top y_t} - 1\right]N(\mathrm{d}y, \mathrm{d}t)\right]. \tag{38}$$

From (37) and (38), we see that the choice of $\theta_{S,t} = (-\sigma_{1,t} \ -\rho\sigma_{1,t})^\top$, $\gamma = \ell_1$ and $\nu = 0$ in the (17) in Theorem 1 yields the Radon-Nikodým derivative

$$\begin{aligned} \left.\frac{d\mathbb{Q}_S}{d\mathbb{Q}}\right|_t &= \mathscr{E}_{\mathbb{Q}}\left[-\int_0^t (\Sigma^{-1}\theta_{P,u})^\top \mathrm{d}W_u\right] \times \mathscr{E}_{\mathbb{Q}}\left[\sum_{n=1}^{N_t} \ell_1^\top Y_{T_n}\right] \\ &= \mathscr{E}_{\mathbb{Q}}\left[\int_0^t \sigma_{1,u}\mathrm{d}W_{1,u} + \sum_{n=1}^{N_t} Y_{1,T_n}\right]. \end{aligned} \tag{39}$$

Thus from Girsanov's Theorem for Wiener processes, there exist standard Brownian motion processes $\widehat{W}_{1,t}$ and $\widehat{W}_{2,t}$ under the reciprocal forward measure $\mathbb{Q}_S$ such that

$$\mathrm{d}\widehat{W}_{1,t} = -\sigma_{1,t}\mathrm{d}t + \mathrm{d}W_{1,t}, \tag{40}$$

and

$$\mathrm{d}\widehat{W}_{2,t} = -\rho\sigma_{1,t}\mathrm{d}t + \mathrm{d}W_{2,t}. \tag{41}$$

Furthermore from Theorem 1, it follows that the Poisson process N_t has a new non-homogeneous arrival intensity

$$\hat{\lambda}_t = \lambda\mathbb{E}_{\mathbb{Q}}[e^{\ell_1^\top Y_t}] = \lambda\mathbb{E}_{\mathbb{Q}}[e^{Y_{1,t}}] \tag{42}$$

and the moment generating function of the jump-size distributions under the reciprocal forward measure $\mathbb{Q}_S$ is

$$M_{\mathbb{Q}_S,Y_t}(u) = \frac{M_{\mathbb{Q},Y_t}(u+\ell_1)}{M_{\mathbb{Q},Y_t}(\ell_1)}. \tag{43}$$

We also define the compensated Poisson counting measure under the reciprocal forward measure $\mathbb{Q}_S$ as

$$\hat{q}_S(\mathrm{d}y, \mathrm{d}t) = N(\mathrm{d}y, \mathrm{d}t) - \hat{\lambda}_t m_{\mathbb{Q}_S,t}(\mathrm{d}y)\mathrm{d}t. \tag{44}$$

Thus the dynamics for the reciprocal forward price as a martingale in the reciprocal forward measure $\mathbb{Q}_S$ are

$$\frac{\mathrm{d}F_{t,T}^{-1}}{F_{t-,T}^{-1}} = \sigma_{2,t}\mathrm{d}\widehat{W}_{2,t} - \sigma_{1,t}\mathrm{d}\widehat{W}_{1,t} + \int_{\mathbb{R}^2}[e^{-\ell^\top y_t} - 1]\hat{q}_S(\mathrm{d}y, \mathrm{d}t). \tag{45}$$

The dynamics of the discounted money market account in the forward measure are

$$\mathrm{d}\left(\frac{M_t}{S_t e^{\int_0^t \zeta_u \mathrm{d}u}}\right) = \frac{M_t}{S_{t-}e^{\int_0^t \zeta_u \mathrm{d}u}}\left[-\sigma_{1,t}\mathrm{d}\widehat{W}_{1,t} + \int_{\mathbb{R}^2}[e^{-y_{1,t}} - 1]\hat{q}_S(\mathrm{d}y, \mathrm{d}t)\right]. \tag{46}$$

From (45), the reciprocal forward price can be expressed as a doléans-dade martingale under the measure $\mathbb{Q}_S$ as

$$F_{t,T}^{-1} = F_{0,T}^{-1}\mathscr{E}_{\mathbb{Q}_S}\left(\int_0^t \sigma_{2,u}d\widehat{W}_{2,u} - \int_0^t \sigma_{1,u}d\widehat{W}_{1,u}\right)\mathscr{E}_{\mathbb{Q}_S}\left(-\sum_{n=1}^{N_t}(Y_{1,T_n} - Y_{2,T_n})\right). \tag{47}$$

From (32) and (45), we see that the forward price $F_{t,T}$ and the reciprocal forward price $F_{t,T}^{-1}$ both have the same Wiener volatility. From (27) and (40), we see that

$$d\widehat{W}_{1,t} = -(\sigma_{1,t} - \rho\sigma_{2,t})\mathrm{d}t + d\widetilde{W}_{1,t}, \tag{48}$$

and similarly from (28) and (41),

$$\mathrm{d}\widehat{W}_{2,t} = -(\rho\sigma_{1,t} - \sigma_{2,t})\mathrm{d}t + \mathrm{d}\widetilde{W}_{2,t}. \tag{49}$$

6 Pricing a European Call Option

We consider a European call option $X_t(S_t)$ with strike price K and maturing at T, the same time as the bond $P_{t,T}$. The usual risk-neutral valuation of the option price at time t is

$$X_t(S_t) = M_t \mathbb{E}_{\mathbb{Q}}\left[\frac{(S_T - K)^+}{M_T}\middle|\mathscr{F}_t\right], \tag{50}$$

where M_t is the money market account. The change of numéraire result in Geman et al. (1995) is frequently used to decompose the conditional expectation on the right hand side of (50) into two terms which can be intepreted as the conditional probabilities of the option being in-the-money at maturity. Although the result is normally applied to stock price processes modelled as continuous semi-martingales, we note that the result is also applicable to a stock and bond market driven by jump-diffusion dynamics, as in our model. The following theorem is an adaptation of the change of numéraire result in Theorem 2 in Geman et al. (1995) to our model.

Theorem 2 *Suppose the dynamics of the stock and zero-coupon bond prices are (2) and (3) respectively, and assumptions 1 and 2 are satisfied. The money market account is* $M_t = \exp(\int_0^t r_u \mathrm{d}u)$. *Then the risk-neutral valuation of a European style call option (50) is equivalent to*

$$X_t(S_t) \equiv X_t(F_{t,T}) = S_t e^{-\int_t^T \zeta_u \mathrm{d}u}\mathbb{Q}_S\left[\frac{1}{F_{t,T}} < \frac{1}{K}\middle|\mathscr{F}_t\right] - KP_{t,T}\mathbb{Q}_P\left[F_{t,T} > K\middle|\mathscr{F}_t\right], \tag{51}$$

where $F_{t,T}$ *is the forward price of the stock given by (22).*

Proof Without loss of generality, we provide the proof for the option price at time $t = 0$. We denote the event that the option is in-the-money at maturity by

$$\mathscr{A} = \{S_T > K\} = \left\{F_{T,T} > K\right\}. \tag{52}$$

Thus

$$\begin{aligned} X_0(S_0) &= \mathbb{E}_{\mathbb{Q}}\left[\frac{(S_T - K)^+}{M_T}\right] = \mathbb{E}_{\mathbb{Q}}\left[\frac{(S_T - K)}{M_T}\mathbb{1}_{\mathscr{A}}\right] \\ &= S_0 e^{-\int_0^T \zeta_u \mathrm{d}u}\mathbb{E}_{\mathbb{Q}}\left[\frac{S_T e^{\int_0^T \zeta_u \mathrm{d}u}}{S_0 M_T}\mathbb{1}_{\mathscr{A}}\right] - KP_{0,T}\mathbb{E}_{\mathbb{Q}}\left[\frac{1}{P_{0,T}M_T}\mathbb{1}_{\mathscr{A}}\right]. \end{aligned} \tag{53}$$

In (53), we note that $\frac{S_T e^{\int_0^T \zeta_u \mathrm{d}u}}{S_0 M_T}$ and $\frac{1}{P_{0,T} M_T} = \frac{P_{T,T}}{P_{0,T} M_T}$ are the Radon-Nikodým derivatives (39) and (26) respectively. Hence it follows that

$$\begin{aligned} X_0(S_0) &= S_0 e^{-\int_0^T \zeta_u \mathrm{d}u} \mathbb{E}_{\mathbb{Q}_S}[\mathbb{1}_{\mathscr{A}}] - K P_{0,T} \mathbb{E}_{\mathbb{Q}_P}[\mathbb{1}_{\mathscr{A}}] \\ &= S_0 e^{-\int_0^T \zeta_u \mathrm{d}u} \mathbb{Q}_S[S_T > K] - K P_{0,T} \mathbb{Q}_P[S_T > K] \\ &= S_0 e^{-\int_0^T \zeta_u \mathrm{d}u} \mathbb{Q}_S\left[\frac{1}{F_{T,T}} < \frac{1}{K}\right] - K P_{0,T} \mathbb{Q}_P\left[F_{T,T} > K\right]. \end{aligned}$$

□

Now we provide an extension of the option pricing formula given by Merton (1976) where the stock price is driven by a jump-diffusion process. In our extension, we include a stochastic bond with HJM-type dynamics given by (3). A nice formula in terms of a Poisson average of log-normal distributions can be obtained if we assume that the jump-sizes for both the stock and the bond are jointly distributed as a bivariate normal. From this point onwards, we make the following assumption:

Assumption 4 Under the risk-neutral measure $\mathbb{Q}$, the jump-sizes $Y_t = (Y_{1,t}, Y_{2,t})^\top$ over the infinitesimally small interval $[t, t + \mathrm{d}t)$ are distributed as a bivariate normal with mean vector $\alpha_t = (\alpha_{1,t}, \alpha_{2,t})^\top$ and with a covariance matrix given as

$$\Sigma_{Y_t} = \begin{pmatrix} \delta_{1,t}^2 & \rho_y \delta_{1,t} \delta_{2,t} \\ \rho_y \delta_{1,t} \delta_{2,t} & \delta_{2,t}^2 \end{pmatrix}, \tag{54}$$

where the variance of $Y_{1,t}$ is $\delta_{1,t}^2$, the variance of $Y_{2,t}$ is $\delta_{2,t}^2$, and the correlation between $Y_{1,t}$ and $Y_{2,t}$ is ρ_y.

Remark 2 From Assumption 3, the conditions $\lim_{t \to T} \alpha_{2,t} = 0$ and $\lim_{t \to T} \sigma_{2,t} = 0$ in Assumption 4 must be satisfied.

The next theorem provides for an option pricing formula when the jump-sizes of the stock and the bond are assumed to be correlated bivariate normals. For ease of notation, denote the covariance matrix of the stock and zero-coupon bond log returns arising from the Wiener components as

$$\Sigma_{W,t} = \begin{pmatrix} \sigma_{1,t}^2 & \rho \sigma_{1,t} \sigma_{2,t} \\ \rho \sigma_{1,t} \sigma_{2,t} & \sigma_{2,t}^2 \end{pmatrix} \tag{55}$$

The option pricing formula requires the joint-density of the jump-arrival times $0 < T_1 < T_2 < \cdots < T_n \leq T$ given $N_T = n$ under both the forward measure $\mathbb{Q}_P$ and the reciprocal forward measure $\mathbb{Q}_S$. A general expression for the joint-density is given in Lemma A.1 in the Appendix. We denote $g_{\mathbb{Q}_P}(t_1, \ldots, t_n)$ the joint-density under the forward measure and $g_{\mathbb{Q}_S}(t_1, \ldots, t_n)$ the joint-density under the reciprocal forward measure.

Theorem 3 *Suppose the dynamics of the stock and zero-coupon bond prices are (2) and (3) respectively, and assumptions 1, 2, 3 and 4 are satisfied. Then the risk-neutral valuation of a European style call option (50) is equivalent to*

$$
\begin{aligned}
X_0(S_0) = & \sum_{n=0}^{\infty} \exp(-\lambda T) \frac{\lambda^n}{n!} \\
& \times \left[S_0 e^{-\int_0^T (\zeta_u + \lambda \kappa_{1,u}) \mathrm{d}u} \int_0^T e^{\alpha_{1,u} + \frac{\delta_{1,u}^2}{2}} \mathrm{d}u \right. \\
& \times \int_0^T \cdots \int_0^{t_3} \int_0^{t_2} \Phi(d_{1,t_1,\cdots,t_n}) g_{\mathbb{Q}_S}(t_1, \cdots, t_n) \mathrm{d}t_1 \cdots \mathrm{d}t_n \\
& - K P_{0,T} e^{-\lambda \int_0^T \kappa_{2,u} \mathrm{d}u} \int_0^T e^{\alpha_{2,u} + \frac{\delta_{2,u}^2}{2}} \mathrm{d}u \\
& \left. \times \int_0^T \cdots \int_0^{t_3} \int_0^{t_2} \Phi(d_{2,t_1,\cdots,t_n}) g_{\mathbb{Q}_P}(t_1, \cdots, t_n) \mathrm{d}t_1 \cdots \mathrm{d}t_n \right],
\end{aligned}
\tag{56}
$$

where

$$
d_{1,t_1,\cdots,t_n} = \frac{\ln\left(\frac{S_0 e^{-\int_0^T \zeta_u \mathrm{d}u}}{K P_{0,T}}\right) + \int_0^T \left(\frac{\sigma_{\hat{W},u}^2}{2} + \hat{\lambda}_u \hat{\kappa}_u\right) \mathrm{d}u + \sum_{0<t_1<\cdots<t_n\le T} \ell^\top (\alpha_{t_i} + \Sigma_{Y_{t_i}} \ell_1)}{\sqrt{\int_0^T \sigma_{t,t_1,\cdots,t_n}^2 \mathrm{d}t}},
$$

$$
d_{2,t_1,\cdots,t_n} = \frac{\ln\left(\frac{S_0 e^{-\int_0^T \zeta_u \mathrm{d}u}}{K P_{0,T}}\right) - \int_0^T \left(\frac{\sigma_{\hat{W},u}^2}{2} + \tilde{\lambda}_u \tilde{\kappa}_u\right) \mathrm{d}u - \sum_{0<t_1<\cdots<t_n\le T} \ell^\top (\alpha_{t_i} + \Sigma_{Y_{t_i}} \ell_2)}{\sqrt{\int_0^T \sigma_{t,t_1,\cdots,t_n}^2 \mathrm{d}t}};
$$

with

$$
\begin{aligned}
\kappa_{1,u} &= \exp\left(\alpha_{1,u} + \frac{\delta_{1,u}^2}{2}\right) - 1, \qquad \kappa_{2,u} = \exp\left(\alpha_{2,u} + \frac{\delta_{2,u}^2}{2}\right) - 1, \\
\hat{\lambda}_u &= \lambda(1 + \kappa_{1,u}), \qquad \tilde{\lambda}_u = \lambda(1 + \kappa_{2,u}), \\
\hat{\kappa}_u &= \exp\left(-\ell^\top(\alpha_u + \Sigma_{Y_u} \ell_1) + \frac{\ell^\top \Sigma_{Y_u} \ell}{2}\right) - 1, \\
\tilde{\kappa}_u &= \exp\left(\ell^\top(\alpha_u + \Sigma_{Y_u} \ell_2) + \frac{\ell^\top \Sigma_{Y_u} \ell}{2}\right) - 1,
\end{aligned}
$$

$$\sigma^2_{W,u} = \ell^\top \Sigma_{W,u} \ell, \qquad \sigma^2_{t,t_1,\cdots,t_n} = \sigma^2_{W,t} + \frac{1}{T} \sum_{0<t_1<\cdots<t_n\leq T} \ell^\top \Sigma_{Y_{t_i}} \ell.$$

Note. *In Theorem 3, expressions of the form* $\sum_{0<t_1<\cdots<t_n\leq T} a_{t_i}$ *denote the sum* $a_{t_1} + a_{t_2} + \cdots + a_{t_n}$ *where* $0 < t_1 < \cdots < t_n \leq T$.

Proof We use the result (51) from Theorem 2. Now the probability of the option being in-the-money under the reciprocal forward measure is

$$\begin{aligned}
\mathbb{Q}_S\left[\frac{1}{F_{T,T}} < \frac{1}{K}\right] &= \mathbb{E}_{\mathbb{Q}_S}\left[\mathbb{E}_{\mathbb{Q}_S}\left[\mathbb{1}_{\mathscr{A}}|N_T\right]\right] \\
&= \sum_{n=0}^{\infty} \exp\left(-\int_0^T \hat{\lambda}_u \mathrm{d}u\right) \frac{(\int_0^T \hat{\lambda}_u \mathrm{d}u)^n}{n!} \\
&\quad \times \mathbb{Q}_S\left(\frac{1}{F_{T,T}} < \frac{1}{K}\middle| N_T = n\right) \\
&= \sum_{n=0}^{\infty} \exp\left(-\int_0^T \hat{\lambda}_u \mathrm{d}u\right) \frac{(\int_0^T \hat{\lambda}_u \mathrm{d}u)^n}{n!} \\
&\quad \times \mathbb{Q}_S\left(\ln\left(\frac{F_{0,T}}{F_{T,T}}\right) < \ln\left(\frac{F_{0,T}}{K}\right)\middle| N_T = n\right). \qquad (57)
\end{aligned}$$

We note that under the reciprocal forward measure $\mathbb{Q}_S$, the log-return of the reciprocal forward price $\ln\left(\frac{F_{0,T}}{F_{T,T}}\right)$, conditional on $N_T = n$ and $T_1 = t_1, \ldots, T_n = t_n$, is normally distributed as

$$N\left(-\int_0^T \left(\hat{\lambda}_u \hat{\kappa}_u + \frac{\sigma^2_{W,u}}{2}\right)\mathrm{d}u - \sum_{0<t_1<\cdots<t_n\leq T} \ell^\top(\alpha_{t_i} + \Sigma_{Y_{t_i}}\ell_1), \int_0^T \sigma^2_{t,t_1,\cdots,t_n}\mathrm{d}t\right),$$

from (47). Thus

$$\begin{aligned}
&\mathbb{Q}_S\left[\frac{1}{F_{T,T}} < \frac{1}{K}\right] \\
&= \sum_{n=0}^{\infty} e^{-\int_0^T \hat{\lambda}_u \mathrm{d}u} \frac{(\int_0^T \hat{\lambda}_u \mathrm{d}u)^n}{n!} \int_0^T \cdots \int_0^{t_3} \int_0^{t_2} \Phi(d_{1,t_1,\cdots,t_n}) g_{\mathbb{Q}_S}(t_1, \cdots, t_n) \mathrm{d}t_1 \cdots \mathrm{d}t_n.
\end{aligned} \qquad (58)$$

On the other hand, the probability of the option being in-the-money under the forward measure is

$$\begin{aligned}
\mathbb{Q}_P\left[F_{T,T} > K\right] &= \mathbb{E}_{\mathbb{Q}_P}\left[\mathbb{E}_{\mathbb{Q}_P}\left[\mathbb{1}_{\mathscr{A}}|N_T\right]\right] \\
&= \sum_{n=0}^{\infty} \exp\left(-\int_0^T \tilde{\lambda}_u \mathrm{d}u\right) \frac{(\int_0^T \tilde{\lambda}_u \mathrm{d}u)^n}{n!} \mathbb{Q}_P\left(F_{T,T} > K \middle| N_T = n\right) \\
&= \sum_{n=0}^{\infty} \exp\left(-\int_0^T \tilde{\lambda}_u \mathrm{d}u\right) \frac{(\int_0^T \tilde{\lambda}_u \mathrm{d}u)^n}{n!} \\
&\times \mathbb{Q}_P\left(\ln\left(\frac{F_{T,T}}{F_{0,T}}\right) > \ln\left(\frac{K}{F_{0,T}}\right) \middle| N_T = n\right) \\
&= \sum_{n=0}^{\infty} \exp\left(-\int_0^T \tilde{\lambda}_u \mathrm{d}u\right) \frac{(\int_0^T \tilde{\lambda}_u \mathrm{d}u)^n}{n!} \\
&\times \mathbb{Q}_P\left(-\ln\left(\frac{F_{T,T}}{F_{0,T}}\right) < -\ln\left(\frac{K}{F_{0,T}}\right) \middle| N_T = n\right). \qquad (59)
\end{aligned}$$

We note that under the forward measure $\mathbb{Q}_P$, the log-return of the forward price $\ln\left(\frac{F_{T,T}}{F_{0,T}}\right)$, conditional on $N_T = n$ and $T_1 = t_1, \cdots, T_n = t_n$, is normally distributed as

$$N\left(-\int_0^T \left(\tilde{\lambda}_u \tilde{\kappa}_u + \frac{\sigma_{W,u}^2}{2}\right)\mathrm{d}u + \sum_{0<t_1<\cdots<t_n\le T} \ell^\top(\alpha_{t_i} + \Sigma_{Y_{t_i}}\ell_2), \int_0^T \sigma_{t,t_1,\cdots,t_n}^2 \mathrm{d}t\right)$$

from (34). Thus

$$\begin{aligned}
&\mathbb{Q}_P\left[F_{T,T} > K\right] \\
&= \sum_{n=0}^{\infty} e^{-\int_0^T \tilde{\lambda}_u \mathrm{d}u} \frac{(\int_0^T \tilde{\lambda}_u \mathrm{d}u)^n}{n!} \int_0^T \cdots \int_0^{t_3} \int_0^{t_2} \Phi(\mathrm{d}_{2,t_1,\cdots,t_n}) g_{\mathbb{Q}_P}(t_1, \cdots, t_n) \mathrm{d}t_1 \cdots \mathrm{d}t_n. \qquad (60)
\end{aligned}$$

The substitution of (58) and (60) into (51) for the case when $t = 0$ and further simplification completes the proof. □

Remark 3 Under Merton's (1976) jump-diffusion model, the interest rate $r_t = r$ is assumed constant. The stock price (2) in Merton's model has constant parameters $\sigma_{1,t} = \sigma_1$, and relative jump-sizes Y_{1,T_n} that are independent and identically distributed normals $N(\alpha_1, \delta_1^2)$. In place of (3), the bond price dynamics is given by

$$\mathrm{d}P_{t,T} = rP_{t,T}\mathrm{d}t, \qquad (61)$$

with $P_{T,T} = 1$ and all other parameters set to zero. It is easy to see that the forward measure $\mathbb{Q}_P$ and the risk-neutral measure $\mathbb{Q}$ are identical under Merton's (1976) model. The forward price volatility is the same as the volatility of the stock price since the non-stochastic bond does not have any volatility. Merton's (1976) option pricing formula can be obtained as a special case of our model.

Remark 4 In place of a stock and bond, it is possible to derive a pricing formula for a European style exchange option where the dynamics of both stock prices are driven by jump-diffusion processes. A similar change of numéraire approach is taken in Cheang and Chiarella (2011) in the derivation of the pricing formula for the exchange option under jump-diffusion dynamics.

7 Conclusion

This paper has extended the results of Geman et al. (1995) and Merton (1976) for pricing a European call option on the stock in the case where the stock price and bond price returns both exhibit jump-diffusion characteristics. This allows us to incorporate a pricing model for the call option where the bond price dynamics are also discontinuous. Merton's (1976) jump-diffusion model for option pricing is a special case of our model.

Appendix

In this appendix, we state a result involving the distribution of the arrival times $(T_1, T_2, \ldots, T_n)$ of a non-homogeneous Poisson process N_T conditioned on n arrivals $N_T = n$ over the interval $(0, T]$.

Lemma A.1 *Consider a non-homogeneous Poisson process N_t with intensity λ_t under some probability measure $\mathbb{P}$. Let the time of the ith arrival be $T_i = t_i$ where $0 < t_1 < t_2 < \cdots < t_n \leq T$. Conditioned on $N_T = n$, the joint density of the times of the arrivals is*

$$g_{\mathbb{P}}(t_1, \cdots, t_n) = \begin{cases} \frac{n! \prod_{i=1}^{n} \lambda_{t_i}}{\left[\int_0^T \lambda_t \mathrm{d}t\right]^n} \text{ for } 0 < t_1 < t_2 < \cdots < t_n \leq T, \\ 0 \qquad \text{otherwise.} \end{cases} \tag{A.1}$$

The proof can be found in standard texts on point processes, e.g. Daley and Vere-Jones (1988). In the context of Theorem 3 in this paper, for $0 < t_1 < t_2 < \cdots < t_n \leq T$, the joint density $g_{\mathbb{Q}_S}(t_1, \ldots, t_n) = n! \frac{\prod_{i=1}^{n} \hat{\lambda}_{t_i}}{\left[\int_0^T \hat{\lambda}_t \mathrm{d}t\right]^n}$ and $g_{\mathbb{Q}_P}(t_1, \ldots, t_n) = n! \frac{\prod_{i=1}^{n} \tilde{\lambda}_{t_i}}{\left[\int_0^T \tilde{\lambda}_t \mathrm{d}t\right]^n}$.

References

Applebaum, D. (2004). *Lévy processes and stochastic calculus*. Cambridge: Cambridge University Press.

Ball, C., & Torus, W. (1985). On jumps in common stock prices and their impact on call option pricing. *Journal of Finance*, *40*, 155–173.

Björk, T., Kabanov, Y., & Runggaldier, W. J. (1997a). Bond market structure in the presence of marked point processes. *Mathematical Finance*, *7*(2), 211–223.

Björk, T., Di Masi, G., Kabanov, Y., & Runggaldier, W. J. (1997b). Towards a general theory of bond markets. *Finance and Stochastics*, *1*, 141–174.

Black, F., & Scholes, M. (1972). The pricing of options and corporate Liabilities. *Journal of Political Economy*, *81*(3), 637–654.

Cheang, G. H. L., & Chiarella, C. (2011). Exchange options under jump-diffusion dynamics. *Applied Mathematical Finance*, *18*(3), 245–276.

Chiarella, C., & Nikitipoulos-Sklibosios, C. (2003). A class of jump-diffusion bond pricing models within the HJM framework. *Asia-Pacific Financial Markets*, *10*, 87–127.

Colwell, D. B., & Elliott, R. J. (1993). Discontinuous asset prices and non-attainable contingent claims. *Mathematical Finance*, *3*, 295–308.

Cont, R., & Tankov, P. (2004a). *Financial modelling with jump processes*. London: Chapman and Hall/CRC.

Cont, R., & Tankov, P. (2004b). Calibration of jump-diffusion option-pricing models: a robust non-parametric approach. *Journal of Computational Finance*, *7*, 1–49.

Das, S. R. (2002). The surprise element: Jumps in interest rates. *Journal of Econometrics*, *106*, 27–65.

Duffie, D., Pan, J., & Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, *68*(6), 1343–1376.

Dungey, M., McKenzie, M., & Smith, V. (2007). *Empirical evidence on jumps in the term structure of the U.S. treasury market*. Working Paper, CAMA.

Daley, D. J., & Vere-Jones, D. (1988). *An introduction to the theory of point processes*. New York: Springer.

Geman, H., El Karoui, N., & Rochet, J.-C. (1995). Change of numéraire, changes of probability measure and option pricing. *Journal of Applied Probability*, *32*(2), 443–459.

Harrison, J. M., & Pliska, S. R. (1981). Martingales and stochastic integrals in the theory of continous trading. *Stochastic Processes and Applications*, *11*, 215–260.

Heath, D., Jarrow, R., & Morton, A. J. (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica*, *60*(1), 77–105.

Jeanblanc, M., Klöppel, S., & Miyahara, Y. (2007). Minimal F^Q-martingale measures for exponential Lévy processes. *Annals of Applied Probability*, *17*, 1615–1638.

Merton, R. C. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Science*, *4*, 141–183.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, *3*, 125–144.

Miyahara, Y. (2001). Geometric Lévy process & MEMM pricing model and related estimation problems. *Asia-Pacific Financial Markets*, *8*, 45–60.

Protter, P. (1990). *Stochastic integration and differential equations: A new approach*. Berlin: Springer.

Runggaldier, W. J. (2003). Jump diffusion models. In S. T. Rachev (Ed.), *Handbook of heavy tailed distributions in finance* (pp. 169–209). North-Holland: Elsevier.