Statistics and Econometrics for Finance

Yong Zeng
Shu Wu *Eds.*

State-Space Models

Applications in Economics and Finance

Springer

Statistics and Econometrics for Finance

For further volumes:
http://www.springer.com/series/10377

Yong Zeng • Shu Wu
Editors

State-Space Models

Applications in Economics and Finance

Springer

Editors

Yong Zeng
Department of Mathematics and Statistics
University of Missouri at Kansas City
Kansas City, Missouri, USA

Shu Wu
Department of Economics
The University of Kansas
Lawrence, Kansas, USA

ISBN 978-1-4614-7788-4 ISBN 978-1-4614-7789-1 (eBook)
DOI 10.1007/978-1-4614-7789-1
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2013943592

Printed on acid-free paper

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

Preface

Since the seminal papers of Kalman (1960, 1961) and the early development in the field of engineering, state-space models have become an increasingly important tool for research in finance and economics in recent years. This book is a collection of contributed papers that reflect this new phase of theoretical developments of state-space models and their applications in economics and finance. We hope that the breadth of research shared in this volume will serve as an inspiration and a valuable reference for future users of state-space models in these two areas.

A generic state-space model consists of two equations. One describes how observable economic variables relate to potentially unobservable *state variables*, and the other describes how state variables evolve over time. Both state and observable variables can be either discrete- or continuous-time stochastic processes.

This book is divided into four parts. In Parts I and II, we mainly consider discrete-time state-space models. Let Y_t be an $n \times 1$ observable variable and X_t be a $k \times 1$ state variable. Then, a state-space model can be written as follows:

$$Y_t = f(X_t, \theta, \varepsilon_t) \tag{1}$$

$$X_{t+1} = g(X_t, \theta, \eta_{t+1}) \tag{2}$$

where θ is an $m \times 1$ vector of model parameters, and ε_t and η_t are independent and identically distributed random shocks (or noises). $f(\cdot)$ and $g(\cdot)$ are $n \times 1$ and $k \times 1$ (non)linear functions, respectively. Because of its flexibility, many discrete-time models in economics and finance can be represented in this state-space form. These models include autoregressive moving average (ARMA) models, regression models with time-varying coefficients, dynamic factor models (DFM), models with stochastic volatility (SV), regime-switching models, and hidden Markov models (HMM). In these models, the parameter θ is typically unknown. Therefore, one central question in many applications is to estimate θ and to conduct statistical inferences of the model (e.g., hypothesis testing). Researchers are also highly interested in obtaining unbiased and efficient estimates of the underlying state variables X_t.

In Parts III and IV, we include models where the state variable, $X(t)$, can be a continuous-time finite-state Markov chain as that in hidden Markov or

regime-switching models of Chaps. 8–11. $X(t)$ can also be a general continuous-time Markov process described by a stochastic differential equation as those in Chaps. 12–15. The Markov processes include a geometric Brownian motion (GBM) and a jump-diffusion process with regime switching among others. The observation process, $Y(t)$, can be a continuous-path process (Chaps. 10 and 11), an equally spaced time series (Chap. 14), or an irregularly spaced point process (Chaps. 8, 13, and 15). These models have found many applications in finance such as pricing of credit risk, optimal trading rules and hedging, optimal annuity purchasing or dividend policies, optimal learning in financial markets, and modeling (ultra) high-frequency data.

Below, we provide a more detailed description of each part.

Part I includes three chapters on *Particle Filtering and Parameter Learning in Nonlinear State-Space Models*. The introduction of particle filters has had a major impact on the development of nonlinear and non-Gaussian state-space models. This technique has expanded the range of practical applicability of state-space models to cases with high-dimension state space.

In Chap. 1, Tze Leung Lai and Vibhav Bukkapatanam provide a review of the estimation of the latent state variables using particle filters with known or unknown model parameters. They also present a new adaptive particle filter that uses a computationally efficient Markov Chain Monte Carlo estimate of the posterior distribution of the state-space model parameters in conjunction with sequential state estimation. They describe several applications in finance and economics, including frailty models of portfolio default probabilities, SV models with contemporaneous price and volatility jumps, and hidden Markov models for high-frequency transaction data.

In Chap. 2, Maria Paula Rios and Hedibert Freitas Lopes explore kernel smoothing and conditional sufficient statistics extensions of the auxiliary particle filters (Pitt and Shepard 1999) and bootstrap filters (Gordon, Salmond and Smith 1993). Using simulated data from SV models with Markov switching, they show that the Liu-West particle filter degenerates and has the largest Monte Carlo error, while their auxiliary particle filter extended with sufficient statistics (APF + SS) has a much better performance. Their APF + SS filter takes advantage of recursive sufficient statistics that are sequentially tracked and whose behavior resembles that of a latent state with conditionally deterministic updates. They also assess the performance of the APF + SS filter in sequential estimation in examples with real data.

In Chap. 3, Alexandre J. Chorin, Matthias Morzfeld, and Xuemin Tu review the implicit particle filter. The key idea is to concentrate the particles on the high-probability regions of the target probability density function (pdf) so that the number of particles required for a good approximation of the pdf remains manageable even if the state space has high dimensions. They explain how this idea is implemented, discuss special cases of practical importance, and show the relations of the implicit particle filter with other data assimilation methods. They further illustrate the method with examples such as SV, stochastic Lorenz attractor, stochastic

Kuramoto–Sivashinsky equation, and data assimilations.

Part II includes four chapters on the application of *Linear State-Space Models in Macroeconomics and Finance*.

In Chap. 4, Yulei Luo, Jun Nie, and Eric Young explicitly solve a linear-quadratic macroeconomic model under model uncertainty (due to concerns of model misspecification) and state uncertainty (due to limited information constraint). They show that the model can be mapped to a state-space representation that can be used to quantify the key parameters of model uncertainty. They demonstrate through examples how this framework can be used to study a range of interesting questions in macroeconomics and international finance such as explaining current account dynamics and resolving the international consumption puzzle.

In Chap. 5, Pym Manopimoke estimates a state-space model of the inflation dynamics in Hong Kong. The model allows her to decompose Hong Kong inflation into a stochastic trend and a stationary cycle component that can be driven by both domestic and foreign economic variables such as output gaps. This empirical model is consistent with economic theories of inflation and output, offering new insight into the determination of trend and cyclical inflation in Hong Kong. This is an example of the power of state-space models in empirical macroeconomic research.

In recent years, vector autoregression models (VARs) have become a primary tool for investigating dynamic relationship between multiple economic variables. One challenge, however, is that such relationships are often evolving over time as a result of shifts in government policies or structural changes in the economy. In Chap. 6, Taeyoung Doh and Michael Connolly show that the state-space representation is a useful tool to estimate VARs with time-varying coefficients and/or SV. They show that these models can better capture the changing relationships between important macroeconomic variables.

Chapter 7 is an application to finance. Jun Ma nd Mark Wohar use a state-space model to address one important issue regarding sources of stock market volatility. They argue that the existing empirical studies have focused on point estimation and lack robust statistical inference. The authors show that the small signal-to-noise ratio has made the market data contain too little useful information for researchers to reach robust conclusions about the relative importance of different sources of stock market volatility.

Part III includes five chapters on *Hidden Markov Models* (*HMM*), *Regime Switching*, and *Mathematical Finance*.

Chapters 8 and 9 are on hidden Markov models and their applications to finance. In Chap. 8, Robert Elliott and Tak Kuen Siu discuss an intensity-based model of portfolio credit risk using a collection of hidden Markov-modulated single jump processes. The model is a dynamic version of a frailty model casted in state-space form, able to describe dependent default risks among firms that are exposed to a common hidden dynamic frailty factor. The authors develop filtering equations and filter-based estimates of the model in recursive forms. They also obtain the joint default probability of reference entities in a credit portfolio as well as the

variance dynamics for both observations and hidden states. In Chap. 9, Xiaojing Xi and Rogemar Mamon develop a weak hidden Markov model (WHMM) for the term structure of interest rates where the means and volatilities of bond yields are governed by a second-order Markov chain in discrete time. The authors use the multivariate filtering technique in conjunction with the EM algorithm to estimate the model parameters. They assess the goodness of fit of the model based on out-of-sample forecasts and apply AIC to determine the optimal number of regimes in their model. They apply the model to a data set of daily Treasury yields in the USA. The empirical results show that their WHMM outperforms the standard HMM in terms of out-of-sample forecasts.

Chapters 10 and 11 are applications of regime-switching models to insurance risk and optimal trading rule. Models with regime switching usually don't have analytical solutions to the associated stochastic control problems. In Chap. 10, Zhuo Jin and George Yin discuss numerical methods for solving stochastic optimization problems involving regime-switching models. They propose a numerical solution to the system of HJB equations based on Markov chain approximation. They show how these regime-switching models can be applied to analyze optimal annuity purchasing and optimal dividend payment strategy problems. In Chap. 11, Eunju Sohn and Qing Zhang study an optimal trading rule problem where the underlying asset price is governed by a mean-reverting process with regime switching. The investor's objective is to buy and sell the asset so as to maximize the overall return. The authors consider the case in which the jump rates of the Markov chain can go to infinite. They study the asymptotic properties of the limit value functions and establish a limiting problem which is easier to solve. They show that the solution to the limiting problem can be used to construct a trading rule that is nearly optimal.

In Chap. 12, Mingming Wang and Allanus Tsoi discuss Constant Proportion Portfolio Insurance (CPPI) problem with jump diffusion. They also consider the associated problem of hedging using both the PDE/PIDE and martingale approaches. In particular, they consider the mean-variance hedging problem when the contingent claim is a function of the CPPI portfolio value.

Part IV includes three chapters on *Nonlinear State-Space Models for High-Frequency Financial Data*.

In Chap. 13, combining classical Kyle and Glosten–Milgrom models, Yoonjung Lee proposes a new state-space modeling framework under asymmetric information. The model is able to describe the interactions among some important variables in financial markets such as the price impact of a trade, the duration between trades, and the degree of information asymmetry. In the model, a private signal is partially revealed through trades, while new public information arrives continuously at the market. In order to set a competitive price that rationally incorporates these two sources of information, the market maker utilizes Bayesian learning. The author derives the corresponding nonlinear filtering equation using anticipative Girsanov transformation. She further proves the existence and uniqueness of the ask and bid prices using an SPDE approach. The pricing rule depends on the actual sequence of order arrivals, not just the total number of buy/sell orders. The price impact of a

trade tends to decrease when the duration between trades gets longer. The speed at which the information gets incorporated into the price depends on the quality of the private signal and the trading rate of informed traders.

Chapter 14 is concerned with volatility estimation and prediction. A popular approach is to use the high-frequency data to estimate volatilities and then fit a low-frequency AR volatility model for forecasting. While the empirical performance of this approach is good, there is a lack of theoretical foundation. In this chapter, Yazhen Wang and Xin Zhang show that, for rather general underlying price and volatility processes, the realized volatility estimators approximately follow a heterogeneous autoregressive model, hence providing theoretical justifications of the popular approach. An important feature of the model is that the two- or multi-scaled realized volatility estimators employed are based on a state-space model, where the prices from high-frequency transactions may include market microstructure noise.

Chapter 15 is concerned with estimating models for ultra-high frequency data. The class of models has a random-arrival-time state-space form that explicitly accommodates market microstructure noises in asset price. Although the model is able to capture stylized facts of tick data, the nonlinear state-space model structure makes parameter estimation a challenge. Cai Zhu and James Huang apply particle Markov Chain Monte Carlo (PMCMC) method to estimate a couple models when the underlying intrinsic value processes follow a GBM or a jump-diffusion process. They show that the PMCMC method is able to yield reasonable estimates of the model parameters and further discuss numeric methods that are able to enhance the efficiency of the algorithm.

We would like to express our gratitude to all the contributors of the book chapters for their efforts in making their research accessible to a wide range readers. We hope the book can lead to more interdisciplinary research among economists, mathematicians, and statisticians. We also would like to thank Yaozhong Hu of University of Kansas, Neng Wang of Columbia University, and Zhenxiao Wu of National University of Singapore for their generous help during the preparation of this book.

We want to thank Brian Foster, a former Springer editor, for his enthusiasm and help in the early stage of this book project. We also want to thank Hannah Bracken, Marc Strauss, Nicolas Philipson, and William Curtis of Springer New York. Their continuing support and commitment throughout the project are highly appreciated. We owe a special thanks to Hannah Bracken for her superb editorial assistance.

Yong Zeng's research is supported by National Science Foundation under the grants of TG-DMS100019 and DMS-1228244 and a grant from the University of Missouri Research Board in 2011. He gratefully acknowledges those grant supports. Finally, we would like to thank our families for their continuous support, which makes the completion of this project possible.

Kansas City, MO
Lawrence, KS

Yong Zeng
Shu Wu

Contents

List of Contributors

Vibhav Bukkapatanam
Department of Management Science and Engineering, Stanford University, Stanford, CA, USA, e-mail: vibhav@stanford.edu

Alexandre J. Chorin
Department of Mathematics, University of California Lawrence, Berkeley National Laboratory, Berkeley, CA, USA,
e-mail: chorin@math.berkeley.edu

Michael Connolly
Federal Reserve Bank of Kansas City, Kansas City, MO, USA,
e-mail: Michael.Connolly@kc.frb.org

Taeyoung Doh
Federal Reserve Bank of Kansas City, Kansas City, MO, USA,
e-mail: Taeyoung.Doh@kc.frb.org

Robert J. Elliott
Haskayne School of Business, University of Calgary, Calgary, AB, Canada
School of Mathematical Sciences, University of Adelaide, Adelaide, SA, Australia
Centre for Applied Financial Studies, University of South Australia, Adelaide, SA, Australia, e-mail: relliott@ucalgary.ca

Jian Hui Huang
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, e-mail: majhuang@inet.polyu.edu.hk

Zhou Jin
Center for Actuarial Studies, Department of Economics, The University of Melbourne, VIC, Australia, e-mail: zjin@unimelb.edu.au

Tze Leung Lai
Department of Statistics, Stanford University, Stanford, CA, USA,
e-mail: lait@stanford.edu

Yoonjung Lee
Boston, MA, USA, e-mail: yoonjung.lee.lin@gmail.com

Hedibert Freitas Lopes
Booth School of Business, University of Chicago, Chicago, IL, USA,
e-mail: hlopes@chicagobooth.edu

Yulei Luo
Department of Economics, The University of Hong Kong, Pokfulam, Hong Kong, China, e-mail: yluo@econ.hku.hk

Jun Ma
Department of Economics, Finance and Legal Studies, University of Alabama, Tuscaloosa, AL, USA, e-mail: jma@cba.ua.edu

Rogemar Mamon
Department of Statistics and Actuarial Sciences, University of Western Ontario, London, ON, Canada, e-mail: rmamon@stats.uwo.ca

Pym Manopimoke
Department of Economics, University of Kansas, Lawrence, KS, USA,
e-mail: pymm@ku.edu

Matthias Morzfeld
Lawrence Berkeley National Laboratory, Berkeley, CA, USA,
e-mail: mmo@math.lbl.gov

Jun Nie
Federal Reserve Bank of Kansas City, Kansas City, MO, USA,
e-mail: jun.nie@kc.frb.org

Maria Paula Rios
Booth School of Business, University of Chicago, Chicago, IL, USA,
e-mail: maria@chicagobooth.edu

Tak Kuen Siu
Cass Business School, City University London, London, UK,
e-mail: Ken.Siu.1@city.ac.uk

Eunju Sohn
Department of Science and Mathematics, Columbia College of Chicago, Chicago, USA, e-mail: esohn@colum.edu

Allanus Tsoi
Department of Mathematics, University of Missouri, Columbia, MO, USA,
e-mail: tsoia@missouri.edu

Xuemin Tu
Department of Mathematics, University of Kansas, Lawrence, KS, USA,
e-mail: xtu@math.ku.edu

Mingming Wang
School of Insurance, The University of International Business and Economics, Beijing, China, e-mail: wiryming@gmail.com

Yazhen Wang
Department of Statistics, University of Wisconsin-Madison, Madison, WI, USA, e-mail: yzwang@stat.wisc.edu

Xiaojing Xi
Department of Applied Mathematics, University of Western Ontario, London, ON, Canada, e-mail: xxi2@uwo.ca

Mark E. Wohar
Department of Economics, University of Nebraska at Omaha, Omaha, NE, USA, e-mail: mwohar@unomaha.edu

George Yin
Department of Mathematics, Wayne State University, Detroit, MI, USA, e-mail: gyin@math.wayne.edu

Eric R. Young
Department of Economics, University of Virginia, Charlottesville, VA, USA, e-mail: ey2d@virginia.edu

Qing Zhang
Department of Mathematics, University of Georgia, Athens, GA, USA, e-mail: qingz@math.uga.edu

Xin Zhang
Department of Statistics, University of Wisconsin-Madison, Madison, WI, USA, e-mail: zhangxin@stat.wisc.edu

Cai Zhu
Department of Finance, Business School, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, e-mail: czhuaa@ust.hk

Part I
Particle Filtering and Parameter Learning in Nonlinear State-Space Models

Chapter 1
Adaptive Filtering, Nonlinear State-Space Models, and Applications in Finance and Econometrics

Tze Leung Lai and Vibhav Bukkapatanam

1.1 Introduction

The Kalman filter, which is applicable to linear Gaussian models, and its modifications such as extended Kalman filters, Gaussian sum filters, and unscented Kalman filters for nonlinear state-space models are widely used in engineering. Without relying on local linearization techniques or functional approximations, particle filters are able to handle a large class of nonlinear non-Gaussian state-space models and have become increasingly popular in engineering applications in the past decade. Filtering in state-space models involves sequential computation of the posterior distribution of the latent state x_t given observations $y_1, \ldots, y_t$. Smoothing involves the estimation of the hidden state x_t given observations $y_1, \ldots, y_n$, with $1 \leq t \leq n$. More details are given in Sect. 1.2.

State-space models typically involve unknown parameters that have to be estimated from the data by either maximum likelihood or Bayesian methods. Replacing these unknown parameters in a particle filter by their sequential estimates leads to an adaptive particle filter; see Liu and West (2001) in [31], Storvik (2002) in [38], Carvalho et al. (2010) in [9], Polson et al. (2008) in [35], and Andrieu et al. (2010) in [2]. Section 1.3 reviews existing methods for sequential parameter estimation in state-space models. Section 1.4 describes a new adaptive filter that combines a novel Markov Chain Monte Carlo (MCMC) scheme for sequential parameter estimation with an efficient particle filter to estimate the state x_t. An important advantage of the new approach is that it yields a consistent estimate of the Monte Carlo standard error.

T.L. Lai (✉)
Department of Statistics, Stanford University, Stanford, CA, USA
e-mail: lait@stanford.edu

V. Bukkapatanam
Department of Management Science and Engineering, Stanford University, Stanford, CA, USA
e-mail: vibhav@stanford.edu

Y. Zeng and S. Wu (eds.), *State-Space Models: Applications in Economics and Finance*, Statistics and Econometrics for Finance 1, DOI 10.1007/978-1-4614-7789-1_1,

Particle filters have powerful and far-reaching applications in state-space models in finance and econometrics, some of which are described in Sect. 1.5. Section 1.6 gives some concluding remarks.

1.2 Particle Filters in Nonlinear State-Space Models

A general state-space model, also called a hidden Markov model (HMM), is defined by the evolution and observation density functions

$$\begin{aligned} x_t|x_{t-1} &\sim f(x_t|x_{t-1},\theta) \\ y_t|x_t &\sim g(y_t|x_t,\theta) \end{aligned} \tag{1.1}$$

with respect to measures μ and ν, respectively, where θ is a vector of parameters of the model. In the Bayesian formulation, the initial state x_0 has prior density $f(x_0|\theta)$, and θ has the prior density $\pi(\theta)$ with respect to some measure on the parameter space. In the HMM, x_t is the latent state and y_t is the observed data at time t. The filtering problem is to sequentially estimate the posterior distribution $p(x_t,\theta|\mathscr{Y}_t)$, where $\mathscr{Y}_t = (y_1, y_2, \ldots, y_t)$ is the set of observations up to time t. Particle filters approximate this posterior density by using a set of particles $(x_t,\theta)^{(i)}$ with weights $\widetilde{w}_t^{(i)}$ $(i = 1,2,\ldots,N)$ summing to 1, so that

$$p^N(x_t,\theta|\mathscr{Y}_t)(\cdot) = \sum_{i=1}^{N} \widetilde{w}_t^{(i)} \delta_{(x_t,\theta)^{(i)}}(\cdot) \tag{1.2}$$

in which δ_z is the Dirac delta function with point mass at z. We now describe different approaches for sampling x_s $(s \le t)$ sequentially to form the particle filter in (1.2) when θ is fixed in advance. The weights $\widetilde{w}_t^{(i)}$ in (1.2) can be converted to $w_t^{(i)} = 1/N$ by a resampling step that samples with replacement N particles from $\{(x_t,\theta)^{(i)} : 1 \le i \le N\}$ with respective weights $\widetilde{w}_t^{(i)}$, as in the original proposal by Gordon et al. (1993) in [23] reviewed below. This is called *bootstrap resampling*. An alternative resampling scheme is introduced in Sect. 1.2.3. Resampling can help to mitigate the degeneracy of particles, which will be discussed in Sects. 1.3 and 1.4. Since resampling is optional and also can be used occasionally, we do not include it in the description of the basic algorithm in Sect. 1.2.1. Moreover, we assume the parameter vector to be known and therefore omit it in the rest of this section.

1.2.1 Bootstrap Filter

Originally proposed by Gordon et al. (1993) in [23], the bootstrap filter chooses samples from the prior distribution of the states. The Bayesian update equation for the posterior in a general filter can be written as

$$p(x_t|\mathscr{Y}_t) = \frac{g(y_t|x_t)p(x_t|\mathscr{Y}_{t-1})}{\int g(y_t|x_t')p(x_t'|\mathscr{Y}_{t-1})d\mu(x_t')} \tag{1.3}$$

where

$$p(x_t|\mathscr{Y}_{t-1}) = \int f(x_t|x_{t-1})p(x_{t-1}|\mathscr{Y}_{t-1})d\mu(x_{t-1}) \tag{1.4}$$

The bootstrap filter samples from the filtering distribution $p(x_t|\mathscr{Y}_{t-1})$ by first propagating the particles which approximate $p(x_{t-1}|\mathscr{Y}_{t-1})$ using the evolution density $f(x_t|x_{t-1})$, and reweighing the particles thus obtained by the likelihood ratio weights, as summarized in Algorithm 1.

Algorithm 1 : Bootstrap Filter

Initialize particles $\{x_0^{(i)}\}_{i=1}^N$ and the corresponding weights $\{\widetilde{w}_0^{(i)} = 1/N\}_{i=1}^N$
for t=1, 2, ..., T **do**
  for i=1, 2, ..., N **do**
    a) Propagate particle $x_{t-1}^{(i)}$ to $x_t^{(i)}$ using the evolution density $f(x_t|x_{t-1}^{(i)})$
    b) Update particle weights according to $\widetilde{w}_t^{(i)} \propto w_{t-1}^{(i)}\, g(y_t|x_t^{(i)})$
  end for
end for

1.2.2 Auxiliary Particle Filter

The auxiliary particle filter proposed by Pitt and Shephard (1999) in [33] generates samples from the filtering distribution with density function

$$p(x_t, x_{t-1}|\mathscr{Y}_t) \propto p(x_t|y_t, x_{t-1})p(y_t|x_{t-1})p(x_{t-1}|\mathscr{Y}_{t-1}) \tag{1.5}$$

Particles approximating $p(x_{t-1}|\mathscr{Y}_{t-1})$ are first resampled using weights proportional to the predictive density $p(y_t|x_{t-1})$, and the resampled particles are propagated forward using $p(x_t|y_t, x_{t-1})$. This is convenient only if $p(x_t|y_t, x_{t-1})$ is not difficult to sample from and $p(y_t|x_{t-1})$ is easily available, which is often not the case. Accordingly Pitt and Shephard in [33] suggest to replace $p(y_t|x_{t-1})$ by $p(y_t|\lambda(x_{t-1}))$, in which $\lambda(x_{t-1})$ is the mean, median, or mode of the distribution of x_t given x_{t-1}, and to propagate the resampled particles by sampling from the proposal density $f(x_t|x_{t-1})$, instead of directly from $p(x_t|x_{t-1}, y_t)$. The method is summarized in Algorithm 2.

Algorithm 2 : Auxiliary Particle Filter

Initialize particles $\{x_0^{(i)}\}_{i=1}^N$ and the corresponding weights $\{w_0^{(i)} = 1/N\}_{i=1}^N$
for t=1, 2, ..., T **do**
  for i=1, 2, ..., N **do**
    a) Resample particle $\widetilde{x}_{t-1}^{(i)}$ from $\{x_{t-1}^{(1)}, \ldots, x_{t-1}^{(N)}\}$ using the weight $\widetilde{w}_t^{(i)} \propto p(y_t|\lambda(x_{t-1}^{(i)}))$
    b) Propagate particle $\widetilde{x}_{t-1}^{(i)}$ to $\widetilde{x}_t^{(i)}$ using $f(\widetilde{x}_t|\widetilde{x}_{t-1}^{(i)})$
    c) Resample $x_t^{(i)}$ from $\{\widetilde{x}_t^{(1)}, \ldots, \widetilde{x}_t^{(N)}\}$ using the weight $w_t^{(i)} \propto g(y_t|\widetilde{x}_t^{(i)})/g(y_t|\lambda(x_{t-1}^{(i)}))$
  end for
end for

1.2.3 Residual Bernoulli Resampling

Bootstrap resampling has been described in the paragraph preceding Sect. 1.2.1. In fact, the name *bootstrap filter* in Sect. 1.2.1 came from bootstrap resampling that Gordon et al. in [23] used to convert weighted particles to particles with equal weights. *Residual Bernoulli resampling* has been proposed as an alternative to bootstrap resampling and has been shown to often lead to smaller variance for the associated particle filter than the bootstrap resampling scheme. The method is summarized in Algorithm 3.

Algorithm 3 : Residual Resampling Scheme

Input: A set of particles $\{(\widetilde{w}_t^{(i)}, \widetilde{\mathscr{X}_t^{(i)}}), i = 1, 2, \ldots, M\}$
Output: A new set of particles $\{(\frac{1}{M}, \mathscr{X}_t^{(i)}), i = 1, 2, \ldots, M\}$
Set $R = \sum_{i=1}^M \lfloor M\widetilde{w}_t^{(i)} \rfloor$
for i=1, 2, ..., M **do**
  – Set $\widehat{w}_t^{(i)} = \frac{M\widetilde{w}_t^{(i)} - \lfloor M\widetilde{w}_t^{(i)} \rfloor}{M-R}$
end for
for j=1, 2, ..., M **do**
  – Sample $\hat{N}_j \sim \texttt{mult}(M-R, \widehat{w}_t^{(1)}, \widehat{w}_t^{(2)}, \ldots, \widehat{w}_t^{(M)})$
  – Set $N_j = \lfloor M\widetilde{w}_n^{(j)} \rfloor + \hat{N}_j$
  – Set $\mathscr{X}_t^{(j)} = \widetilde{\mathscr{X}_t^{(N_j)}}$
end for

1.3 Particle Filters with Sequential Parameter Estimation

An important problem which has been studied extensively in recent filtering literature is that of joint parameter estimation and filtering for general state-space models. Traditional methods which incorporate the parameters as part of the latent state vector suffer from severe degeneracy problems due to the absence of state evolution dynamics for the subvector of latent states representing parameters. Methods to address this issue have been considered in [2, 9, 31, 35, 38], and [34]. They are summarized below.

1.3.1 Liu and West's Filter

Liu and West (2001) in [31] suggest to use a kernel smoothing approximation to the posterior density $p(\theta|\mathscr{Y}_{t-1})$ of the unknown parameter θ via a mixture of multivariate normals and to combine it with an auxiliary particle filter described in Algorithm 2. Let $\{x_{t-1}^{(i)}, \theta_{t-1}^{(i)}\}_{i=1}^{N}$ be a set of particles with weights $w_{t-1}^{(i)}$ $(i = 1,\ldots,N)$, which approximate $p(x_{t-1}, \theta|\mathscr{Y}_{t-1})$. They approximate the posterior density for θ by

$$p(\theta|\mathscr{Y}_{t-1}) = \sum_{j=1}^{N} w_{t-1}^{(j)} N(\theta; m^{(j)}, \Sigma_{t-1})$$

where

$$\begin{aligned} m^{(j)} &= a\theta_{t-1}^{(j)} + (1-a)\bar{\theta} \\ \bar{\theta} &= \sum_{j=1}^{N} \frac{\theta_{t-1}^{(j)}}{N} \\ \Sigma_{t-1} &= (1-a^2) \sum_{j=1}^{N} \frac{(\theta_{t-1}^{(j)} - \bar{\theta})(\theta_{t-1}^{(j)} - \bar{\theta})'}{N} \end{aligned}$$

The constant a measures the extent of shrinkage of the individual $\theta_{t-1}^{(j)}$ to the overall mean $\bar{\theta}$. It is a tuning parameter whose choice is discussed in [31]. The mixture approximation generates new samples from the current posterior and attempts to avoid particle degeneracy. The method is summarized in Algorithm 4.

Algorithm 4 : Liu and West's Filter

Output: The filtering particles $\{(x_t^{(i)}, \theta_t^{(i)})\}_{i=1}^{N}$ and the parameter posterior $p(\theta|\mathscr{Y}_t)$, $t = 1,\ldots,T$
for t=1,2,...,T **do**
  for i=1,2,...,N **do**
    a) Resample $(\tilde{x}_{t-1}, \tilde{\theta}_{t-1})^{(i)}$ from $\{(x_{t-1}^{(j)}, \theta_{t-1}^{(j)})\}_{j=1}^{N}$ with weights $w_t^{(i)} \propto p(y_t|\lambda(x_{t-1}^{(i)}), m^{(i)})$
    b) Propagate $\tilde{\theta}_{t-1}^{(i)}$ to $\hat{\theta}_t^{(i)}$ using $N(\cdot; m^{(i)}, \Sigma_{t-1})$
    c) Propagate $\tilde{x}_{t-1}^{(i)}$ to $\hat{x}_t^{(i)}$ using $p(x_t|\tilde{x}_{t-1}^{(i)}, \tilde{\theta}_t^{(i)})$
    d) Resample $(x_t, \theta_t)^{(i)}$ from $\{(\hat{x}_t, \hat{\theta}_t)^{(j)}\}_{j=1}^{N}$ with weights $w_t^{(i)} \propto p(y_t|\hat{x}_t^{(i)}, \hat{\theta}_t^{(i)}) / p(y_t|\lambda(\tilde{x}_{t-1}^{(i)}), m^{(i)})$
  end for
end for

1.3.2 Storvik's Filter

Storvik (2002) in [38] considered sequential parameter estimation for a class of state-space models in which the posterior parameter density $p(\theta|\mathscr{X}_t,\mathscr{Y}_t)$ can be written as $p(\theta|s_t)$, where s_t is a low-dimensional set of sufficient statistics that can be recursively updated by $s_t = \mathscr{S}(s_{t-1},x_t,y_t)$. The method is summarized in Algorithm 5.

Algorithm 5 : Storvik's Filter

Output: The filtering particles $\{(x_t^{(i)},\theta_t^{(i)})\}_{i=1}^N$ and the parameter posterior $p(\theta|s_t)$, $t=1,\ldots,T$
for t=1,2,...,T **do**
  for i=1,2,...,N **do**
    a) Propagate $x_{t-1}^{(i)}$ to $\tilde{x}_t^{(i)}$ using $q(x_t|\tilde{x}_{t-1}^{(j)},\theta,\mathscr{Y}_t)$
    b) Resample $(x_t,s_{t-1})^{(i)}$ from $\{(\tilde{x}_t,s_{t-1})^{(j)}\}_{j=1}^N$ with weights $w_t^{(i)} \propto \frac{p(y_t|\tilde{x}_{t-1}^{(i)},\theta)p(\tilde{x}_t^{(i)}|x_{t-1}^{(i)},\theta)}{q(\tilde{x}_t^{(i)}|x_{t-1}^{(i)},\theta,\mathscr{Y}_t)}$
    c) Compute sufficient statistics $s_t^{(i)} = \mathscr{S}(s_{t-1}^{(i)},x_t^{(i)},y_t)$
    d) Sample $\theta_t^{(t)}$ from $p(\theta|s_t^{(i)})$
  end for
end for

1.3.3 Particle Learning

Carvalho et al. (2010) in [9] propose the *particle learning* method which utilizes a resample-propagate scheme similar to auxiliary particle filters and show that the proposed method outperforms the filter of Liu and West of [31] in some comparative studies. Assuming the availability of conditional sufficient statistics s_t to represent the posterior of the parameter vector θ, and conditional sufficient statistics $s_{t,x}$ recursive state and parameter updates. The particles are now represented at each time by $z_t^{(i)}(x_t,s_t,s_{t,x},\theta)^{(i)}$ and the Bayesian updating equation can be written as

$$p(z_t|\mathscr{Y}_t) = \int p(s_t|x_t,s_{t-1},y_t)p(x_t|z_{t-1},y_t)p(z_{t-1}|\mathscr{Y}_t)dx_t dz_{t-1} \tag{1.6}$$

where

$$p(z_{t-1}|\mathscr{Y}_t) \propto p(\mathscr{Y}_t|z_{t-1})p(z_{t-1}|\mathscr{Y}_{t-1}) \tag{1.7}$$

Denoting the updating formulas for s_t and $s_{t,x}$ by $s_t = \mathscr{S}(s_{t-1},x_t,y_t)$ and $s_{t,x} = \mathscr{K}(s_{t-1,x},\theta,y_t)$, Algorithm 6 summarizes the resampling and propagation steps in their filter.

Algorithm 6 : Particle Learning

Output: The filtering density approximated by particles $z_t^{(i)}$ and the parameter posterior $p(\theta|s_t)$
for t=1,2,...,T **do**
1) Resample: $\hat{z}_{t-1}^{(i)}$ from the particles $\{z_{t-1}^{(j)}\}_{j=1}^N$ using weights $w_t^{(i)} \propto p(y_t|z_{t-1}^{(i)})$
2) Propagate: $\hat{x}_{t-1}^{(i)}$ to $x_t^{(i)}$ using the distribution $p(x_t|\hat{z}_{t-1}^{(i)}, y_t)$
3) Propagate: Parameter sufficient statistics $s_t^{(i)} = \mathscr{S}(\hat{s}_{t-1}^{(i)}, x_t^{(i)}, y_t)$
4) Propagate: Sample $\theta^{(i)}$ from $p(\theta|s_t^{(i)})$
5) Propagate: State sufficient statistics $s_{t,x}^{(i)} = \mathscr{K}(\hat{s}_{t-1,x}^{(i)}, \theta^{(i)}, y_t)$
end for

1.3.4 Particle MCMC

Hybrid methods that combine particle filters with MCMC schemes have been considered in the literature. Important recent developments in this direction are [2, 35], and [34]. Polson et al. (2008) in [35] use a rolling window MCMC algorithm that approximates the target posterior distribution by a mixture of lag-k smoothing distributions. They recast the filtering problem as a sequence of smaller smoothing problems which can be solved using standard MCMC approaches as in [7] and [8]. They exploit, whenever possible, a sufficient statistic structure as in [19] and [38] to perform parameter updates and develop an algorithm with linear computational cost. Andrieu et al. (2010) in [2] introduce the particle MCMC (PMCMC) methods to perform inference on the unknown parameter vector θ. Pitt et al. (2012) in [34] provide further analytic results on PMCMC and show that using auxiliary particle filters in PMCMC schemes may help reduce computation time.

1.4 A New Approach to Adaptive Particle Filtering

In this section, we describe a new adaptive filtering technique, recently introduced in [11, 12], for joint parameter and latent state filtering in particle filters, which provides substantial improvement over previous approaches. The authors in [12] propose an efficient MCMC method to estimate the posterior distribution of the parameters, which can be used in conjunction with traditional particle filter methods that assume the parameters to be known. Bukkapatanam et al. (2012) in [5] provide further development of the methodology and its applications to economics and finance, which will be summarized in Sect. 1.5.

Chan and Lai (2012a) in [11] begin by considering the case where the parameter vector θ is known so that it can be omitted from the notation for particle filters, as in Sects. 1.2.1–1.2.3. They consider more generally the estimation of $\psi_T = \mathbb{E}[\psi(\mathscr{X}_T)|\mathscr{Y}_T]$ instead of $\mathbb{E}[\psi(x_T)|\mathscr{Y}_T]$, where $\mathscr{X}_T = (x_1, \ldots, x_T)$. When bootstrap resampling is performed at every stage, they show that the bootstrap filter estimate $\widehat{\psi}_T$ of $\psi_T = \mathbb{E}[\psi(\mathscr{X}_T)|\mathscr{Y}_T]$ has a martingale representation

$$m(\widehat{\psi}_T - \psi_T) = \sum_{j=1}^{m} (\varepsilon_1^j + \cdots + \varepsilon_{2T-1}^j) + O_p(1)$$

where $\{\varepsilon_k^j,\ 1 \le k \le 2T-1\}$ is a martingale difference sequence for $j = 1, \ldots, M$. They use this to derive the central limit theorem

$$\sqrt{M}(\widehat{\psi}_T - \psi_T) \implies N(0, \sigma^2) \quad \text{as } M \to \infty$$

for the particle filter and a consistent estimate

$$\widehat{\sigma}^2 = \frac{1}{M} \sum_{j=1}^{M} \Big(\sum_{i: A_{T-1}^i = j} \frac{\widetilde{w}_T^{(i)}}{\bar{w}_T} [\psi(\mathscr{X}_T^i) - \widehat{\psi}_T] \Big)^2$$

of σ^2, where $\bar{w}_t = M^{-1} \sum_{j=1}^{M} \widetilde{w}_t^{(j)}$, and A_t^i denotes the ancestral origin of the particle $\mathscr{X}_t^i$, thus $A_t^i = j$ if the first component of $\mathscr{X}_t^i$ is x_1^j, recalling that the first generation of the M particles consists of $x_1^1, \ldots, x_1^m$ before resampling. Chan and Lai (2012a) in [11] also extend these results to the case where occasional resampling is used. Conditional on $\mathscr{Y}_T$, the mean squared error of estimating ψ_T by the particle filter $\widehat{\psi}_T$ is

$$\begin{aligned} E[\{\psi(\mathscr{X}_T) - \widehat{\psi}_T\}^2 | \mathscr{Y}_T] \\ = \underbrace{\mathbb{E}[\{\psi(\mathscr{X}_T) - \psi_T\}^2 | \mathscr{Y}_T]}_{(I)} + \underbrace{\mathbb{E}[\{\psi_T - \widehat{\psi}_T\}^2 | \mathscr{Y}_T]}_{(II)} \end{aligned} \tag{1.8}$$

The preceding discussion shows how (II) can be consistently estimated. To estimate (I), write it as $\mathbb{E}[\psi^2(\mathscr{X}_T)|\mathscr{Y}_T] - \psi_T^2$. The first term can be estimated by particle filters (with ψ^2 replacing ψ) and the second term is estimated by $\widehat{\psi}_T^2$.

1.4.1 A New MCMC Approach to Sequential Parameter Estimation

As described in the previous section, several hybrid schemes that combine particle filters and MCMC schemes have been proposed in literature. A shortcoming of most of the existing methods is the prohibitively lengthy computational time, which makes it very difficult to carry out simulation studies of their performance. To significantly reduce the computational burden of these hybrid schemes, Chan and Lai (2012b) in [12] use the following *state substitution* method to carry out MCMC iterations.

In what follows, we refer to the posterior density $p(\theta|\mathscr{Y}_t)$ of a parameter vector θ given the observed data as the *target density*. Since $\mathscr{Y}_t$ is observed, we shall treat it as a constant and simply denote $p(\theta|\mathscr{Y}_t)$ by $p(\theta)$. This distribution is approximated by a sequentially selected set of n atoms and the weights associated with the atoms. The MCMC scheme chooses these representative atoms by a Markov updating procedure

Algorithm 7 : Efficient MCMC Algorithm involving State Substitutions

Output: Atoms $S_K = \{\theta_K^1, \ldots, \theta_K^n\}$ approximating the target density
for k=1,2,...,K **do**
Sample θ_{k-1}^{n+1} from $q(\cdot, S_{k-1})$
Let $S_{k-1,i} = (S_{k-1} \cup \{\theta_{k-1}^{n+1}\}) \setminus \{\theta_{k-1}^i\} \quad i = 1, 2, \ldots, n+1$
Generate J_k with the following distribution

$$P(J_k = i) = \frac{\lambda_{k.i}}{\sum_{j=1}^{n+1} \lambda_{k.j}} \tag{1.9}$$

where

$$\lambda_{k.i} = \frac{q(\theta_{k-1}^i; S_{k-1,i})}{p(\theta_{k-1}^i)} \tag{1.10}$$

if $J_k = n+1$ **then**

$$\theta_k^i = \theta_{k-1}^i \quad \forall\, i = 1, \ldots, n$$

else

$$\theta_k^i = \begin{cases} \theta_{k-1}^i \text{ for } i = 1, \ldots, n \text{ if } i \neq J_k \\ \theta_{k-1}^{n+1} \text{ if } i = J_k \end{cases} \tag{1.11}$$

end if
end for

involving state substitutions so that the associated distribution of the weights on the atoms converges weakly to the target distribution. Let $\Theta = \{\theta : p(\theta) > 0\}$ be an open subset of $\mathbb{R}^d$. Let $q(\cdot; S)$ be a proposal distribution whose form depends on a given set of S of parameters. Let $S_0 = \{\theta_0^1, \ldots, \theta_0^n\}$ be an initial set of n parameter values. The method is summarized in Algorithm 7. Note that since (1.10) is a quotient, the target density only needs to be specified up to a normalizing constant. The substitution idea represented by (1.11) attempts to use θ_{k-1}^{n+1} that is newly generated from the proposal distribution to substitute $\theta_{k-1}^{J_k}$ that tends to have a larger likelihood ratio of the proposal density to the target density. The estimate of $\mathbb{E}(\psi(\theta)|\mathscr{Y}_t)$ is a weighted average of the form

$$\widehat{\psi}^{(k)} = \frac{\sum_{i=1}^n \sum_{k=1}^K \tilde{\lambda}_k^{-1} \psi(\theta_{k-1}^i)}{n \sum_{k=1}^K \tilde{\lambda}_k^{-1}} \tag{1.12}$$

where

$$\tilde{\lambda}_K = \left(\frac{1}{n} \sum_{i=1}^n \lambda_K^i \right) \vee \lambda_* \tag{1.13}$$

and λ_* is a positive number to ensure that the weights $\tilde{\lambda}_k^{-1}$ are not too large.

1.4.2 Adaptive Particle Filters and Asymptotic Theory

Using the MCMC procedure of Sect. 1.4.1 for parameter estimation, Chan and Lai (2012b) in [12] introduce an adaptive particle filter algorithm that is described in this section. Consider the state-space model

$$X_t \sim f_\theta(\cdot|X_{t-1})$$
$$Y_t \sim g_\theta(\cdot|X_t)$$

where $f_\theta(\cdot|X_0) = f_\theta(\cdot)$, and the prior distribution of the parameter vector θ has density function π. We initialize the adaptive filter by sampling $\theta_0^1, \ldots, \theta_0^n$ independently from π. Let $S_0^0 = \{\theta_0^1, \ldots, \theta_0^n\}$, $\hat{p}_0(\theta) = \pi(\theta)$, $w_{0,m}^i = 1$ for $i = 1, \ldots, n$, and $m = 1, \ldots, M$ and where M denotes the number of particles in the particle filter. The parameter and state update steps are described in Algorithm 8, in which K denotes the number of MCMC iterations and T is the number of time steps. The adaptive particle filter estimate $\mathbb{E}[\psi_t(\theta, \mathscr{X}_t)|\mathscr{Y}_t]$ is given by

$$\widehat{\psi}_t = \frac{1}{n}\sum_{i=1}^{n} \widehat{\psi}_t^i \tag{1.14}$$

with

$$\widehat{\psi}_t^i = \frac{\sum_{k=1}^{K} \widetilde{\lambda}_k^{-1} \left(\sum_{m=1}^{M} \nu_{t,m}^i \psi_t(\theta_k^i, \mathscr{X}_{t,m}^i)\right)}{\sum_{k=1}^{K} \widetilde{\lambda}_k^{-1}} \tag{1.15}$$

where using the notation in Algorithm 8,

$$\nu_{t,m}^i = \frac{\widetilde{w}_{t,m}^i}{\sum_{l=1}^{M} \widetilde{w}_{t,l}^i} \tag{1.16}$$

We now focus on the case $\psi_T = \mathbb{E}[\psi(x_T, \theta)|\mathscr{Y}_T]$, for ψ involving only x_T instead of the entire trajectory $\mathscr{X}_T$. For large T, the adaptive filter is computationally intensive and the MCMC algorithm may also take a long time to converge. On the other hand, under certain regularity conditions, the MLE converges to θ as $T \to \infty$. We propose to break the data into smaller batches of size T_0 and carry out the preceding algorithm for each batch, initializing with the prior density for the first batch and using the kernel density, e.g. $b^{-d}\sum_{i=1}^{n} \phi((\theta - \theta_K^i)/b)$, instead of the atoms to approximate the posterior for the subsequent batches, where ϕ denotes the $N(0, I_d)$ density, $d = dim(\theta)$, and $b = b_K$ is the bandwidth such that $b_K \to 0$ as $K \to \infty$. After a certain number κ of batches (such that κT_0 is large enough; the covariance matrix of the atoms in S_K of the MCMC algorithm gives a good diagnostic check on whether that is the case), we can estimate θ by recursive maximum likelihood (gradient-type EM involving particle filters for the E-step) and replace the unknown θ in the particle filter that assumes θ to be known by the recursive estimate $\widehat{\theta}_t$ at stage t.

Algorithm 8 : Adaptive Filter

Input: $S_0^0, \widehat{p}_0, w_{0,n}^i$
Output: Atoms $S_K = \{\theta_K^1, \ldots, \theta_K^n\}$ approximating the target density and the filtered states
for $t = 1, \ldots, T$ **do**
  for $i = 1, \ldots, n$ **do**
    for $m = 1, \ldots, M$ **do**
      Sample the new particles $\widetilde{X}_{t,m}^i$ as below and let $\widetilde{\mathscr{X}}_{t,m}^i = (\mathscr{X}_{t-1,m}^i, \widetilde{X}_{t,m}^i)$

$$\widetilde{X}_{t,m}^i \sim f_{\theta_{k-1}^i}(\cdot, X_{t-1,m}^i)$$

      Update the weights $\widetilde{w}_{t,m}^i = w_{t-1,m}^i \times g_{\theta_{k-1}^i}(Y_t | \widetilde{X}_{t,m}^i)$
    end for
    Compute the Effective Sample Size $ESS_t^i = \frac{(\sum_{m=1}^M \widetilde{w}_{t,m}^i)^2}{\sum_{m=1}^M (\widetilde{w}_{t,m}^i)^2}$
    if $ESS_t^i < c$ and $t < T$ **then**
      Resample $\widetilde{\mathscr{X}}_{t,m}^i$ using bootstrap resampling with weights $\widetilde{w}_{t,m}^i$ to get $\mathscr{X}_{t,m}^i$
    else
      Set $\mathscr{X}_{t,m}^i = \widetilde{\mathscr{X}}_{t,m}^i$ and $w_{t,m}^i = \widetilde{w}_{t,m}^i$
    end if
    Update the target density $\widehat{p}_t(\theta_{k-1}^i) = \widehat{p}_{\tau_t^i}(\theta_{k-1}^i) \frac{\sum_{m=1}^M \widetilde{w}_{t,m}^i}{M}$ where τ_t^i is the most recent resampling time before time t for atom θ_{k-1}^i. $\tau_t^i = 0$ if no resampling has occurred
  end for
  for $k = 1, \ldots, K$ **do**
    Generate a new atom θ_{k-1}^{n+1} from $q(\cdot; \bar{\gamma}_k)$, where $\bar{\gamma}_k = n^{-1} \sum_{i=1}^n \gamma(\theta_{k-1}^i)$ for some $\gamma : \Theta \to \Gamma$ with $\Gamma \subset \mathbb{R}^{d_\Gamma}$
    for u=1,. . . ,t **do**
      Perform particle filter update steps with atom set to θ_{k-1}^{n+1}
    end for
    Update the target density $\widehat{p}_t(\theta_{k-1}^{n+1}) = \widehat{p}_{\tau_t^i}(\theta_{k-1}^{n+1}) \frac{\sum_{m=1}^M w_{t,m}^i}{M}$ using the weights of particles generated for this atom
    Substitute θ_{k-1}^{n+1} into some θ_{k-1}^j using the MCMC Algorithm (7)
  end for
  Get the set of atoms for the next particle update step S_k
end for

Bukkapatanam et al. (2012) in [5] have developed an asymptotic theory for the adaptive filter described in Algorithm 8 using the results of [12] on the asymptotic bias and asymptotic variance of the MCMC scheme in Sect. 1.4.1 as $n \to \infty$ and $K \to \infty$, in the case of fixed T. For large T, they use the batch idea in the preceding paragraph. The theory of stochastic approximation can be used to show that the recursive maximum likelihood estimate $\widehat{\theta}_T$ is asymptotically normal as $T \to \infty$. Analogous to (1.8), the asymptotic variance of the adaptive particle filter $\widehat{\psi}$ can be decomposed as the sum

$$V_T + (1 + o_p(1))(\sigma^2/T + \sigma_\theta^2/M) \tag{1.17}$$

where σ_θ^2/M is the asymptotic Monte Carlo variance of the particle filter (with M particles) when θ is known, σ^2/T is the additional variance due to replacing θ in the particle filter by $\widehat{\theta}_T$, and

$$V_T = \mathbb{E}\left[\{\psi(x_T, \theta) - \mathbb{E}(\psi(x_T, \theta)|\mathscr{Y}_T)\}^2|\mathscr{Y}_T\right] \tag{1.18}$$

1.5 Applications in Finance and Economics

We begin this section with an application of the adaptive particle filter in Sect. 1.4.2 to filtering in dynamic frailty models, introduced by [16] in the wake of the Financial Crisis of 2007–2008, for corporate defaults. We then describe a variety of other applications of nonlinear state-space models and adaptive particle filters in the recent finance and economics literature.

1.5.1 Frailty Models for Corporate Defaults

In an influential study of default probabilities in corporate debt portfolios, Duffie et al. (2009) in [15] find that conventional estimators of portfolio default probabilities are downward biased and show that firms are exposed to a common dynamic latent factor driving default even after controlling for observable factors, which are used by conventional estimators to compute the firm-by-firm default probabilities. This latent variable, which in essence captures the other factors that the modeler may have failed to consider, is shown to cause a substantial increase in the conditional probability of large portfolio default losses. An example of one such factor which was not included in many of the mortgage portfolio default loss models, and thereby contributed to a large loss in US mortgage portfolios (>$800 billion), is the degree to which borrowers and mortgage brokers provided proper documentation of borrower's credit qualities. Duffie et al. (2009) in [15] model this unobserved covariate, called *frailty*, as a dynamic process and use the stochastic EM algorithm proposed in [42] to perform maximum likelihood estimation of the model parameters, and MCMC methods to estimate the conditional distribution of the latent frailty process.

Specifically, they model the intensity $\lambda_i(t)$ of default of firm i at time t by

$$\lambda_i(t) = \exp(\eta_0 + \eta_1 \cdot V_t + \eta_2 \cdot U_{it} + \eta F_t) \tag{1.19}$$

where $\cdot$ denotes the inner product of two vectors, U_{it} is a firm-specific observable covariate vector at time t, V_t consists of the observable macroeconomic covariates at time t, F_t is a dynamic frailty covariate that is not observable, and $\eta_0, \eta_1, \eta_2,$ and η are unknown parameters. The latent frailty F_t is assumed to follow an

Ornstein–Uhlenbeck (OU) process

$$dF_t = \kappa(\mu - F_t)dt + dB_t, \quad F_0 = 0 \tag{1.20}$$

where B_t is a standard Brownian motion, $\kappa > 0$ is the mean reversion parameter, and μ is the steady-state mean of the OU process. The observations at time t are $Y_t = \{(T_i \wedge (t - e_i)^+, \delta_{it}, U_{it}, V_t)\}_{i=1}^I$, where $T_i = \tau_i \wedge c_i$, with τ_i being the default time of firm i (measured from the firm's entry time e_i into the study), c_i being the censoring variable caused by the firm's exit from the study due to merger, acquisition, or other failure, and δ_{it} is the default indicator (taking the value 0 or 1) so that $\delta_{it} = 1$ if $T_i \wedge t = \tau_i$. It is assumed that τ_i and c_i are independent.

The MCMC methods used in [15] to compute the posterior distribution of the latent frailty process F_t involve both Gibbs sampling and random walk Metropolis–Hastings steps [32]. Together with the stochastic EM algorithm used to estimate the unknown parameters $\eta_0, \eta_1, \eta_2, \eta, \kappa,$ and μ, the procedure is very computationally intensive. We use the adaptive particle filter in Sect. 1.4.2 instead in the following simulation study. The adaptive particle filter involves $n = 1{,}000$ atoms, $M = 5{,}000$ particles, and $K = 1{,}000$ MCMC iterations. Because the adaptive filter is much faster than the procedure of [15], we can indeed carry out the simulation study with 100 simulations with relative ease.

The simulation study considers $I = 500$ firms, with $e_1 = \cdots = e_I = 0$, and a time period of $T = 30$ years. The parameters of the OU process (1.20) are $\kappa = 0.125$ and $\mu = 1$. We assume a scalar firm-specific covariate U_{it} and a single macroeconomic variable V_t. Following [17], we assume that these covariates are observed only on a monthly basis and are generated by independent AR(1) processes

$$\begin{aligned} V_t &= 0.9V_{t-1} + 0.6 + 1.8\varepsilon_t \\ U_{it} &= U_{it-1} + 0.04(\mu_i - U_{it-1}) + 0.3\xi_{it} \end{aligned} \tag{1.21}$$

with $V_1 \sim N(6, 1.8^2)$, $\mu_i \sim N(2, 0.5^2)$, $\varepsilon_t \sim N(0,1)$, $\xi_{it} \sim N(0,1)$, and $U_{i1} \sim N(\mu_i, 0.3^2)$. The parameters in the default intensity (1.19) are $(\eta_0, \eta_1, \eta_2, \eta) = (-2, -1, -0.3, 0.5)$. In each of the 100 simulation runs, we generate F_t by (1.20), (U_{it}, V_T) by (1.21), and use the default intensity (1.19) to generate the firm's default times using the *thinning* algorithm for nonhomogeneous Poisson processes [37]. For simplicity we only assume "administrative censoring" by fixed time t in $T_i \wedge t$ and no additional stochastic censoring variables c_i. We assume the prior distribution π of the parameter vector $\theta = (\eta_0, \eta_1, \eta_2, \eta, \kappa, \mu)$ to be $N(\mu_0, \Sigma_0)$, where $\mu_0 = (-3, -1.5, -0.5, 0.8, 0.15, 1.3)$ and $\Sigma_0 = diag(1, 0.8^2, 0.4^2, 0.5^2, 0.1^2, 0.6^2)$. This yields $\widehat{p}_0 = \pi$ in Algorithm 8. Note that the observation density in this HMM is

$$g_\theta(Y_t \mid F_t) = \prod_{i=1}^{I} (\lambda_i(T_i\,; \theta))^{\delta_{it}} \exp(-\Lambda_i(T_i\,; \theta)) \tag{1.22}$$

where $\Lambda_i(t_i\,; \theta) = \int_0^t \lambda_i(s\,; \theta) ds$ is the cumulative hazard function that can be evaluated by numerical quadrature; we use $\lambda_i(t\,; \theta)$ to highlight the dependence on

the parameter vector in (1.19) and (1.20). For the particle filters in Algorithm 8, we perform bootstrap resampling only when the effective sample size $ESS < 50$. Figure 1.1 compares the actual frailty path F_t (at times $t = 1,2,\ldots,30$), from one of the simulated samples with the filtered frailty paths $\mathbb{E}(F_t \mid \mathscr{Y}_t)$ estimated by two different particle filters: (1) the adaptive particle filter implemented by Algorithm 8, and (2) the bootstrap filter that assumes the parameter vector θ to be known. Figure 1.1 shows that the adaptive particle filter performs on par with the "oracle" bootstrap filter and is able to retrieve the latent frailty estimate.

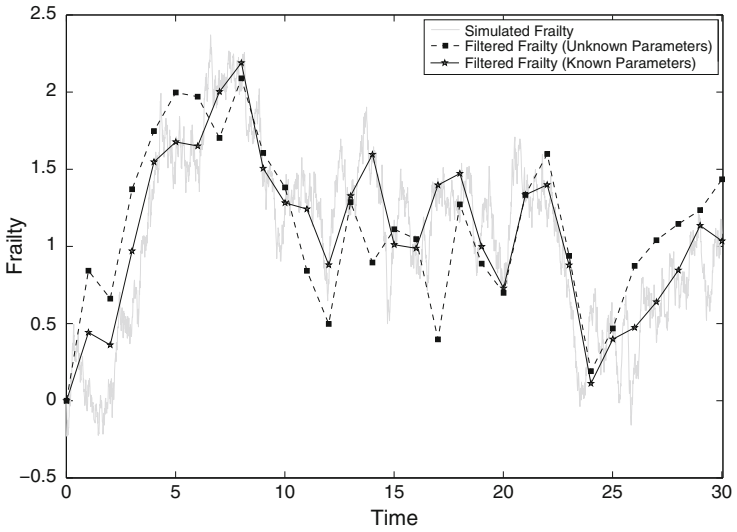

Fig. 1.1 Filtered frailty paths of the adaptive and the bootstrap particle filters

Our simulation study also shows that the adaptive particle filter estimates the parameter vector θ quite well. The following table gives the means and standard deviations of the parameter estimates over the 100 simulations. The parameter estimates $\widehat{\theta}$ are given by the weighted averages of the atoms at the termination of the MCMC iterations in Algorithm 8 that also gives the weights through the target density.

We can also use the adaptive particle filter to predict the default probabilities for the subsequent time periods. In particular, our simulation study considers 1-year ahead default probability forecasts of the adaptive particle filter and compares them to the actual probabilities. We first apply Algorithm 8 to an expanding window of training data from time $t = 1$ up to time $t = 15,16,\ldots,29$, respectively. Using the estimated parameters from the adaptive particle filter at time t, we simulate 10,000 OU paths for the latent frailty, corresponding to the period $[t,t+1)$, for $t = 15,\ldots,29$. We then estimate the probability of default $\hat{\pi}_{i,t+1}$ of each firm i in the time period $[t,t+1)$ and compare it with the true value $\pi_{i,t+1}$ obtained from the thinning algorithm. For a single firm in one of the simulated data sets, we plot the 1-year ahead

	η_0	η_1	η_2	μ	κ	η
θ	-2	-1	-0.3	1	0.125	0.5
$\mathbb{E}(\widehat{\theta}_{15})$	-2.583	-1.837	-0.386	0.997	0.123	0.485
$SD(\widehat{\theta}_{15})$	0.883	0.445	0.105	0.104	0.005	0.031
$\mathbb{E}(\widehat{\theta}_{20})$	-2.435	-1.859	-0.389	0.985	0.127	0.479
$SD(\widehat{\theta}_{20})$	0.634	0.745	0.127	0.129	0.009	0.052
$\mathbb{E}(\widehat{\theta}_{25})$	-2.396	-1.635	-0.368	1.028	0.124	0.493
$SD(\widehat{\theta}_{25})$	0.608	0.593	0.095	0.073	0.004	0.046
$\mathbb{E}(\widehat{\theta}_{30})$	-2.353	-1.467	-0.346	1.013	0.126	0.487
$SD(\widehat{\theta}_{30})$	0.581	0.623	0.086	0.055	0.006	0.077

probability forecast of the adaptive particle filter, and the true default probability over time $t = 15, 16, \ldots, 29$ in Fig. 1.2. The figure shows that the adaptive particle filter produces good estimates for the 1-year ahead default probability.

1.5.2 Stochastic Volatility with Contemporaneous Jumps

Stochastic volatility models have been studied extensively in the recent literature. There is strong consensus about the need to include discontinuities in asset prices via return jumps, while incorporating stochastic time variation in the volatility of the continuous shocks to returns, via stochastic volatility. While these features yield fatter tails in the return distribution, they do not completely explain the rapid increases in volatility experienced in history.

Duffie et al. (2000) in [16] propose adding jumps to both the returns and the stochastic volatility which, as noted in [18], serves two different but complementary purposes. Jumps in returns generate large, sudden movements in asset prices which are infrequently observed. On the other hand, jumps in volatility lead to fast changes in the level of volatility, and hence the distribution of asset prices, due to volatility persistence. An issue which has to be addressed in this case is that of jumps in returns and volatility occurring contemporaneously. Duffie et al. (2000) in [16] propose contemporaneous arrivals with correlated jump sizes. Jacod and Todorov (2009) in [26] and Todorov and Tauchen in [39, 40] find evidence of a specification with a high likelihood of contemporaneous jump arrivals using intraday observations on the VIX and the S&P 500 index.

A discrete-time version of the stochastic volatility model with contemporaneous jumps (SVCJ) in both the returns and the volatility can be written as

$$\begin{pmatrix} y_t \\ V_t \end{pmatrix} = \begin{pmatrix} \mu \\ \kappa\theta + (1-\kappa)V_{t-1} \end{pmatrix} + \sqrt{V_{t-1}} \begin{pmatrix} \varepsilon_t^Y \\ \sigma_V \varepsilon_t^V \end{pmatrix} + \begin{pmatrix} \xi_t^Y \\ \xi_t^V \end{pmatrix} J_t \tag{1.23}$$

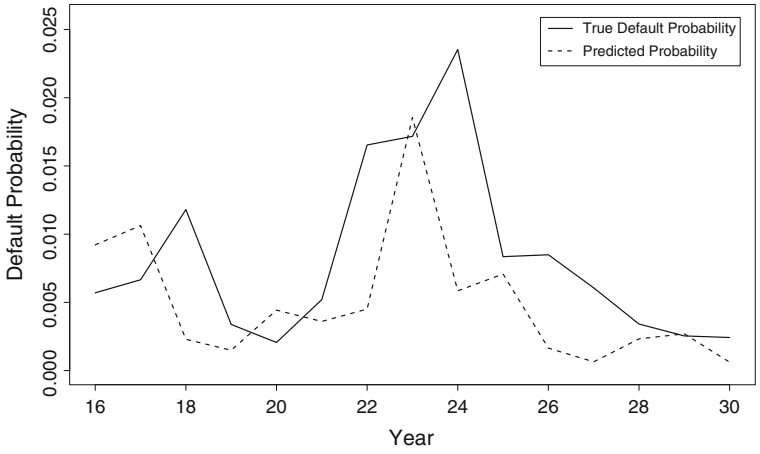

Fig. 1.2 One-year ahead predictions

where

$$\begin{pmatrix} \varepsilon_t^Y \\ \varepsilon_t^V \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right) \tag{1.24}$$

and

$$J_t \sim \text{Bernoulli}(\lambda) \quad \xi_t^Y \sim N(\mu_Y, \sigma_Y^2) \quad \xi_t^V \sim \text{Exp}(\mu_V) \tag{1.25}$$

In this specification, y_t is the observed asset log-return at time t, V_t is the latent stochastic volatility, and the jumps in returns and volatility are contemporaneous without any correlation in the jump sizes. The complete specification of the latent state x_t includes not just only the unknown V_{t-1} but also any random variable generated in the process of obtaining the V that may have an influence on the distribution of y_t, in this case J_t and ε_t^V. We thus have the latent state vector $x_t = (V_{t-1}, J_t, \varepsilon_t^V)$. The evolution density of the state-space model is given by (1.23)–(1.25), and the observation density is given by

$$\varepsilon_t^Y | \varepsilon_t^V \sim N(\rho \varepsilon_t^V, 1 - \rho^2) \tag{1.26}$$

and

$$y_t | x_t \sim \begin{cases} N(\mu + \sqrt{V_{t-1}}\rho\varepsilon_t^V + \mu_Y, V_{t-1}(1-\rho^2) + \sigma_y^2) & \text{if } J_t = 1 \\ N(\mu + \sqrt{V_{t-1}}\rho\varepsilon_t^V, V_{t-1}(1-\rho^2)) & \text{if } J_t = 0 \end{cases} \tag{1.27}$$

Parameter estimation and sequential state filtering in this model can be performed using the adaptive particle filter as given in Sect. 1.4.2, which is considerably more efficient than the MCMC methods of [18] and [28].

1.5.3 State-Space Models for High-Frequency Transaction Data

In the last decade, there has been an explosion of computer-based trading systems which place quotes, make trading decisions, and manage existing orders after submission using automated computer algorithms. In the US equities market, Brogaard (2010) in [4] and Securities & Exchange Commission in [41] estimate that 60–75 % of the daily trading volume is due to the so-called high frequency traders (HFTs). In addition to the complete automation of major financial exchanges, there has been a proliferation of electronic communication networks and dark pools of liquidity in the financial markets. With quote and trade update frequencies in the order of a few milliseconds, an important issue facing an HFT is the online estimation of market microstructure parameters. These estimates could feed into adaptive market making algorithms that dynamically adjust the bid/ask prices and sizes quoted by the market maker at the various levels of the order book, while adjusting for market liquidity and existing inventory. Much of the literature on estimation of market microstructure parameters has focused on nonparametric statistical methods. Notable exceptions are Zeng (2003) and (2004) in [43, 44] and Hu et al. (2010) in [25]. They use a nonlinear filtering framework with marked point process observations for high-frequency transaction data and derive filtering equations to characterize the evolution of likelihoods, Bayes factors, and posterior probabilities. Most of the existing models, however, do not consider limit order book dynamics and other features that are of vital importance in implementing a market making system which needs to adapt to changes in a large number of securities in real time. In the financial econometrics literature on market microstructure [1, 3, 24, 45], it is customary to model the latent efficient price process as being observed in the presence of additive microstructure noise also referred to as "market frictions." These models assume that the frictions are static throughout the day and do not take into consideration the discreteness of observed prices and the dynamic nature of market frictions. In addition, the assumption of additive microstructure noise severely restricts the scope of the microstructure model.

Bukkapatanam and Lai (2012) in [6] have recently introduced a general nonlinear market microstructure model that overcomes these limitations. Using the methodology in Sect. 1.4, they have developed an efficient adaptive particle filter and used it to update the spot and cross volatility estimates at each new transaction. Their microstructure model incorporates price discreteness and market information from the live limit order book. The ability to adjust in real time to changes in the limit order book is an important feature of their method and allows it to be applied to practical market making strategies. The filtering approach to estimating cross volatility is compatible with asynchronous transaction data. Their simulation studies show the effectiveness of their approach in estimating the underlying spot and cross volatilities of the price process. They also apply the adaptive particle filter to NYMEX and CME futures exchange data and demonstrate its superior real time performance.

1.5.4 Other Applications in Finance and Economics

Carvalho and Lopes (2007) in [10] analyze a Markov switching stochastic volatility model using the particle filter of [31]. Several Lévy-type stochastic volatility models are studied in [27, 30, 36] and [29]. DaSilva et al. (2009) in [14] model the monthly Brazilian unemployment rate from March 2002 to December 2009 using a dynamic beta regression model analyzed using MCMC. This is a special case of a dynamic generalized linear model, where the observational distributions belong to an exponential family. Flury and Shephard (2011) in [22] apply the particle MCMC framework to problems arising in macroeconomics and finance. Creal (2012) in [13] surveys sequential Monte Carlo methods in economics and finance. He points out that building on the foundation of nonlinear microeconomic models for learning and strategic interaction amount agents, "macroeconomists formulate their structural models as dynamic stochastic general equilibrium (DSGE) models," and that the recent popularity of particle filters in the economics literature began with their applications to estimation in DSGE models by Fernandez-Villaverde and Rubio-Ramrez (2005, 2007) in [20, 21].

1.6 Conclusion

Sequential Monte Carlo methods, also known as particle filters, have far-reaching and powerful applications in modern time series analysis problems involving state-space models. Without relying on local linearization techniques or functional approximations, particle filters are able to handle a large class of nonlinear non-Gaussian state-space models and have become increasingly popular in engineering and econometric applications in the past decade. We provide a brief overview of particle filters for state-space models and their applications in finance and economics. We also introduce a new adaptive particle filter that uses a novel Markov Chain Monte Carlo scheme to approximate the posterior distribution of the model parameters.

References

1. Aït-Sahalia, Y.; Mykland, P. A., and Zhang, L. How often to sample a continuous-time process in the presence of market microstructure noise. Review of Financial Studies, **18**, 351–416 (2005)
2. Andrieu, C., Doucet, A., Holenstein, R.: Particle Markov Chain Monte Carlo methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **72**(3), 269–342 (2010)
3. Barndorff-Nielsen, O., Hansen, P., Lunde, A., Shephard, N.: Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise. Econometrica **76**, 1481–1536 (2008)
4. Brogaard, J.: High frequency trading and its impact on market quality. Working Paper, Kellogg School of Management, Northwestern University (2010)

5. Bukkapatanam, V., Chan, H., Lai, T.: Adaptive particle filters in nonlinear state-space models and their econometric applications. Working Paper, Department of Statistics, Stanford University (2012)
6. Bukkapatanam, V., Lai, T.: State-space models of dynamic market microstructure and estimation with high-frequency transactions data. Working Paper, Department of Statistics, Stanford University (2012)
7. Carlin, B.P., Polson, N.G., Stoffer, D.: A Monte Carlo approach to nonnormal and nonlinear state-space modeling. Journal of the American Statistical Association **87**, 493–500 (1992)
8. Carter, C.K., Kohn, R.: On Gibbs sampling for state space models. Biometrika **81**, 541–553 (1994)
9. Carvalho, C., Johannes, M., Lopes, H., Polson, N.: Particle learning and smoothing. Statistical Science pp. 88–106 (2010)
10. Carvalho, C., Lopes, H.: Simulation-based sequential analysis of Markov switching stochastic volatility models. Computational Statistics and Data Analysis **51**, 4526–4542 (2007)
11. Chan, H., Lai, T.: A general theory of particle filters in hidden Markov models and some applications. Annals of Statistics (2013). To appear
12. Chan, H., Lai, T.: A new approach to Markov Chain Monte Carlo with applications to adaptive particle filters. Working Paper, Department of Statistics, Stanford University (2012)
13. Creal, D.: A survey of sequential Monte Carlo methods for economics and finance. Econometric Reviews **31**(3), 245–296 (2012)
14. DaSilva, C., Migon, H., Correira, L.: Bayesian beta dynamic model and applications. Working paper, Department of Statistics, Federal University of Rio de Janeiro (2009)
15. Duffie, D., Eckner, A., Horel, G., Saita, L.: Frailty correlated default. Journal of Finance **64**(5), 2089–2123 (2009)
16. Duffie, D., Pan, J., Singleton, K.: Transform analysis and asset pricing for affine jump diffusions. Econometrica **68**, 1343–1376 (2000)
17. Duffie, D., Saita, L., Wang, K.: Multi-period corporate default prediction with stochastic covariates. Journal of Financial Economics **83**, 635–665 (2007)
18. Eraker, B., Johannes, M., Polson, N.: The impact of jumps in volatility and returns. Journal of Finance **58**, 1269–1300 (2003)
19. Fearnhead, P.: Markov Chain Monte Carlo, sufficient statistics, and particle filters. Journal of Computational and Graphical Statistics **11**(4), 848–862 (2002)
20. Fernández-Villaverde, J., Rubio-Ramírez, J.: Estimating dynamic equilibrium economies: linear versus nonlinear likelihood. Journal of Applied Econometrics **20**(7), 891–910 (2005)
21. Fernández-Villaverde, J., Rubio-Ramírez, J.: Estimating macroeconomic models: A likelihood approach. Review of Economic Studies **74**(4), 1059–1087 (2007)
22. Flury, T., Shephard, N.: Bayesian interference based only in simulated likelihood: Particle filter analysis of dynamic economic models. Econometric Theory **27**(Special Issue 05), 933–956 (2011)
23. Gordon, N., Salmond, D., Smith, A.: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEEE Proceedings F, Radar and Signal Processing **140**(2), 107–113 (1993)
24. Hansen, P., Lunde, A.: Realized variance and market microstructure noise. Journal of Business and Economic Statistics **24**(2) (2006)
25. Hu, X., Kuipers, D., Zeng, Y.: Econometric analysis via filtering for ultra-high frequency data. Working Paper, Department of Mathematics and Statistics, University of Missouri at Kansas City (2010)
26. Jacod, J., Todorov, V.: Testing for common arrivals of jumps for discretely observed multidimensional processes. Annals of Statistics **37**, 1792–1838 (2009)
27. Jasra, A., Stephens, D., Doucet, A., Tsagaris, T.: Inference for Lévy-driven stochastic volatility models via adaptive sequential Monte Carlo. Working paper, Institute for Statistical Mathematics, Tokyo, Japan (2008)
28. Johannes, M., Polson, N.: MCMC methods for continuous-time financial econometrics. In: Handbook of Financial Econometrics, vol. 2, pp. 1–66. North Holland (2009)
29. Li, H.: Sequential Bayesian analysis of time-changed infinite activity derivatives pricing models. Working paper, ESSEC Business School, Paris/Singapore (2009)

30. Li, H., Wells, M., Yu, C.: A Bayesian analysis of return dynamics with Lévy jumps. Review of Financial Studies **21**, 2345–2378 (2008)
31. Liu, J., West, M.: Combined parameter and state estimation in simulation-based filtering. In: Sequential Monte Carlo Methods in Practice. Springer (2001)
32. Liu, J.S.: Monte Carlo Strategies in Scientific Computing. Springer (2001)
33. Pitt, M., Shephard, N.: Filtering via simulation: Auxiliary particle filters. Journal of the American Statistical Association **94**(446), 590–599 (1999)
34. Pitt, M., Silva, R., Giordani, P., Kohn, R.: On some properties of Markov Chain Monte Carlo simulation methods based on the particle filter. Journal of Econometrics **171**(2), 134 – 151 (2012)
35. Polson, N., Stroud, J., Müller, P.: Practical filtering with sequential parameter learning. Journal of the Royal Statistical Society, Series B **70**, 413–428 (2008)
36. Raggi, D., Bordignon, S.: Sequential Monte Carlo methods for stochastic volatility models with jumps. Working paper, Department of Economics, University of Bologna (2006)
37. Ross, S.: Simulation. Elsevier, Burlington, MA. (2006)
38. Storvik, G.: Particle filters for state-space models with the presence of unknown static parameters. Signal Processing, IEEE Transactions on **50**(2), 281 –289 (2002)
39. Todorov, V., Tauchen, G.: Activity signature functions for high-frequency data analysis. Journal of Econometrics **54**, 125–138 (2010)
40. Todorov, V., Tauchen, G.: Volatility jumps. Journal of Business and Economic Statistics **29**, 356–371 (2011)
41. of USA, S..E.C.: Concept release on equity market structure. Release No. 34-61358; File No. S7-02-10, SEC. 17 CFR PART 242 (2010)
42. Wei, G.C.G., Tanner, M.A.: A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. Journal of the American Statistical Association **85**(411), pp. 699–704 (1990)
43. Zeng, Y.: A partially observed model for micromovement of asset prices with Bayes estimation via filtering. Mathematical Finance **13**(3), 411–444 (2003)
44. Zeng, Y.: Estimating stochastic volatility via filtering for the micromovement of asset prices. Automatic Control, IEEE Transactions on **49**(3), 338 – 348 (2004)
45. Zhang, L., Mykland, P., Aït-Sahalia, Y.: A tale of two time scales: Determining integrated volatility with noisy high-frequency data. Journal of the American Statistical Association **100**, 1394–1411 (2005)

Chapter 2
The Extended Liu and West Filter: Parameter Learning in Markov Switching Stochastic Volatility Models

Maria Paula Rios and Hedibert Freitas Lopes

2.1 Introduction

Since the seminal chapter by Gordon, Salmond and Smith (1993) with its *Bootstrap Filter* (*BF*), simulation-based sequential estimation tools, commonly known as sequential Monte Carlo (SMC) methods or particle filters (PF), have been receiving increasing attention in its application to nonlinear and non-Gaussian state-space models. There has been a particular emphasis on the application of such methods in state estimation problems in target tracking, signal processing, communications, molecular biology, macroeconomics, and financial time series (see compendium edited by Doucet, De Freitas and Gordon (2001)).

Nonetheless, only recently sequential parameter estimation started to gain more formal attention, with Liu and West (2001) (LW, hereafter) being one of the first contributions to the area. Their main contribution was the generalization of the SMC filter of Pitt and Shephard (1999), namely the *Auxiliary Particle Filter* (APF). LW incorporated sequential parameter learning in the estimation. Amongst other recent contributions in this direction are the *Practical Filter* of Polson, Stroud and Muller (2008) and the *Particle Learning* scheme of Carvalho, Johannes, Lopes and Polson (2010). The former relies on sequential batches of short MCMC runs while the latter relies on a recursive data augmentation argument, both of which aimed at replenishing the particles for both states and parameters. They also rely on the idea of sequential sufficient statistics for sequential parameter estimation (Storvik (2002) and Fearnhead, (2002)).

Implementation of the LW filter in various disciplines has shown that this methodology produces degenerate parameter estimates as discussed in Carvalho *et al.* (2010). Here we use volatility models to evidence the latter. One appreciates that the LW parameter estimates collapse to a point as further discussed

M.P. Rios • H.F. Lopes (✉)
Booth School of Business, University of Chicago 5807 S Woodlawn Avenue, Chicago, IL 60637, USA
e-mail: maria@chicagobooth.edu; hlopes@chicagobooth.edu

Y. Zeng and S. Wu (eds.), *State-Space Models: Applications in Economics and Finance*, Statistics and Econometrics for Finance 1, DOI 10.1007/978-1-4614-7789-1_2,

(see Figs. 2.4 and 2.5). Parameter degeneracy limits the applicability of the LW methodology. In particular, without proper parameter estimates one cannot make accurate forecasts, which are desired in many of the applications where filters are implemented.

To overcome the limitations of the LW filter, we explore three more filters of similar nature. Using the APF and BF as starting points for the propagation and resampling of the latent state, we incorporate sequential parameter learning techniques to extend these two filters to accommodate for parameter estimation. The first algorithm relies on the kernel smoothing idea that LW present when introducing their filter (see Liu and West (2001)). The second methodology relies on parameter estimation via recursive computation of conditionally sufficient statistics. In short, we construct four filters[1] that are hybrids between the BF, APF, kernel smoothing, and sufficient statistics.

Throughout the chapter we emphasize our analysis on two filters of particular interest, the LW filter and the so-called APF + SS filter. The latter is the extension of the APF filter that incorporates conditional sufficient statistics (SS) in the fixed parameter estimation.

To highlight the shortcomings of the LW filter and the applicability and improvements the APF + SS filter and the other two filters introduced, we focus only on one of the many applications where this technique is relevant. In this chapter we revisit the work of Carvalho and Lopes (2007). They used the LW filter SMC for state filtering and sequential parameter estimation in Markov switching stochastic volatility (MSSV) models. Using Carvalho and Lopes (2007) as reference, we implement the filters to the estimation of MSSV models. We empirically show, using simulated and real data, that LW filter degenerates, has larger Monte Carlo error, and in general terms underperforms when compared to the other filters of interest.

2.1.1 Volatility Models

Bayesian filters are a general technique that have a broad application scope. As shown in Carvalho *et al.* (2010), particle learning techniques can be implemented in Gaussian dynamic linear models (GDLM) and conditional dynamic linear models (CDLM). In this chapter, however, we focus only on one of the possible applications of the filters of interest. In particular we estimate fixed parameters and latent states in MSSV models.

Over the years, stochastic volatility models have been considered a useful tool for modeling time-varying variances, mainly in financial applications where agents are constantly facing decisions dependent on measures of volatility and risk. Bayesian estimation of stochastic volatility models can be found Jacquier *et al.* (1994) and Kim *et al.* (1998). Comprehensive reviews of stochastic volatility models can be found in Ghysels *et al.* (1996).

[1] Two of the filters we construct have been previously described by Liu and West (2001) and Storvik (2002).

2.1.1.1 Log-Stochastic Volatility

The building block for the MSSV models is the standard univariate log-stochastic volatility model, SV hereon, (see, for example, Jacquier *et al.* (1994), or Ghysels *et al.* (1996)), where (log) returns r_t and log-volatility states λ_t follow a state-space model of the form,

$$r_t = \exp\{\lambda_t/2\}\varepsilon_t \tag{2.1}$$
$$\lambda_t = \alpha + \eta\lambda_{t-1} + \tau\eta_t \tag{2.2}$$

where the errors, ε_t and η_t, are independent standard normal sequences. We also assume the initial log-volatility follows $\lambda_0 \sim N(m_0, C_0)$. The parameter vector, θ_{sv}, consists of the volatility mean reversion parameters $\psi = (\alpha, \eta)$ and the volatility of volatility τ. It is worth mentioning that the model assumes conditional independence of the r_t, $t = 1, \ldots, T$ variables.

2.1.1.2 Markov Switching Stochastic Volatility

Jumps have been a broadly studied characteristic of financial data (see, for example, Eraker *et al.* (2003)). So *et al.* (1998) suggest a model that allows for occasional discrete shifts in the parameter determining the level of the log-volatility through a Markovian process. They claim that this model not only is a better way to explain volatility persistence but is also a tool to capture changes in economic forces, as well as abrupt changes due to unusual market forces.

So *et al.*'s (1998) approach generalizes the SV model to include jumps by allowing the state space equation to follow a process that changes according to an underlying regime that determines the parameters for λ_t. For this, assume that s_t is an unobserved discrete random variable with domain $\{1, 2, \ldots, k\}$. Assuming a k-state first-order Markov process, we define the transition probabilities as

$$p_{j,l} = P(s_t = l | s_{t-1} = j) \qquad \text{for } j, l = 1, \ldots, k \tag{2.3}$$

with $\sum_{j=1}^{k} p_{ij} = 1$ for $i = 1, \ldots, k$. As suggested in So *et al.* (1998) (2.2) can be generalized to include such regime changes in the α parameter. Carvalho and Lopes (2007) suggest that α in this model corresponds to the level of the log-volatility and in order to allow occasional changes the model introduces different values α's following the described first-order Markovian process.

Again, let r_t be the observed process, just like it was defined for the SV model, with observations $r_1, \ldots, r_t$ conditionally independent and identically distributed. To keep consistency with the previously defined SV model, same notation and the normality and independence assumptions on the error terms will also be used here. This means that the observation r_t, $t = 1, \ldots, T$ is normal with time-varying log-volatilities $\lambda_1, \ldots, \lambda_T$. More specifically,

$$r_t = \exp\{\lambda_t/2\}\varepsilon_t \tag{2.4}$$

$$\lambda_t = \alpha_{s_t} + \eta\lambda_{t-1} + \tau\eta_t \tag{2.5}$$

Let $\xi = (\alpha, \eta, \tau^2)$, $\alpha = (\alpha_1, \ldots, \alpha_k)$, $p = (p_{11}, p_{1,2}, \ldots, p_{1,k-1}, \ldots, p_{k,1}, \ldots, p_{k,k-1})$, then $\theta_{\text{MSSV}} = (\xi, p)$ is the set of (k^2+2) parameters to estimate at each point in time. For instance, in a two-state model, six parameters must be estimated. It is common in the literature to refer to $S = (s_1, \ldots, s_T)$ and $\lambda = (\lambda_1, \ldots, \lambda_T)$ as the states of the model. The initial value of λ, λ_0, is $\text{N}(m_0, C_0)$.

To avoid identification issues in α, So *et al.* (1998) suggest to re-parameterize it as

$$\alpha_{s_i} = \gamma_1 + \sum_{j=1}^{k} \gamma_j I_{ji} \tag{2.6}$$

where $I_{ji} = 1$ when $s_i \geq j$ and 0 otherwise, $\gamma_1 \in R$ and $\gamma_i > 0$ for all $i > 1$. The model described by (2.4)–(2.6) is known as an MSSV model. As previously discussed the case where $k = 1$ reduces to the SV model presented above.

In this chapter we explore to cases of the MSSV model, $k = 1$ (or log-stochastic volatility) and $k = 2$. We fit these two models to the simulated and real data that we explore in Sects. 2.3 and 2.4.

2.1.2 Particle Filters: A Brief Review

Particle filters are SMC methods that basically rely on a *sampling importance resampling* (SIR) argument in order to sequentially reweigh and/or resample particles as new observations arrive. More specifically, let the general state-space model be defined by

$$\text{Observation equation} : p(y_{t+1}|x_{t+1}) \tag{2.7}$$

$$\text{State equation} : p(x_{t+1}|x_t) \tag{2.8}$$

where, for now, all static parameters are kept known. The observed variables y_t and the latent state variables x_t can be univariate or multivariate, discrete or continuous. Nonetheless, for didactical reasons, we will assume both are continuous scalar quantities.

Particle filters aim at computing/sampling from the filtering density[2]

$$p(x_{t+1}|y^t) = \int p(x_{t+1}|x_t, y^t) p(x_t|y^t) dx_t \tag{2.9}$$

and computing/sampling the posterior density via Bayes' theorem

$$p(x_{t+1}|y^{t+1}) \propto p(y_{t+1}|x_{t+1}) p(x_{t+1}|y^t) \tag{2.10}$$

[2] To avoid confusion, y^t makes reference to all the data observed up to point t, while y_t refers to the data observation at time t.

Put simply, PFs are Monte Carlo schemes whose main objective is to obtain draws $\{x_{t+1}^{(i)}\}_{i=1}^{N}$ from the state posterior distribution at time $t+1$, $p(x_{t+1}|y^{t+1})$, when the only draws available are $\{x_t^{(i)}\}_{i=1}^{N}$ from the state posterior distribution at time t, $p(x_t|y^t)$.

Recent reviews of PFs are Lopes and Tsay (2011), Olsson *et al.* (2008), Doucet and Johansen (2010), and Lopes and Carvalho (2011).

In what follows we briefly review two of the most popular filters for situations where static parameters are known. The BF or *sequential importance sampling with resampling* (SISR) filter and the APF. For these filters we assume that $\{x_0^{(i)}\}_{i=1}^{N}$ is a sample from $p(x_0|y^0)$.

2.1.2.1 The Bootstrap Filter

Gordon, Salmond and Smith's (1993) seminal filter basically uses the transition (2.8) in order to propagate particles, which then are resampled from the model (2.7). The BF can be summarized in the following two steps:

1. *Propagation.* Particles $\tilde{x}_{t+1}^{(i)}$ are drawn from $p(x_{t+1}|x_t^{(i)})$, for $i = 1,\ldots,N$, so the particle set $\{\tilde{x}_{t+1}^{(i)}\}_{i=1}^{N}$ approximates the filtering density $p(x_{t+1}|y^t)$ from (2.9).
2. *Resampling.* The SIR argument converts prior draws into posterior draws by resampling from $\{\tilde{x}_{t+1}^{(i)}\}_{i=1}^{N}$ with weights proportional to the likelihood, $\omega_{t+1}^{(i)} \propto p(y_{t+1}|\tilde{x}_{t+1}^{(i)})$, for $i = 1,\ldots,N$.

If the resampling step is replaced simply by a reweighing step, then the weights are replaced by $\omega_{t+1}^{(i)} \propto \omega_t^{(i)} p(y_{t+1}|\tilde{x}_{t+1}^{(i)})$, where $\omega_0^{(i)}$ is usually set at $1/N$. The SIR scheme samples from the prior and avoids the potentially expensive and/or practically intractable task of point-wise evaluation of $p(x_{t+1}|y^t)$. The flexibility and generality that comes with this *blind* scheme is the usually unbearable price of high Monte Carlo errors. More importantly, it leads to *particle degeneracy*, a Monte Carlo phenomenon where, after a few recursions of steps 1 and 2 above, all particles collapse into a few points and eventually to one single point.

2.1.2.2 The Auxiliary Particle Filter

One of the first *unblinded* filters was proposed by Pitt and Shephard (1999), who resample old draws with weights proportional to the proposal or candidate density $p(y_{t+1}|g(x_t))$, for some function g, such as the mean or the mode of the evolution density, and then propagate such draws via the evolution equation in (2.9). Finally, propagated draws are resampled with weights given by step 2 below. Their argument is based on a Monte Carlo approximation to

$$p(x_{t+1}|y^{t+1}) = \int p(x_{t+1}|x_t, y^{t+1}) p(x_t|y^{t+1}) dx_t \tag{2.11}$$

which is based on the one-step smoothing density $p(x_t|y^{t+1})$. Pitt and Shephard's (1999) APF can be summarized in the following three steps:

1. *Resampling.* The set $\{\tilde{x}_t^{(i)}\}_{i=1}^N$ is sampled from $\{x_t^{(i)}\}_{i=1}^N$ with weights $\{\pi_{t+1}^{(i)}\}_{i=1}^N$, where $\pi_{t+1}^{(i)} \propto p(y_{t+1}|g(x_t^{(j)}))$.
2. *Propagation.* The transition equation $p(x_{t+1}|\tilde{x}_t^{(i)})$ is used to draw $\tilde{x}_{t+1}^{(i)}$, whose corresponding weight is $\omega_{t+1}^{(i)} \propto p(y_{t+1}|\tilde{x}_{t+1})/p(y_{t+1}|g(\tilde{x}_t))$, for $i = 1,\ldots,N$.
3. *Posterior draws.* The set $\{x_{t+1}^{(i)}\}_{i=1}^N$ is sample from $\{\tilde{x}_{t+1}^{(i)}\}_{i=1}^N$ with weights $\{\omega_{t+1}^{(i)}\}_{i=1}^N$.

Our main contributions are twofold. Firstly, by comparing the four filters of interest we highlight the limitations of the LW-type filters for two cases of MSSV models. Secondly, we introduce an extension of the APF filter to overcome such limitations and produce more accurate estimates. The remainder of the chapter is organized as follows. In the next section we introduce the sequential parameter learning strategies that we then incorporate in the two filters previously discussed to extend them to allow for parameter estimation (the LW filter is one of such extensions). Results are analyzed in two sections. Section 2.3 presents and analyzes all the simulated data study while Sect. 2.4 presents real data applications. Section 2.5 concludes.

2.2 Particle Filters with Parameter Learning

We extend the BF and APF filtering strategies to allow for fixed parameter learning. Incorporating two techniques to each of the filters we study the four resulting types of Bayesian filters, which will be compared and evaluated in order to determine which filter outperforms the rest.

2.2.1 Kernel Smoothing

The first strategy that we incorporate for fixed parameter estimation is kernel smoothing, KS hereon, that was introduced in Liu and West (2001).

Liu and West (2001) generalizes the Pitt and Shephard's (1999) APF to accommodate sequential parameter learning. They rely on West's (1993) mixture of normals argument, which assumes that, for a fixed parameter vector θ,

$$p(\theta|y^t) \approx \sum_{i=1}^{N} f_N(\theta; m_t^{(i)}, h^2 V_t) \tag{2.12}$$

where $f_N(\theta; a, b)$ is the density of a multivariate normal with mean a and variance–covariance matrix b evaluated at θ. $\{\theta_t^{(i)}\}_{i=1}^N$ approximates $p(\theta|y^t)$, V_t approximates

the variance of θ given y^t, h^2 is a smoothing factor, and $m_t(\theta^{(i)}) = a\theta_t^{(i)} + (1-a)\bar{\theta}_t$ for $\bar{\theta}_t$ an approximation to the mean of θ given y^t and a a shrinkage factor, usually associated with h through $h^2 = 1 - a^2$.

The performance of filters that implement a KS strategy depends on the choice of the tuning parameter a, which drives both the shrinkage and the smoothness of the normal approximation. It is common practice to use a around 0.98 or higher. Also, the normal approximation can be easily adapted to other distributions, such as the normal-inverse-gamma approximation for conditionally conjugate location-scale models.

2.2.1.1 APF + KS: LW Filter

The first filter we consider is the so-called LW filter. This filter incorporates the KS strategy to the APF filter (see Liu and West (2001)). This is the filter used by Carvalho and Lopes (2007) to sequentially learn about parameters and state in a MSSV model.

Let the particle set $\{(x_t, \theta_t)^{(i)}\}_{i=1}^N$ approximate $p(x_t, \theta|y^t)$, $\bar{\theta}_t$ and V_t estimates of the posterior mean and posterior variance of θ, respectively, with $g(x_t^{(i)}) = E(x_{t+1}|x_t^{(i)}, m(\theta_t^{(i)}))$ and $m_t(\theta^{(i)}) = a\theta_t^{(i)} + (1-a)\bar{\theta}_t$, for $i = 1, \ldots, N$. The LW filter can be summarized in the following three steps:

1. *Resampling.* The set $\{(\tilde{x}_t, \tilde{\theta}_t)^{(i)}\}_{i=1}^N$ is sampled from $\{(x_t, \theta_t)^{(i)}\}_{i=1}^N$ with weights $\{\pi_{t+1}^{(i)}\}_{i=1}^N$, where $\pi_{t+1}^{(i)} \propto p(y_{t+1}|g(x_t^{(i)}), m(\theta_t^{(i)}))$;
2. *Propagation.* For $i = 1, \ldots, N$,
   (a) *Propagating* θ. Sample $\tilde{\theta}_{t+1}^{(i)}$ from $N(m(\tilde{\theta}_t^{(i)}), h^2 V_t)$,
   (b) *Propagating* x_{t+1}. Sample $\tilde{x}_{t+1}^{(i)}$ from $p(x_{t+1}|\tilde{x}_t^{(i)}, \tilde{\theta}_{t+1}^{(i)})$,
   (c) *Computing weights.* $\omega_{t+1}^{(i)} \propto p(y_{t+1}|\tilde{x}_{t+1}^{(i)}, \tilde{\theta}_{t+1}^{(i)}) / p(y_{t+1}|g(\tilde{x}_t^{(i)}), m(\tilde{\theta}_t^{(i)}))$;
3. *Posterior draws.* The set $\{(x_{t+1}, \theta_{t+1})^{(i)}\}_{i=1}^N$ is sampled from $\{(\tilde{x}_{t+1}, \tilde{\theta}_{t+1})^{(i)}\}_{i=1}^N$ with weights $\{\omega_{t+1}^{(i)}\}_{i=1}^N$.

2.2.1.2 BF + KS

The second filter that we analyze in this chapter is the extension of the BF when we include the KS strategy in the fixed parameter estimation. The following algorithm summarizes the BF + KS filter.

Let the particle set $\{(x_t, \theta_t)^{(i)}\}_{i=1}^N$ approximate $p(x_t, \theta|y^t)$, $\bar{\theta}_t$ and V_t estimates of the posterior mean and posterior variance of θ, respectively, with $g(x_t^{(i)}) = E(x_{t+1}|x_t^{(i)}, m(\theta_t^{(i)}))$ and $m_t(\theta^{(i)}) = a\theta_t^{(i)} + (1-a)\bar{\theta}_t$, for $i = 1, \ldots, N$.

1. *Propagation.* For $i = 1, \ldots, N$,
   (a) *Propagating* x_{t+1}. Sample $\tilde{x}_{t+1}^{(i)}$ from $p(x_{t+1}|\tilde{x}_t^{(i)}, \tilde{\theta}_t^{(i)})$.
   (b) *Propagating* θ. Sample $\tilde{\theta}_{t+1}^{(i)}$ from $N(m(\tilde{\theta}_t^{(i)}), h^2 V_t)$.
   (c) *Computing weights.* $\omega_{t+1}^{(i)} \propto p(y_{t+1}|\tilde{x}_{t+1}^{(i)}, \tilde{\theta}_{t+1}^{(i)})$.
2. *Posterior draws.* The set $\{(x_{t+1}, \theta_{t+1})^{(i)}\}_{i=1}^N$ is sampled from $\{(\tilde{x}_{t+1}, \tilde{\theta}_{t+1})^{(i)}\}_{i=1}^N$ with weights $\{\omega_{t+1}^{(i)}\}_{i=1}^N$.

2.2.2 *Sufficient Statistics*

The second method that we consider for sequential parameter learning is the recursive sufficient statistics, SS hereon. This technique can be implemented in situations where the vector of fixed parameters θ admits recursive conditional sufficient statistics (Storvik, (2002), and Fearnhead, (2002)). That is the prior for θ is

$$p(\theta) = p(\theta|s_0) \tag{2.13}$$

One of the main advantages of this estimation technique is that Monte Carlo error is reduced by decreasing the number of parameters in Liu and West's kernel mixture approximation. In addition, tracking sufficient statistics can be seen as replacing the sequential estimation of fixed parameters by the sequential updating of a low-dimensional vector of deterministic states. This is particularly important when sequentially learning about variance parameters. See Carvalho *et al.* (2010) for further discussion. Furthermore, this methodology reduces the variance of the sampling weights, resulting in algorithms with increased efficiency and helps delaying the decay in the particle approximation often found in algorithms based on SIR.[3]

[3] As an illustration, we present the SS for the MSSV $k = 2$ model.
The following two equations define the model.

$$\begin{aligned} r_t|\lambda &\sim N(0, e_t^{\lambda}) \\ \lambda_t|\lambda_{t-1}, \alpha, \eta, \gamma &\sim N(\alpha + \eta\lambda_{t-1} + \gamma s_t \sim N(\alpha + \eta\lambda_{t-1} + \gamma s_t, \tau^2) \end{aligned}$$

Priors and hyperparameter values are defined in Sect. 2.3. Let $x't = (\mathbf{1}, \lambda_{t-1}, s_t)$ and $\theta = (\alpha, \eta, \gamma)$. Therefore, conjugacy leads to

$$\begin{aligned} (\alpha, \eta, \gamma)|\tau^2, x)_{1:t} &\sim N(a_t, \tau A_t)\mathbf{1}_{\gamma>0} \\ \tau^2|x_{1:t} &\sim IG(v_t/2, v_t\tau^2/2) \\ p_{1,1}|s_{1:t} &\sim Beta(n_{11t}, n_{12t}) \\ p_{2,2}|s_{1:t} &\sim Beta(n_{21t}, n_{22t}) \end{aligned}$$

2.2.2.1 APF + SS

The main filter that we showcase in this chapter is the SS extension to the APF filter which we describe below. We call this filter the APF+SS filter and will show its ability to overcome many of the limitations present in other filtering strategies, like the LW filter.

Let the particle set $\{(x_t,\theta,s_t)^{(i)}\}_{i=1}^N$ approximate $p(x_t,\theta,s_t|y^t)$ with $g(x_t^{(i)}) = E(x_{t+1}|x_t^{(i)})$. The APF + SS can be summarized as follows:

1. *Resampling.* The set $\{(\tilde{x}_t,\tilde{\theta},\tilde{s}_t)^{(i)}\}_{i=1}^N$ is sampled from $\{(x_t,\theta,s_t)^{(i)}\}_{i=1}^N$ with weights $\{\pi_t^{(i)}\}_{i=1}^N$, where $\pi_t^{(i)} \propto p(y_{t+1}|g(x_t^{(i)}))$.
2. *Propagation.* For $i = 1,\ldots,N$,
   (a) *Propagating* x_{t+1}. Sample $\tilde{x}_{t+1}^{(i)}$ from $p(x_{t+1}|\tilde{x}_t^{(i)},\tilde{\theta}^{(i)})$.
   (b) *Computing weights.* $\omega_{t+1}^{(i)} \propto p(y_{t+1}|\tilde{x}_{t+1}^{(i)},\tilde{\theta}^{(i)})/p(y_{t+1}|g(\tilde{x}_t^{(i)}),\tilde{\theta}^{(i)})$.
3. *Posterior draws.* The set $\{(x_{t+1},\theta,s_t)^{(i)}\}_{i=1}^N$ is sampled from $\{(\tilde{x}_{t+1},\tilde{\theta},\tilde{s}_t)^{(i)}\}_{i=1}^N$ with weights $\{\omega_{t+1}^{(i)}\}_{i=1}^N$.
4. *Update sufficient statistics.* $s_{t+1}^{(i)} = \mathscr{S}(\tilde{s}_t^{(i)},x_{t+1}^{(i)},y_{t+1})$, for $i = 1,\ldots,N$.
5. *Parameter learning.* Sample $\theta^{(i)} \sim p(\theta|s_{t+1}^{(i)})$, for $i = 1,\ldots,N$.

Both APF + SS and particle learning algorithms (PL) presented in Carvalho *et al.* (2010) are particle filters that resample old particles first and then propagate them. While PL and APF + SS are quite similar when dealing with the parameter sufficient statistics, PL approximates the $\log\chi^2$ distribution of log squared returns by Kim, Shephard and Chib's (1998) mixture of seven normal densities while APF + SS uses Pitt and Shephard's (1999) APF that approximates the predictive density with the likelihood function. Further investigation comparing these algorithms for more general classes of stochastic volatility models is an open area beyond the scope of this chapter.

2.2.2.2 BF + SS

The last filter that we consider in this chapter is the SS extension to the BF, as suggested in Storvik (2002). What we will refer to as the BF + SS filter can be summarized with the following steps:

---

where $x_{1:t} = \{x_1,\ldots,x_t\}$ and $v_t = v_{t-1}+1$. $A_t, a_t, v_t\tau^2$ and n_{ijt} are the sufficient statistics defined recursively by

$$\begin{aligned}
A_t^{-1} &= A_{t-1}^{-1} + x_t x_t' \\
A_t^{-1}a_t &= A_{t-1}^{-1}a_t + x_t\lambda_t \\
v_t\tau^2 &= v_{t-1}\tau_{t-1}^2 + \lambda_t^2 - x_t'a_t\lambda_t + a_{t-1}'A_{t-1}^{-1}a_{t-1} - a_t'A_{t-1}^{-1}a_{t-1} \\
n_{ij,t} &= n_{ij,t-1} + \mathbf{1}_{(s_{t-1}=i,s_t=j)}
\end{aligned}$$

Let the particle set $\{(x_t,\theta,s_t)^{(i)}\}_{i=1}^N$ approximate $p(x_t,\theta,s_t|y^t)$.

1. *Propagation.* For $i=1,\ldots,N$,
   (a) *Propagating* x_{t+1}. Sample $\tilde{x}_{t+1}^{(i)}$ from $p(x_{t+1}|\tilde{x}_t^{(i)},\tilde{\theta}^{(i)})$.
   (b) *Computing weights.* $\omega_{t+1}^{(i)} \propto p(y_{t+1}|\tilde{x}_{t+1}^{(i)},\tilde{\theta}^{(i)})$.
2. *Posterior draws.* The set $\{(x_{t+1},\theta,s_t)^{(i)}\}_{i=1}^N$ is sampled from $\{(\tilde{x}_{t+1},\tilde{\theta},\tilde{s}_t)^{(i)}\}_{i=1}^N$ with weights $\{\omega_{t+1}^{(i)}\}_{i=1}^N$.
3. *Update sufficient statistics.* $s_{t+1}^{(i)} = \mathscr{S}(s_t^{(i)},x_{t+1}^{(i)},\theta_{t+1}^{(i)},y_{t+1})$, for $i=1,\ldots,N$.
4. *Parameter learning.* Sample $\eta^{(i)} \sim p(\eta|s_{t+1}^{(i)},\theta_{t+1}^{(i)})$, for $i=1,\ldots,N$.

2.3 Analysis and Results: Simulation Study

The first part of the analysis presented is a simulation study that provides insight into the behavior of the four filters discussed in this chapter. We are able to identify limitations and benefits of using each approach. The particle filters are compared in four ways: (1) degree of particle degeneracy and estimation accuracy; (2) accuracy in estimating regime-switching parameters; (3) size of the Monte Carlo error, and (4) computational cost. Completing the study we discuss the economic insight that can be inferred from the Bayesian estimates and end with a robustness analysis to control for data set-specific effects.

Our simulation analysis is based on 50, 5,000 particle runs. For each of the 50 iterations, we drew a new set of priors that was used to initiate each one of the four filters of interest. In the two filters that use a KS technique we use a shrinkage/smoothness factor of $a=0.9$. For both the volatility process and the parameters, the median particle is used as the estimate and the 97.5 % and 2.5 % percentile particles are used as the upper and lower bounds of the 95 % confidence band, respectively.

Robustness results are based on ten different data sets and ten runs of the filters, each with a different starting set of prior draws.[4]

In the $k=1$ case, all filters' prior distribution for τ^2 is inverse gamma, i.e. $\tau^2 \sim IG(\nu_0/2,\nu_0\tau_0^2/2)$, with prior mean $\nu_0\tau_0^2/(\nu_0-2)$. For the filter that use sufficient statistics in the estimation (APF + SS and BF + SS), the prior distributions for α and η are conditionally conjugate, i.e. $\eta|\tau^2 \sim TN_{(-1,1)}(b_0,\tau^2 B_0)$ and $\alpha|\tau^2 \sim N(a_0,\tau^2 A_0)$, where $TN_A(a,b)$ is the normal distribution with mean a and variance b and truncated at A. For the filters with kernel smoothing (LW and BF + KS), the prior distributions are $\eta \sim TN_{(-1,1)}(b_0,B_0)$ and $\alpha \sim N(a_0,A_0)$. The difference between these priors has negligible effect on our empirical study.

[4] The same prior draws are used in one run of all four filters, thus ensuring that results in this run are comparable across filter.

Hyperparameter values are set up to ensure uninformative priors. In all application scenarios presented in this chapter we used the following setup: $m_0 = 0$, $C_0 = 1$, $a_0 = b_0 = 0$, $A_0 = B_0 = 3$, $\nu_0 = 4.01$ and $\tau_0^2 = 0.01$. Changes in hyperparameter values were made and no significant change was observed in the results.

In the $k = 2$ case we use the same priors and hyperparameter values for τ^2 and η as the ones just described for the log-stochastic volatility model. Additionally for all filters we have that $p_i \sim Dir(u_{i0})$[5] for $p_i = (p_{i1}, \ldots, p_{ik}), i = 1, \ldots, k$. For the filter kernel smoothing filters' implementation in this scenario we set $\gamma_1 \sim N(a_0, A_0)$ and $\gamma_i \sim TN_{(0,\infty)}(g_0, G_0) \quad i = 2, \ldots, k$. Once more, for the sufficient statistic-based filters priors, we condition on τ^2 for $\gamma_i, i = 1, \ldots k$. That is, we have $\gamma_1 | \tau^2 \sim N(a_0, \tau^2 A_0)$ and $\gamma_i | \tau^2 \sim TN_{(0,\infty)}(g_0, \tau^2 G_0)$ $i = 2, \ldots, k$. All of the hyperparameter values that were already defined for the SV remained unchanged, and the following new values were added: $u_{i0} = (0.5, \ldots, 0.5)$ for $i = 1, \ldots, k$, $g_0 = 0$ and $G_0 = 3$.

2.3.1 Simulated Data

As mentioned before, we focus our investigation on MSSV models, one of the many possible applications in which to implement the filters presented in Sect. 2.2. As mentioned before, we consider two cases of the number of states, k, in the model: (1) the *log-stochastic volatility* or $k = 1$ and (2) the two-state MSSV ($k = 2$).

These two models are simulated for a time frame of 1,000 time periods. For the *log-stochastic volatility* case we use $\alpha = -1$, $\eta = 0.9$, and $\tau^2 = 1$. Time series plots for the return y_t, latent state x_t, and volatility processes are presented in Fig. 2.1. In the $k = 2$ parameter values were chosen to match the values of the first data set used in Carvalho and Lopes (2007). The parameter vector, Θ_2, is determined by $\alpha_1 = -2.5, \alpha_2 = -1, \eta = 0.5, \tau^2 = 1, p_{11} = 0.99$, and $p_{22} = 0.985$. A graphical summary of the processes of interest, y_t, x_t, volatility and the state in which the process is on, s_t, are presented in Fig. 2.2.

2.3.2 Exact Estimation Path

Given the Bayesian nature of the filters analyzed in this chapter, we use an *exact estimation path* as a reference for which the *best* estimation path should be. Likewise, we use the confidence bands obtained in the *exact path estimation* as reference for what sensible confidence bands are for the estimates of interest. The path and bands are obtained by running one of the filters with a large enough number of particles what ensures that both the path and the bands will be replicated by the filter regardless of the prior draws used to initiate the filter. In a non-time constrained world this

[5] $Dir(u_{i0})$ means that the prior distribution is Dirichlet with parameter u_{i0}.

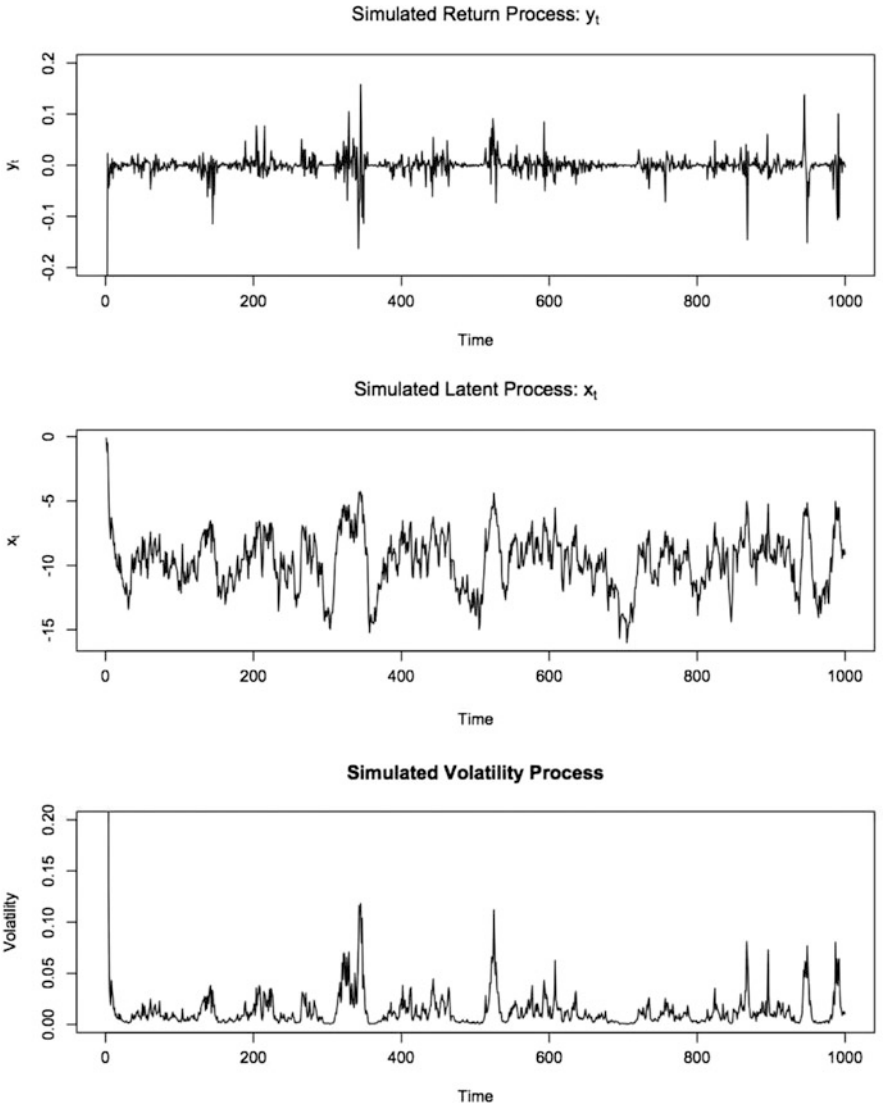

Fig. 2.1 Time series of the simulated return, latent state, and volatility processes for the MSSV model with $k = 1$. Details of the choice of model parameter values can be found in Sect. 2.3.1.

would be the ideal path to use; however, given the current computational capacity, running the filters for sufficiently large number of particles is not time efficient.

Under the premise that these path and bands are perceived as *the true path*, the choice of filter should be irrelevant. Here, the *exact estimation path* is calculated by running a 100,000 particle APF + SS filter. For both the $k = 1, 2$ simulated data

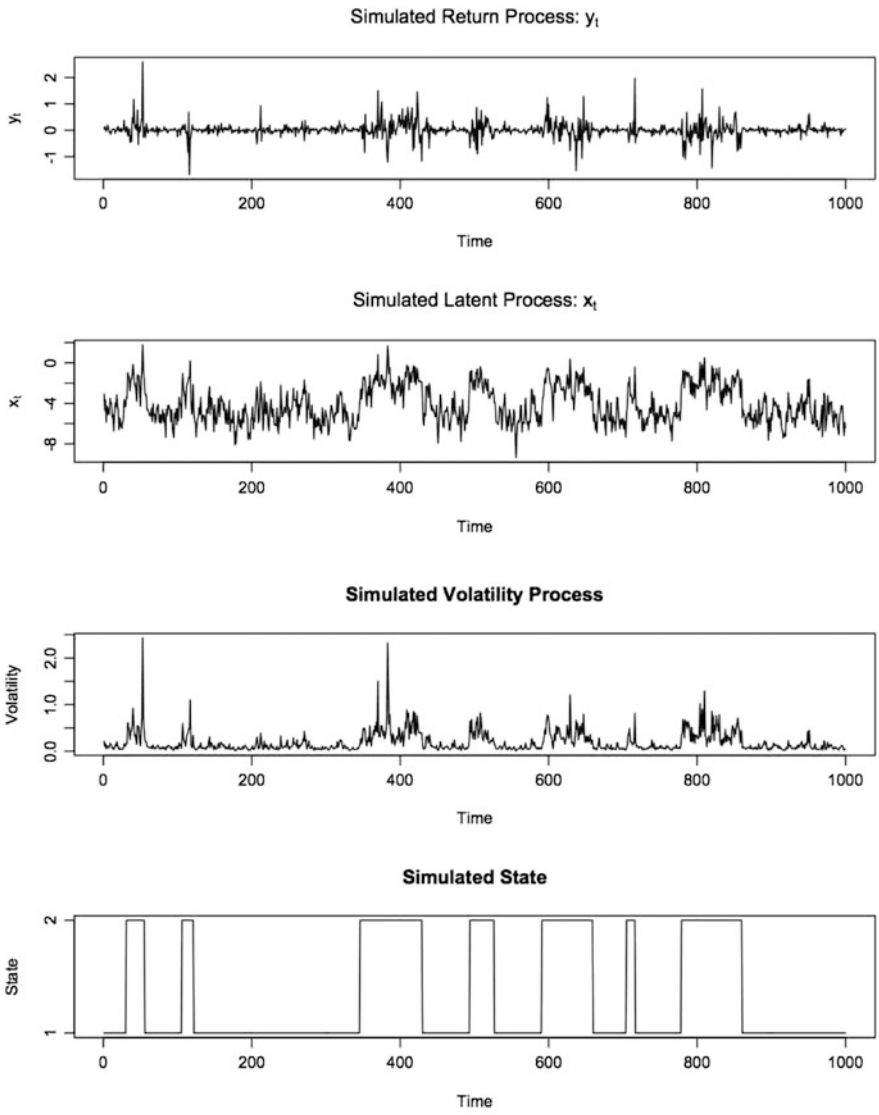

Fig. 2.2 Time series of the simulated return, latent state, and volatility processes for the MSSV model with $k = 2$. Details of the choice of model parameter values can be found in Sect. 2.3.1.

sets, the *exact* parameters and volatility paths and their 95 % confidence bands are estimated. In this chapter setting the *exact* estimates and confidence bands paths obtained will be regarded as the *true paths and bands* and as a result they will be the benchmark used to compare the filters.

2.3.3 Estimate Evaluation

2.3.3.1 Parameter Degeneration and Estimate Accuracy

The behavior of the filter estimates is first analyzed in terms of parameter degeneration and estimate accuracy. Determining how well the filters are able to correctly replicate the volatility processes and estimate the parameter values is paramount to the filters performance.

Correct latent state estimation is the ultimate goal of any filter. In order to evaluate how well the filters presented in this chapter are replicating the volatility process we compare the true simulated process with the filtered estimates. We use a *mean squared error* (MSE) to measure the deviation between the real and estimated processes. The MSE is defined by

$$MSE = \frac{1}{T}\sum_{t=1}^{T}\left(V_t - \hat{V}_t\right)^2 \tag{2.14}$$

where V_t is the real volatility process and $\hat{V}_t$ is the filtered volatility process.

Table 2.1 presents the mean MSE, averaged across the 50 filter repetitions,[6] for all filters and both volatility models. Divergence from the real volatility process is small and similar in all four filters for the two MSSV cases, showing that the filters are able to accurately replicate the latent state x_t, and thus produce good volatility estimates.

Closer inspection of the behavior of the estimated paths reveals that the discrepancies between the real and the estimated volatilities happen when there are sudden increases in volatility. None of the four filters are able to completely capture these peaks. The problem magnifies in the $k = 2$ case.

Table 2.1 Mean MSE between the real and the filtered volatility processes, averaged across the 50 repetitions of each filter.

k	APF + SS	BF + SS	BF + KS	LW
1	0.000520	0.000527	0.000578	0.000552
2	0.019394	0.019613	0.019155	0.020136

Another element that is worth exploring is the variability that exists within the MSE of the 50 repetitions. From Fig. 2.3 one appreciates that the LW filter results have a significantly wider spread. Moreover, we see that the two strategies that

[6] The mean MSE presented in Table 2.1 is averaged across repetitions for each one of the filters. That is:

$$\text{Mean } MSE = \frac{1}{50}\sum_{i=1}^{50} MSE_i = \frac{1}{50}\sum_{i=1}^{50}\frac{1}{T}\sum_{t=1}^{T}\left(V_t - \hat{V}_t\right)^2 \tag{2.15}$$

where MSE_i is the MSE for repetition i of a given filter.

involve an APF in the propagation and sampling of the underlying process are less stable than the two filters that implement a BF strategy. As expected, the lower the variability, the more powerful the claim we can make on the accuracy of the filters, as any run will likely have the same deviations from the real process.

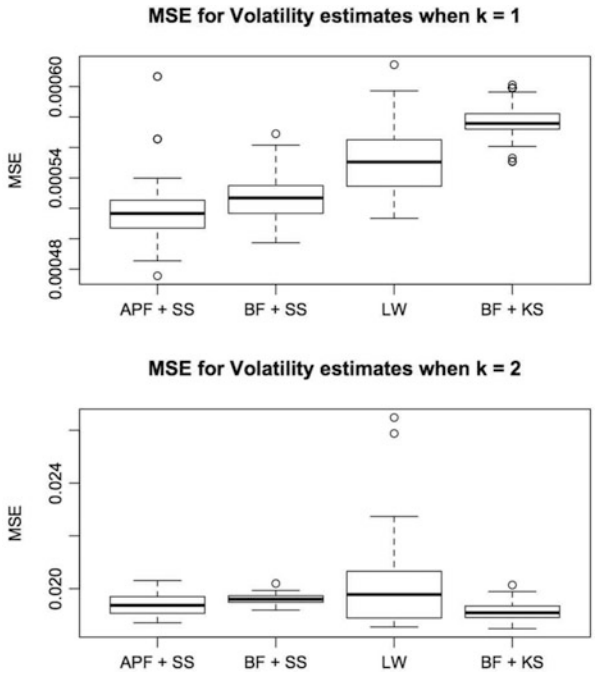

Fig. 2.3 Box plots of the MSE of the estimated volatility process compared to the real simulated volatility process for each filter. The *left plot* presents the results for the MSSV with $k = 1$ and the *right plot* presents the results for the MSSV with $k = 2$.

Switching to parameter accuracy we focus on parameter degeneracy, a phenomenon that appears when the resampling weights concentrate on one or a few mass points and make the parameter estimates and their confidence bands collapse to very narrow ranges and sometimes even to a single point. Our filter comparison uses the *exact path's* 95 % band as a benchmark for reasonable values of estimate's confidence credibility bandwidth. Furthermore, we use the *effective sample size* (*ESS*) to complement this part of the study. ESS is defined as:

$$ESS = \left(\sum_{t=1}^{T} w_t^2 \right)^{-1} \tag{2.16}$$

where w_t is the resampling weight at time t (see Lopes and Carvalho (2011) and Kong, Liu and Wong (1994)). This measure is a good proxy for the number of particles where the weights have mass, thus making them the most likely candidates in the resampling step. As we will further explore, there arguably is a relationship between parameter collapsing and ESS value.

The first component of the parameter degeneration analysis is to understand which filters, and in what proportion, present cases of the latter phenomenon. To this end, we look at how many filter runs have parameter 95 % confidence bands' width narrower than two threshold percentages of the *parameter exact 95 % confidence band's width*. In particular, we are interested in confidence credibility bandwidths narrower than 10 % and 20 % of the benchmark 95 % confidence band's width.

Table 2.2 presents a summary of the results for the two volatility models discussed. In the $k = 1$ case, only the LW filter presents collapsing parameters, with at least 20 % of the filter runs presenting this anomaly. In the $k = 2$ case all four filters have at least one parameter for which the estimates degenerate. Yet, it is again the LW filter the one that presents a more delicate situation with the most collapsing cases. At least 20 % of the filter's runs appear to produce defective parameter estimates. The high proportions of cases with parameter collapses issues found in the LW filter raise a flag on the accuracy and applicability of this widespread filter.

Table 2.2 The left side presents the number of runs that reveals parameter collapses in the 50 runs. A collapsing case is a filter repetition in which the width of the 95 % credibility bands is narrower than 10 % or 20 % of the *w*idth of the exact parameter path. The right side presents the 25 %, 50 %, and 75 % percentiles for the effective sample size of the non-collapsing filters.

		Parameter														Collapse	ESS		
k	Filter	α		α_1		α_2		η		τ^2		$p_{1,1}$		$p_{2,2}$			25 %	50 %	75 %
		0.1	0.2	0.1	0.2	0.1	0.2	0.1	0.2	0.1	0.2	0.1	0.2	0.1	0.2				
1	APF + SS	0	0					0	0	0	0					No	3,614.798	4,226.624	4,484.099
	BF + SS	0	0					0	0	0	0					No	3,352.356	3,915.876	4,322.197
	BF + KS	0	0					0	0	0	0					No	3,321.399	3,882.107	4,303.528
	LW	13	19					13	20	16	28					Yes			
2	APF + SS			0	0	0	5	0	0	0	0	0	1	2	44	No	3,616.132	4,151.599	4,422.509
	BF + SS			0	0	0	0	0	0	0	0	0	0	4	45	No	3,352.903	3,937.380	4,293.923
	BF + KS			0	0	0	0	0	0	0	0	2	16	25	32	No	3,317.990	3,919.167	4,301.583
	LW			11	22	12	32	12	25	22	22	11	23	32	48	Yes			

Graphical examples of the particle degeneration phenomenon for one run of the LW filter for both $k = 1, 2$ are presented in Figs. 2.4 and 2.5, respectively. The dotted lines highlight the points where the minimum ESS happens in the analyzed run. In the two showcased examples the minimum ESS obtained are 1.05 and 1.37 for the $k = 1, 2$ cases, respectively. In other words, it appears that for both cases the LW filter reaches a point where it will give weight to only a very small set of particles, thus only resampling from this limited set. The reader can see in the plots how this clearly translates. To the right of the dotted lines, the estimates collapse to almost a single point and the confidence bands become extremely narrow. Furthermore, for both MSSV models the values where the estimates collapse to are different to the true parameter values. Additionally, due to the band's narrowness, the true value does not lie within the estimates of 95 % confidence band, hence leading to erroneous conclusions about the parameter values. This is a critical estimation accuracy shortcoming of the LW filter.

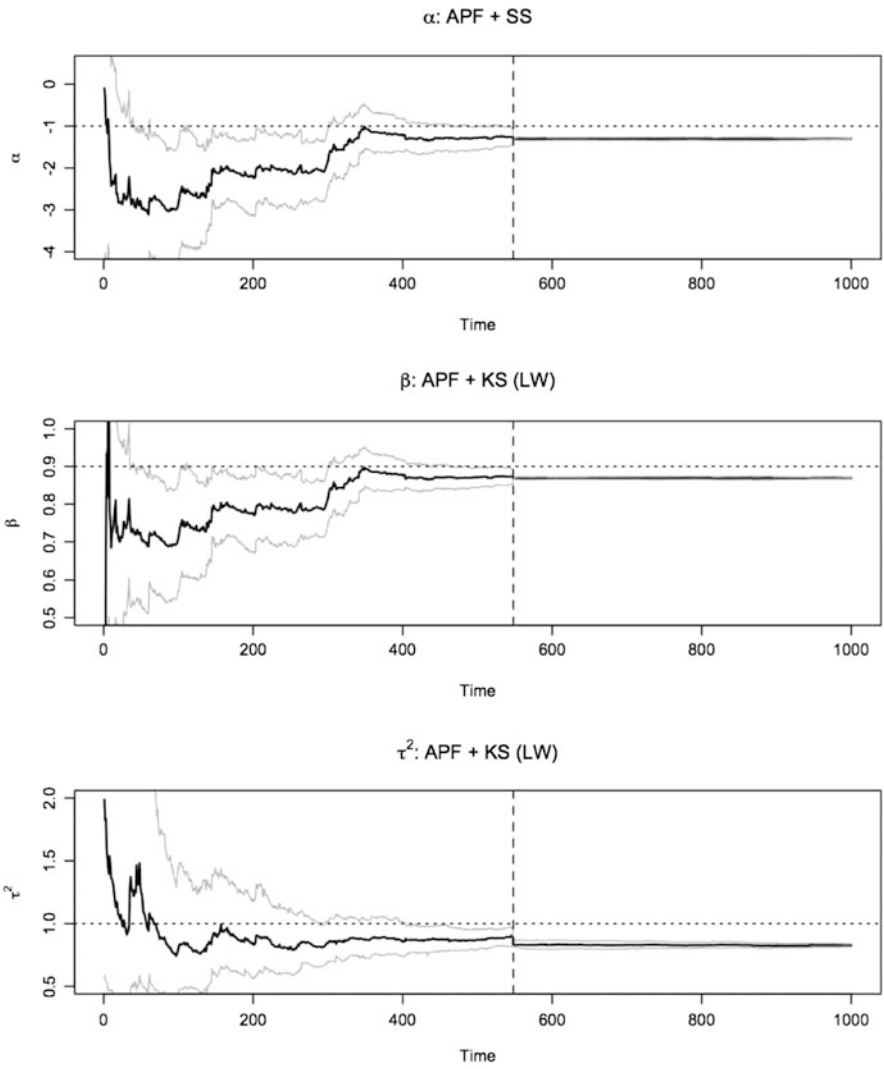

Fig. 2.4 Parameter estimates for LW Filter, in an MSSV $k = 1$ repetition where parameters collapse. *Black lines* represent parameter estimates (median particle for each time period); *gray lines* represent the 97.5 % and 2.5 % quantiles for the particle estimates and the *dotted line* is the true parameter value. The *dashed line* highlights the time period where the min ESS happens in that particular run.

Exploring the ESS behavior for the non-collapsing parameters, Table 2.2 presents[7] the 25 %, 50 %, and 75 % quantiles for the ESS values obtained in the

[7] In spite of the fact that the APF + SS, BF + SS, and BF + KS filters show evidence of parameter collapses in the $p_{2,2}$ parameter of the $k = 2$ MSSV model, we still consider them as non-collapsing filters. This is due to the fact that this phenomenon is only present in one of the parameters.

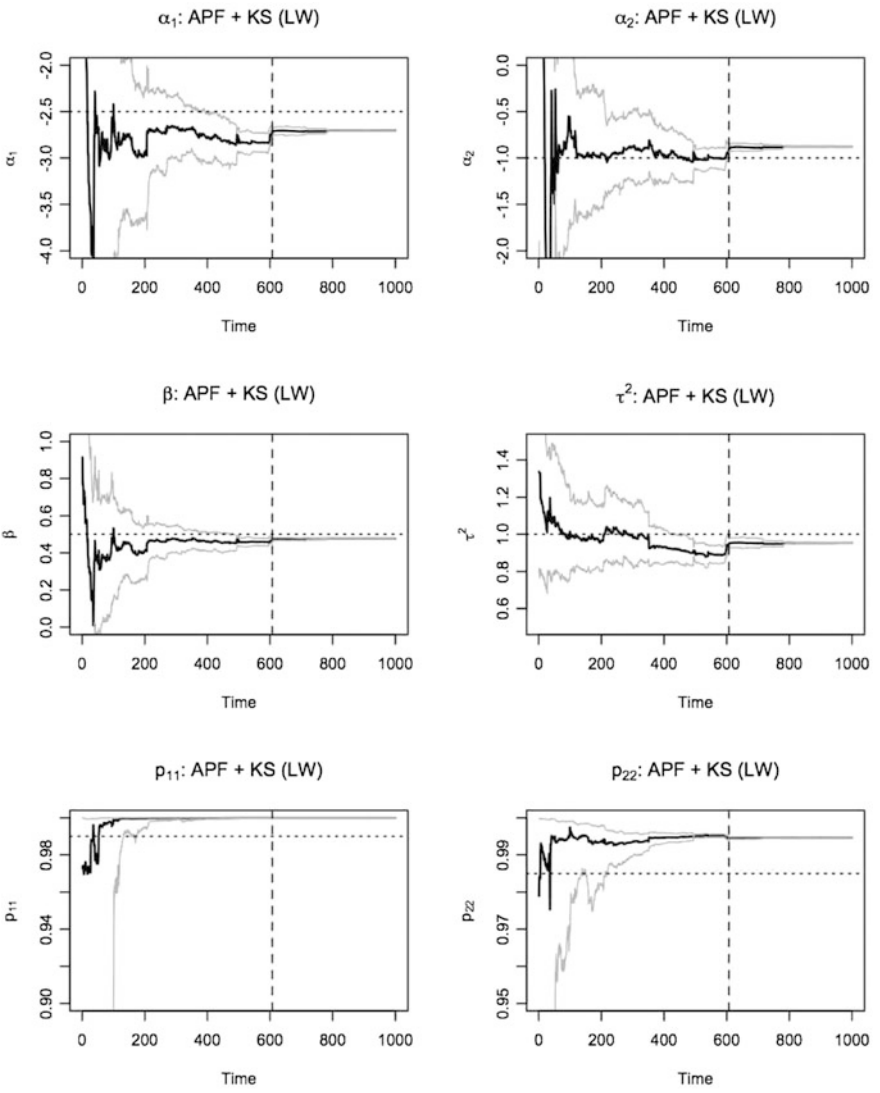

Fig. 2.5 Parameter estimates for LW Filter, in an MSSV $k = 2$ repetition where parameters collapse. *Black lines* represent parameter estimates (median particle for each time period); *gray lines* represent the 97.5 % and 2.5 % quantiles for the particle estimates and the *dotted line* is the true parameter value. The *dashed line* highlights the time period where the min ESS happens in that particular run.

50 replications of each filter. The latter results show that for the most part, the three filters of interest rely on *healthy* amounts of particles to resample from, ensuring variability in the resampling weights which is critical to accurate estimation of the parameters.

2.3.3.2 Regime-Switching Estimation

Particular to the type of volatility models we are estimating in cases where $k \geq 2$, it is important to understand how the filters are able to capture regime changes and track the states in which the model is at. For this we focus on analyzing the two-state MSSV.

Following Bruno and Otranto (2008), Otranto (2001), and Bodart et al. (2005) we use the *quadratic probability score* (*QPS*), developed by Diebold and Rudebusch (1989), to evaluate the filters' abilities to correctly determine the state in which the economy is at. The QPS is defined by

$$QPS = \frac{100}{T} \sum_{t=1}^{T} [Pr(S_t = 2) - D_t]^2 \tag{2.17}$$

where d_t is an indicator variable equal to 1 when the true process is in state 2 and $Pr(S_t = 2)$ is the estimated probability that the process is in the second state. The index varies between 0 and 100. It is equal to 0 in the case of correct assignment of the state variable for all time periods and equals 100 in the opposite case.

Table 2.3 Mean QPS for the $k = 2$ model.

APF + SS	BF + SS	BF + KS	LW
8.96	8.10	7.94	17.95

Table 2.3 presents a summary of the results. For each one of the four filters, the QPS averaged across the 50 replications is presented.[8] The APF + SS, BF + SS, and BF + KS filters all appear to have similar abilities of correctly estimating the correct state. The LW filter, however, produces considerably less accurate estimates of the regime where the economy is. This could be linked to the particle degeneracy found in the LW filter, as inaccurate parameter estimates lead to erroneous state estimates.

The SS-based methods are specially good at tracking regime changes. For certain scenarios and applications this is an extremely important feature. Thus this is another aspect in which we can claim that the LW filter has shortcomings while the APF + SS and BF + SS filters are outperforming.

[8] The mean QPS presented in Table 2.3 is averaged across repetitions for each one of the filters. That is:

$$\text{Mean } QPS = \frac{1}{50} \sum_{i=1}^{50} QPS_i = \frac{1}{50} \sum_{i=1}^{50} \frac{100}{T} \sum_{t=1}^{T} [Pr(S_t = 2) - D_t]^2 \tag{2.18}$$

where QPS_i is the QPS for repetition i of a given filter.

2.3.3.3 Monte Carlo Error

Next, the four filters are evaluated in terms of stability of the produced estimates. In other words, how much Monte Carlo variability is found in the estimates. Under ideal conditions we would like to have parameter estimates that perfectly replicate the estimation paths regardless of the set of prior draws used to initialize the filters. However, this is not a realistic scenario, and all four filters of interest have some Monte Carlo variability, or as we like to call it Monte Carlo error.

In order to analyze the latter variability we once more use the *exact estimate paths* described in Sect. 2.3.2 as benchmark. Assuming that the *exact paths* are what the *true* estimate paths should look like, we analyze the deviations between the different estimate paths that the filters produce and the so-called *exact path*.

A preliminary graphical exploration of the estimates allows to get a first impression of the way that the estimates behave. Figures 2.6–2.8 present the estimate paths for the three parameters in the $k = 1$ MSSV model, while Figs. 2.9–2.14 show the paths for the estimates of the six parameters in the $k = 2$ MSSV model. In these panels, the solid black lines present the *exact path* and the gray lines are the estimate paths for each one of the 50 runs of each filter.

The plots reveal that the two filters that have the least Monte Carlo variability are the two using an SS approach in the fixed parameter estimation. On the other hand, the filter that consistently appears to have the largest variability is the so-called LW filter, which in the $k = 2$ model has considerably greater variation for $\eta, p_{1,1}$, and $p_{2,2}$.

A more rigorous way to analyze the Monte Carlo error is to *measure* the deviation between the *exact path* and the estimate path. To avoid confusion with the MSE previously used, we here use the mean absolute error between the two paths, which we call the Monte Carlo mean absolute error (MCMAE), defined by

$$MCMAE = \frac{1}{T}\sum_{t=1}^{T} |p_t - \hat{p}_t| \tag{2.19}$$

where p_t is the *exact parameter path* and $\hat{p}_t$ is the estimated path at time t. Given that we are looking at 50 runs of each one of the filters we will focus on analyzing the mean across runs of the MCMAE. A summary of the mean MCMAE results for all the parameters in the two MSSV models is shown in Table 2.4.[9]

The results in the latter table corroborate the graphical findings. The APF + SS and BF + SS filters have the smaller deviations. For most parameters the APF + SS has slightly lower Monte Carlo error than the BF + SS. Analyzing the behavior

[9] The mean MCMAE presented in Table 2.4 is averaged across repetitions for each one of the filters. That is:

$$\text{Mean } MCMAE = \frac{1}{50}\sum_{i=1}^{50} MCMAE_i = \frac{1}{50} = \frac{1}{T}\sum_{t=1}^{T} |p_t - \hat{p}_t| \tag{2.20}$$

where $MCMAE_i$ is the MCMAE for repetition i of a given filter.

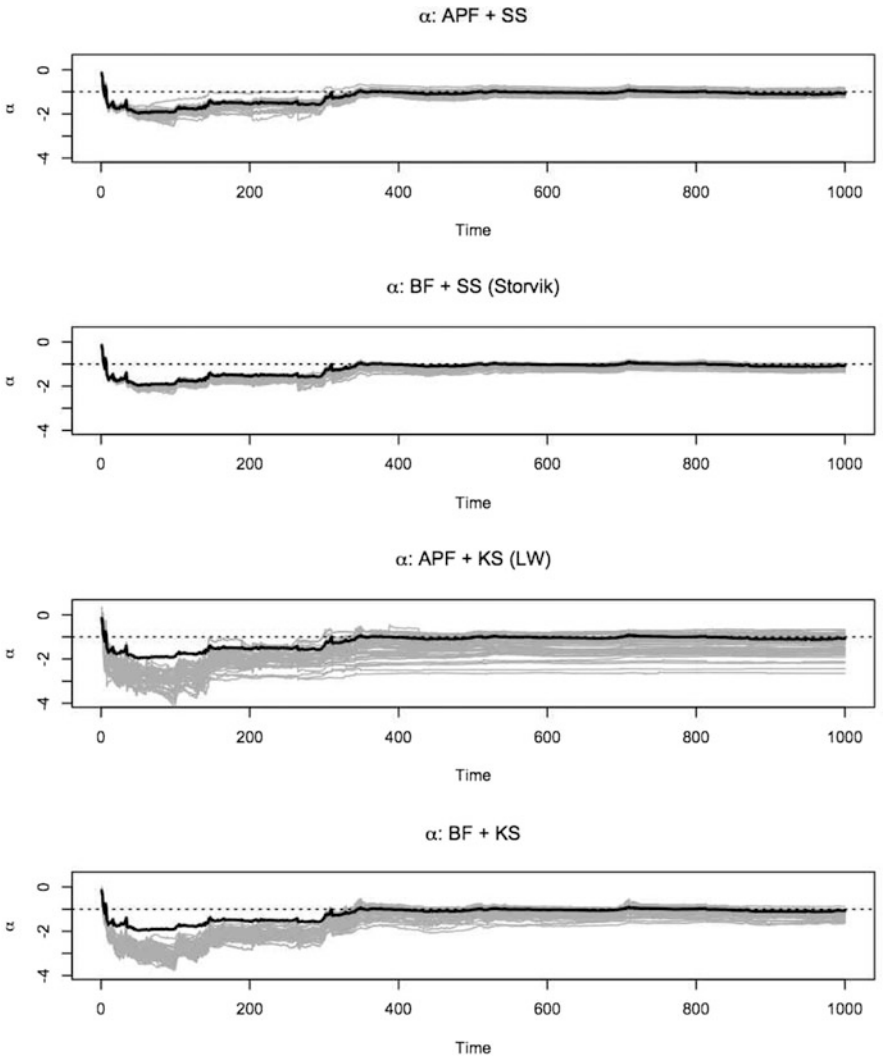

Fig. 2.6 Estimate paths for α in the MSSV $k = 1$ model for the 50 repetitions and the *exact estimation path*. The *solid black line* presents the *exact estimation path*, the *gray lines* are each one of the 50 repetitions of the filter, and the *dotted line* is the true parameter value.

of the kernel smoothing related filters, one appreciates that the LW filter is consistently more variable than the BF + KS filter. The largest deviations are significantly greater; in some cases the LW filter MCMAE is twice as much as the APF + SS' MCMAE. Once more, we see that on this dimension the LW filter appears to underperform the other filters.

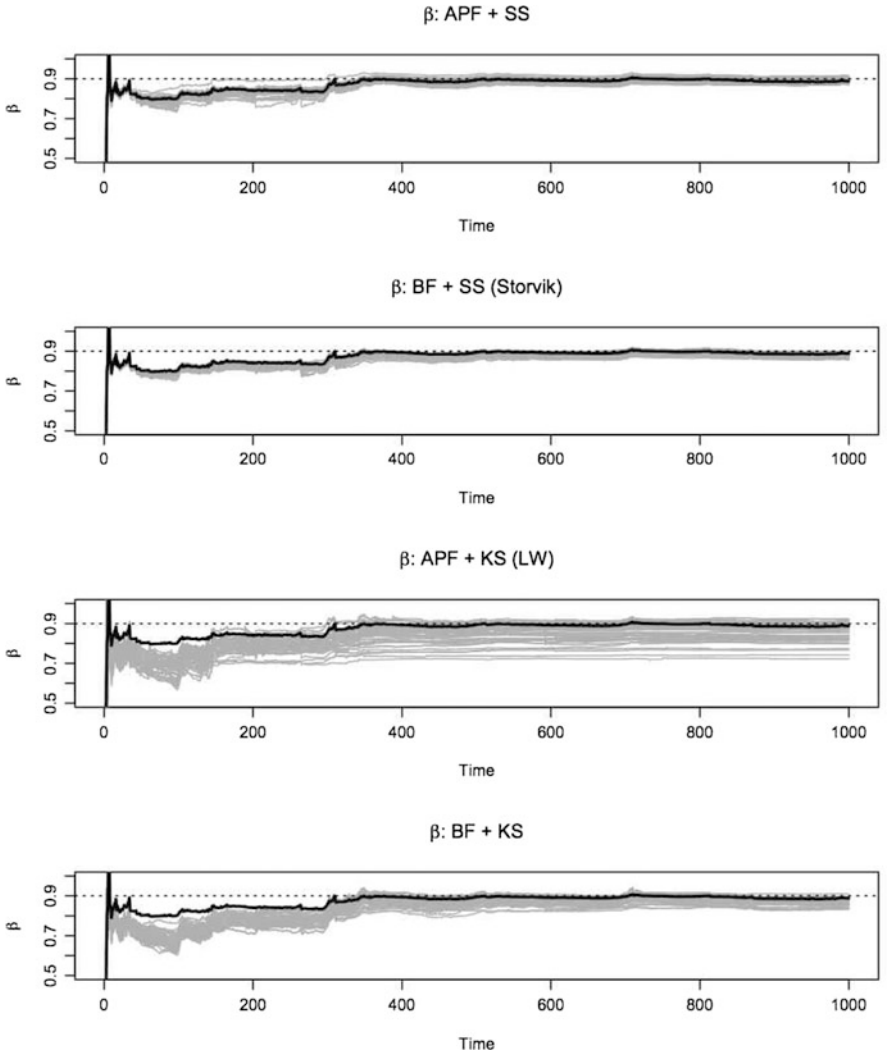

Fig. 2.7 Estimate paths for η in the MSSV $k = 1$ model for the 50 repetitions and the *exact estimation path*. The *solid black line* presents the *exact estimation path*, the *gray lines* are each one of the 50 repetitions of the filter, and the *dotted line* is the true parameter value.

2.3.3.4 Computational Time

The last dimension on which we compare the filters in the simulation study is the amount of time taken to complete one run. Table 2.5 presents the estimation times in seconds, averaged across runs[10] for the four filters and the two models of interest.

[10] The MSSV with $k = 1$ model runs and the MSSV with $k = 2$ model runs were implemented in different machines. In this chapter, the MSSV $k = 2$ was run in a more powerful computer.

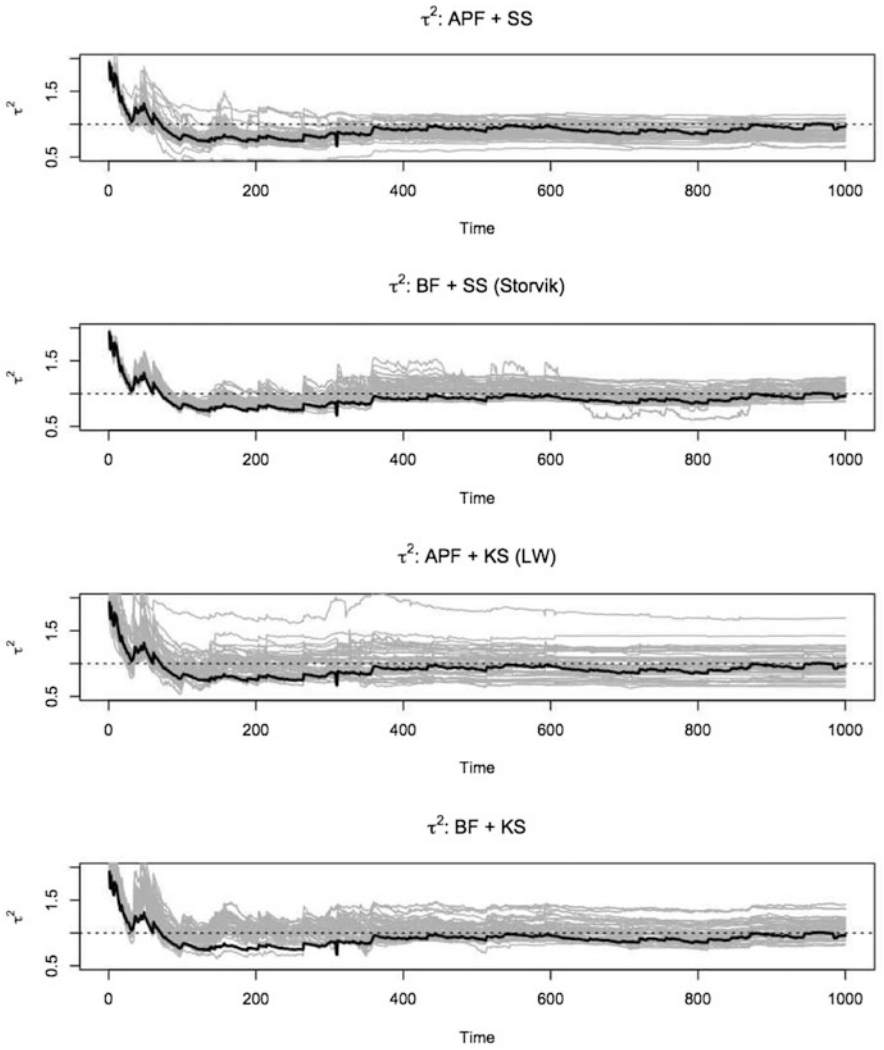

Fig. 2.8 Estimate paths for τ^2 in the MSSV $k = 1$ model for the 50 repetitions and the *exact estimation path*. The *solid black line* presents the *exact estimation path*, the *gray lines* are each one of the 50 repetitions of the filter, and the *dotted line* is the true parameter value.

The filters that implement an SS approach to parameter estimation take significantly more time. The reason for this is the complexity of the operations needed to update the sufficient statistics. Furthermore, we run all of our simulations in R which is known to struggle with loop calculations, which are, unfortunately, unavoidable in the SS updating. Operations that parameter estimation requires in the kernel smoothing technique are considerably simpler, making the LW and BF + KS

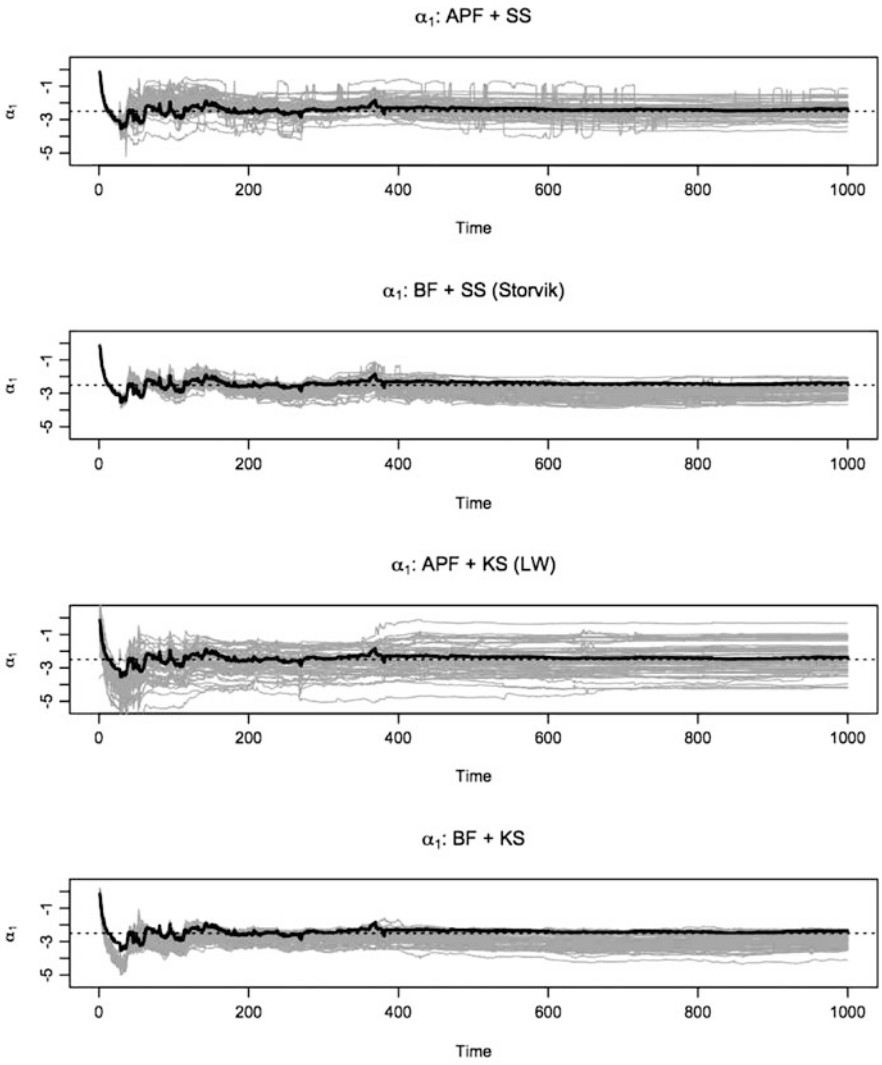

Fig. 2.9 Estimate paths for α_1 in the MSSV $k = 2$ model for the 50 repetitions and the *exact estimation path*. The *solid black line* presents the *exact estimation path*, the *gray lines* are each one of the 50 repetitions of the filter, and the *dotted line* is the true parameter value.

filters much more efficient in computation time terms. Unlike the other dimensions that we have explored so far, the LW filter is one of the filters that outperforms.

There appears to be an interesting trade-off between accuracy and computation time. The more accurate filters appear take longer to estimate. Therefore the question is how much accuracy you are willing to give up for a faster estimation. Another perspective from which this issue can be analyzed is how many particles I will implement. Accuracy and time are closely related to the amount of particles used.

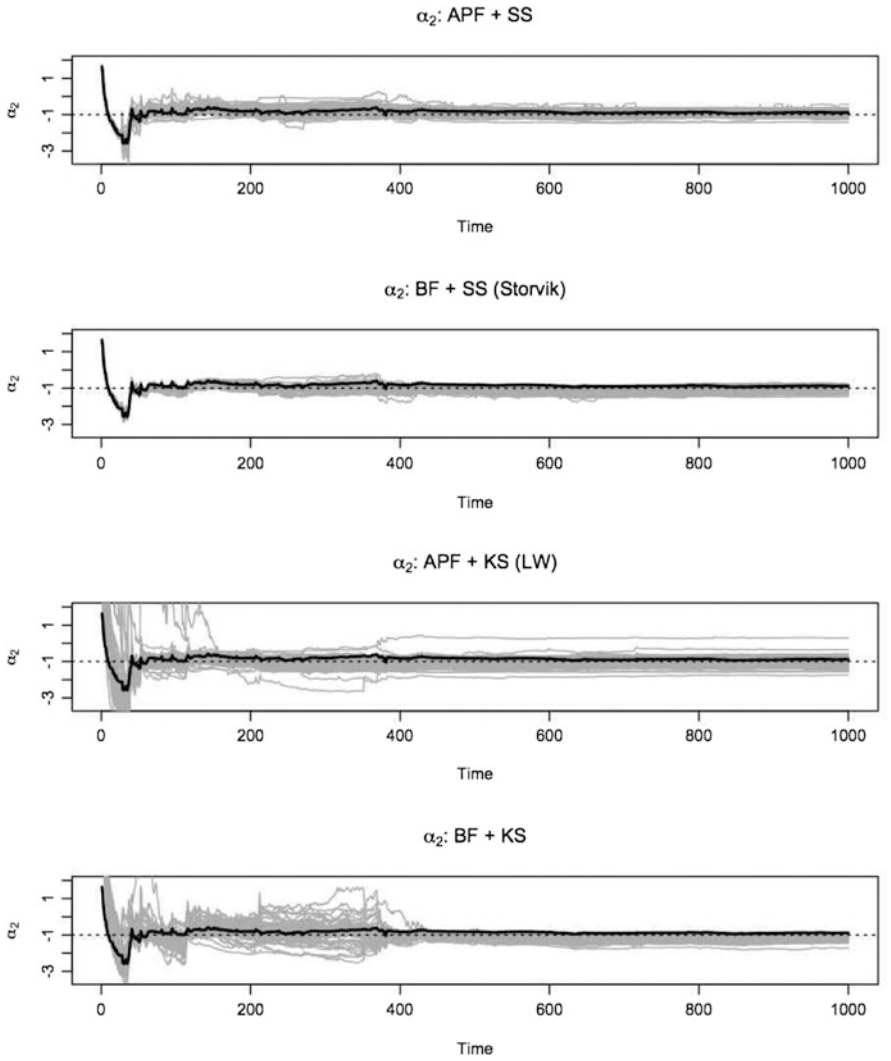

Fig. 2.10 Estimate paths for α_2 in the MSSV $k = 2$ model for the 50 repetitions and the *exact estimation path*. The *solid black line* presents the *exact estimation path*, the *gray lines* are each one of the 50 repetitions of the filter, and the *dotted line* is the true parameter value.

A good compromise to getting faster more accurate estimation could be to increase the number of particles in the LW filter estimation. Another option is to use less particles in an APF + SS filter. Preliminary analysis shows that this filter produces accurate estimates with a smaller number of particles. Details on how the increases in particles would affect accuracy and estimation time is beyond the scope of this chapter.

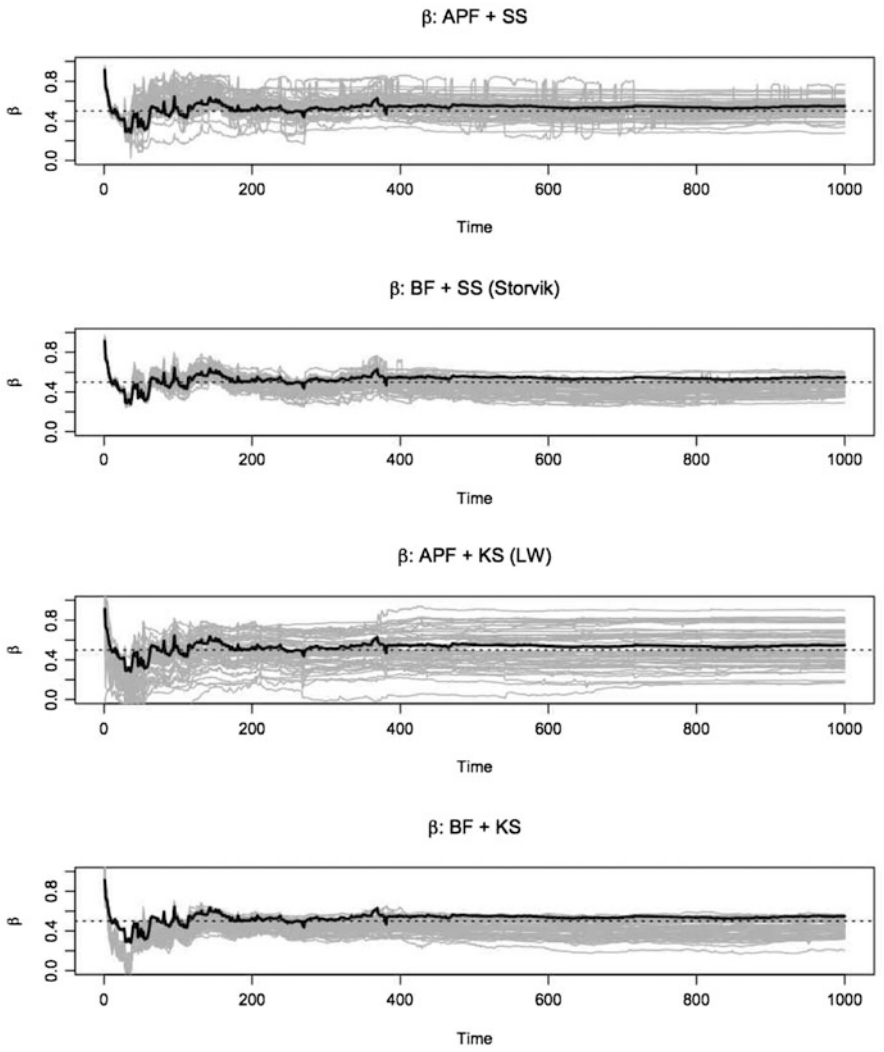

Fig. 2.11 Estimate paths for η in the MSSV $k = 2$ model for the 50 repetitions and the *exact estimation path*. The *solid black line* presents the *exact estimation path*, the *gray lines* are each one of the 50 repetitions of the filter, and the *dotted line* is the true parameter value.

2.3.4 Economic Insight

Exploring in more detail the estimates we observe that they bring interesting economic insight. At every point in time, the Bayesian nature of the results allows to infer information about the distribution of the estimates.

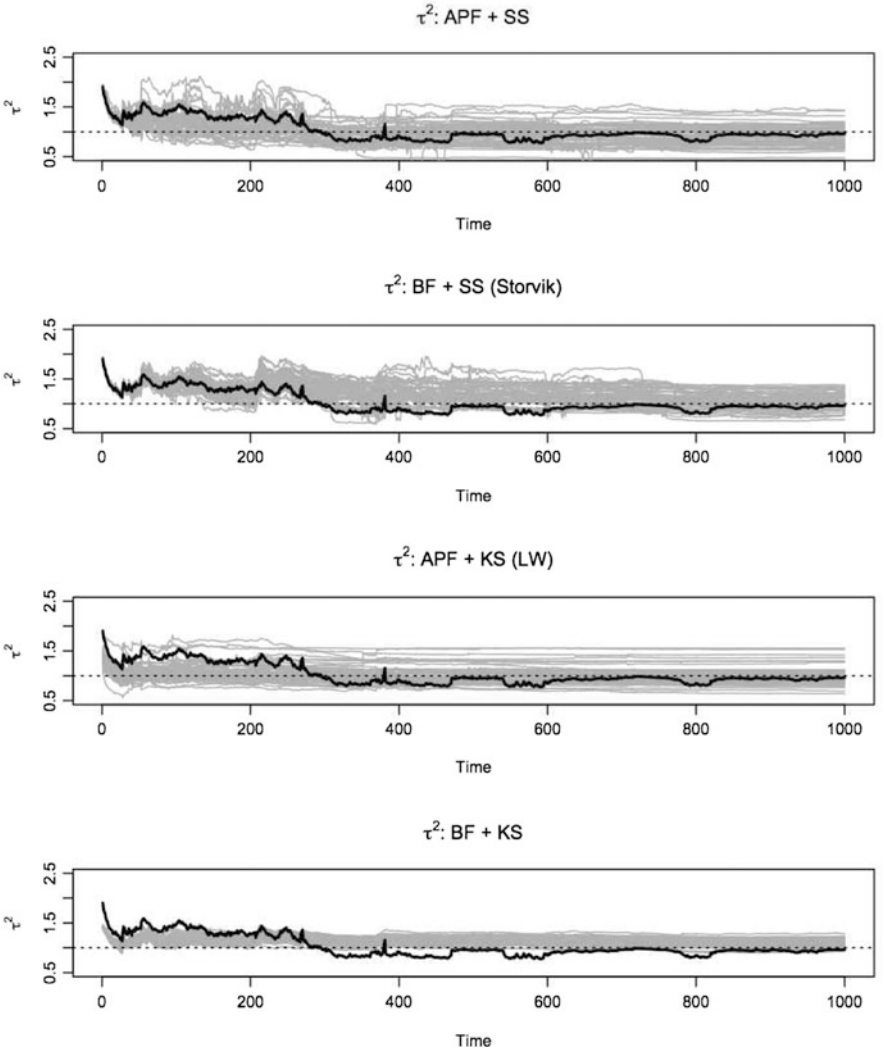

Fig. 2.12 Estimate paths for τ^2 in the MSSV $k = 2$ model for the 50 repetitions and the *exact estimation path*. The *solid black line* presents the *exact estimation path*, the *gray lines* are each one of the 50 repetitions of the filter, and the *dotted line* is the true parameter value.

Using the exact path estimates we can provide the economic interpretation based on the posterior distribution of the parameter. In particular, τ^2 provides interesting information about the volatility process in the two-state MSSV. Figure 2.15 shows that τ^2 moves along the volatility, i.e. when there are economy shifts between regimes. From the two panels in Fig. 2.15 one appreciates that when the volatility process shifts to a higher level state, the τ^2 estimates arguably also switch to a higher volatility of volatility level.

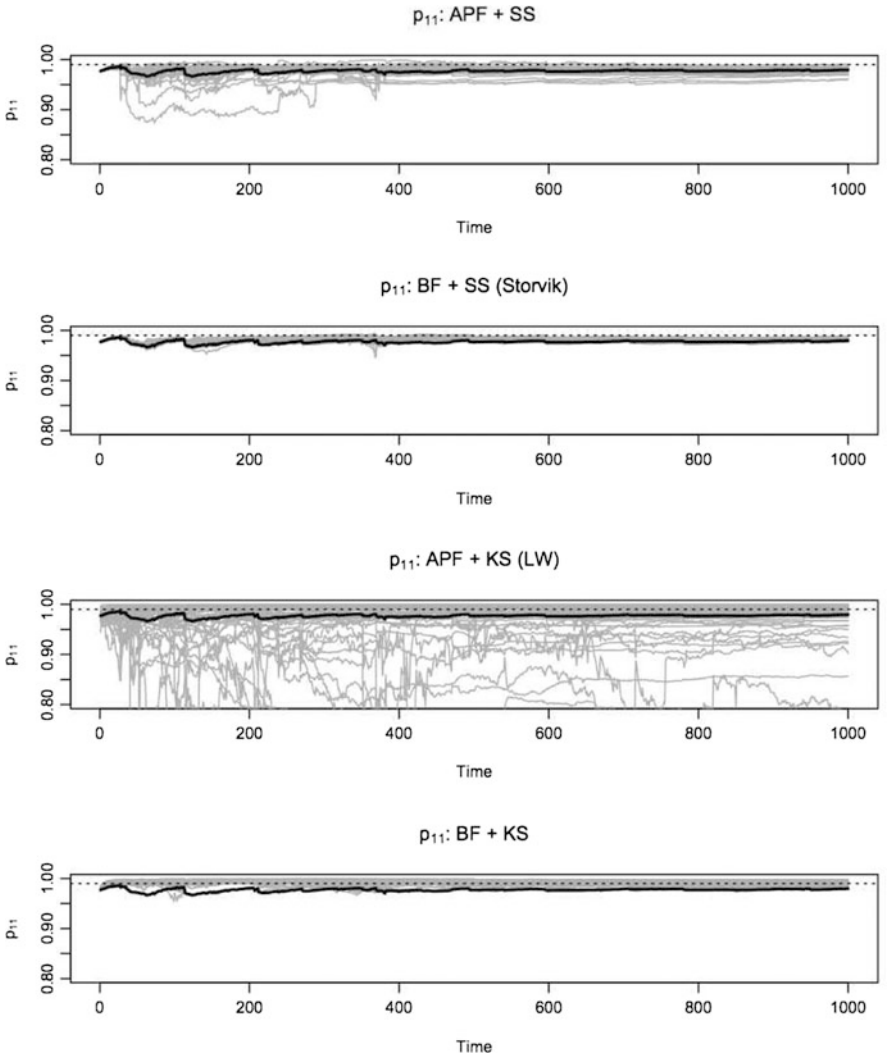

Fig. 2.13 Estimate paths for $p_{1,1}$ in the MSSV $k = 2$ model for the 50 repetitions and the *exact estimation path*. The *solid black line* presents the *exact estimation path*, the *gray lines* are each one of the 50 repetitions of the filter, and the *dotted line* is the true parameter value.

2.3.5 Robustness

Using the same parameter values as the ones discussed in Sect. 2.3.1, ten new data sets were simulated for the $k = 1$ and $k = 2$ cases. Each new data sets was estimated for ten runs. Results were then analyzed on the four dimensions presented above.

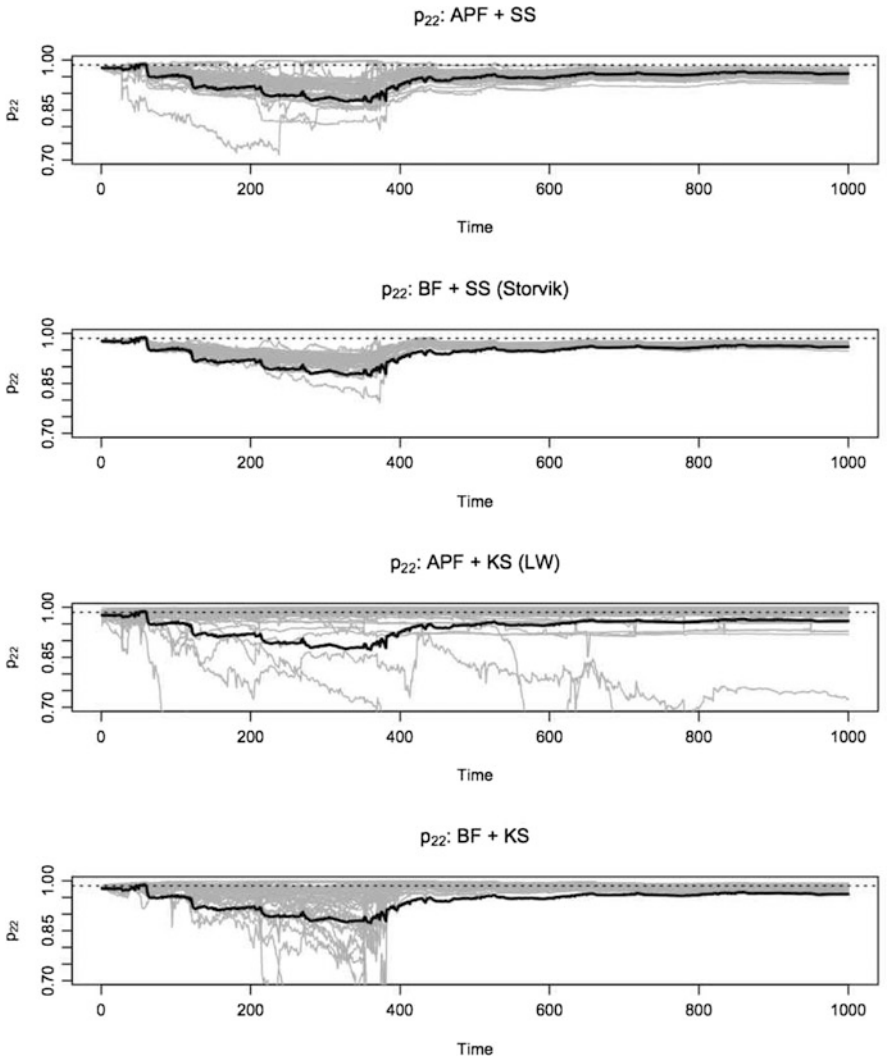

Fig. 2.14 Estimate paths for $p_{2,2}$ in the MSSV $k = 2$ model for the 50 repetitions and the *exact estimation path*. The *solid black line* presents the *exact estimation path*, the *gray lines* are each one of the 50 repetitions of the filter, and the *dotted line* is the true parameter value.

A detailed exploration of the ten data sets and ten runs for each one of the models revealed findings consistent with the ones previously discussed. All the results presented in Sects. 2.3.3.1–2.3.3.4 are robust to the data set chosen.

We see that in the additional tested cases, the APF + SS filter is the filter that appears to outperform the other filters. Likewise, we observed that the LW filter continues to have the same shortcomings. It has collapsing parameters, has the largest

Table 2.4 Mean MCMAE between the *exact path* and the estimated path for the parameters of interest in the MSSV models where $k = 1, 2$. The MCMAE are averaged across the 50 repetitions of each filter.

k	Parameter	APF + SS	BF + SS	BF + KS	LW
1	α	0.102	0.135	0.381	0.508
	η	0.0109	0.0136	0.0396	0.0544
	τ^2	0.0959	0.123	0.845	0.921
2	α_1	0.391	0.472	0.567	0.731
	α_2	0.172	0.253	0.392	0.568
	η	0.0806	0.0928	0.1170	0.1550
	τ^2	0.185	0.225	0.207	0.199
	p_{11}	0.00627	0.00572	0.207	0.08460
	p_{22}	0.0185	0.0177	0.0364	0.0603

Table 2.5 Mean time in seconds taken to estimate the MSSV models with $K = 1$ and 2 by the four different filtering strategies. Computational times are averaged across the 50 filter repetitions.

k	APF + SS	BF + SS	BF + KS	LW
1	5429.02	5447.22	31.58	39.15
2	4432.43	4244.95	159.62	220.55

Monte Carlo error, and has the biggest discrepancies when capturing the regime switches. The latter evidence reassures our previous findings on the applicability and accuracy of the four filters of interest.

2.4 Analysis and Results: Real Data Applications

In the second part of the analysis we use two equity indices and analyze their volatility processes using the outperforming filter, that is the APF + SS filter. The first is the IBOVESPA index[11] which is presented in order to replicate the results presented in Carvalho and Lopes (2007). The second series that we use is the S&P 500 index, where we explore a short and a long series allowing us to explore and highlight more properties of the Bayesian filtering estimation techniques, and the APF + SS filter in particular.

All data analyzed here was obtained from *Bloomberg* using the last price as a proxy for the day's trading price and only including data from days where trading took place. Table 2.6 presents summary statistics of the three analyzed series.

[11] IBOVESPA is an index of about 50 stocks that are traded on the Sao Paulo Stock, Mercantile and Futures Exchange (BOVESPA).

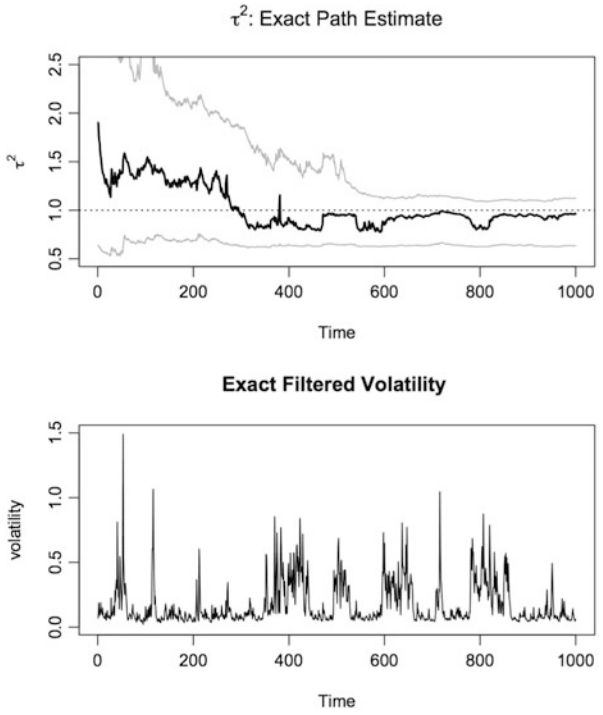

Fig. 2.15 *Top panel* shows the *exact parameter estimate* for τ^2. *Bottom panel* shows the *exact volatility estimate* for the MSSV $k = 2$ model.

Table 2.6 Summary statistics for the three real data sets used in the applications in Sect. 2.4.

Series	Start date	End date	Obs	Mean	SD	Min	Max	Kurtosis	Skewness
IBOVESPA	01/02/1997	01/16/2001	1000	0.000878	0.0294	−0.172	0.288	15.920	0.597
S&P 500 (short)	9/1/2006	8/31/2011	1258	−5.791e−05	0.0163	−0.0947	0.110	10.665	−0.280
S&P 500 (long)	9/1/1971	8/31/2011	10094	0.00024	0.011	−0.229	0.110	29.272	−1.075

2.4.1 IBOVESPA

We implement the proposed filter on the IBOVESPA stock index from 01/02/1997 to 01/16/2001. Figure 2.16 shows the log returns of the index from 01/02/1997 to 01/16/2001 (1,000 observations). This period includes a set of currency crises, such as the Asian crisis in 1997, the Russian crisis in 1998, and the Brazilian crisis in 1999, all of which directly affected emerging countries, like Brazil, generating high levels of uncertainty in the markets and consequently high levels of volatility.

Starting with priors similar to the ones used in Carvalho & Lopes (2007) we estimate the parameters both in the log-linear stochastic volatility and in the two-state MSSV. Model selection was done using Bayes factors that revealed strong

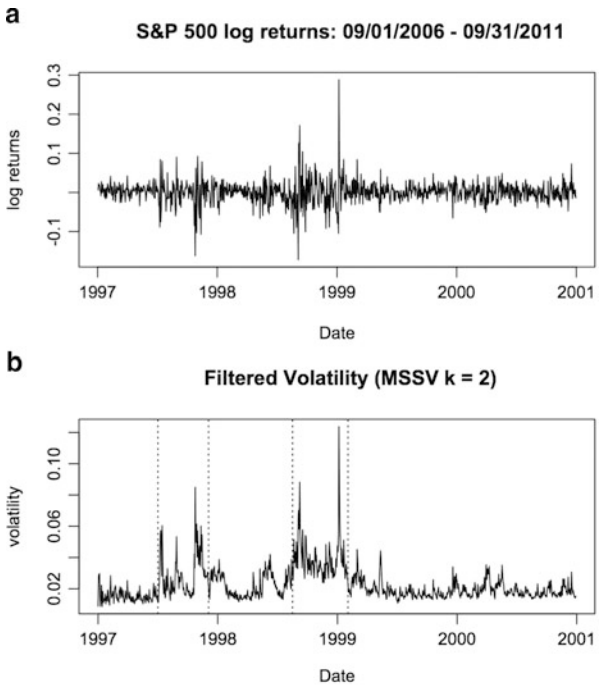

Fig. 2.16 Panel A: log returns for the IBOVESPA stock index between 01/02/1997 and 01/16/2001 for a total of 1,000 observations. Panel B: volatility estimates for the IBOVESPA fitting an MSSV $k = 2$ model. *Dotted lines* highlight the following dates: (1) 07/02/1997: Thailand devalues the Baht by as much as 20 %; (2)12/02/1997: IMF and South Korea set a bailout agreement; (3) 08/19/1998: Russia officially falls into default; 02/02/1999: Arminio Fraga is named President of Brazil's Central Bank.

evidence in favor of the two-state MSSV (log Bayes factor of 10.579). The estimated volatility times series is presented in panel B of Fig. 2.16 where one appreciates, like in Carvalho and Lopes (2007), the structural changes that result in periods of higher volatility. The dotted lines in plot highlight four key dates[12] mentioned in Carvalho and Lopes (2007), which appear to coincide with the regime switches in the model and that agents perceive as moments that started or ended a crisis. This

[12] The following key dates are highlighted in Fig. 2.16 panel B that are presented in Carvalho and Lopes (2007)

(a) 07/02/1997: Thailand devalues the Baht by as much as 20 %
(b) 12/02/1997: IMF and South Korea set a bailout agreement
(c) 08/19/1998: Russia officially falls into default
(d) 02/02/1999: Arminio Fraga is named President of Brazil's Central Bank

behavior matches the one found in Carvalho and Lopes (2007) and corroborates the fact that the sequential estimation is able to identify the structural changes through the discrete state prediction.

Our fixed parameter estimates present behavior similar to the findings by Carvalho and Lopes (2007). Again, the persistence parameter η is not overestimated. The discrete shifts in volatility level the posterior mean for η is no longer close to once. Likewise, the diagonal elements of the transitions probability matrix for the discrete states are high. In our estimation we obtained[13] $E(p_{11}|D_t) = 0.994$ and $E(p_{22}|D_t) = 0.974$, matching the observation of Carvalho and Lopes (2007) that the duration in each regime is quite long with a predominance of the low volatility regime.

2.4.2 S&P 500

As mentioned before, for the S&P 500 exploration we look at two series: (1) a short, 5-year series between 9/1/2006 and 8/31/2011 and (2) a long 40-year series between 9/1/1971 and 8/31/2011. The shorter series is used as second example of the applicability of the filters and for this one we do a backtest to highlight the accuracy of the estimates in a real data setting. The longer series is used to highlight the benefits of using the APF + SS in more extensive data sets.

Both the log linear stochastic volatility model and the two-state MSSV model are implemented for the two time series. The filters are run using uninformative priors with the same hyperparameter values as the ones implemented in the simulation study. Again a Bayes factor will be used to determine which model better fits the data.

2.4.2.1 Short Time Series

The shorter S&P 500 series that we explore in this chapter is summarized graphically in Fig. 2.17 top panel. Here, the reader can appreciate that at the end of 2008 and in the middle of 2011 there are periods with particularly high variability in the returns. This is consistent with the economic climate, as they coincide with the timing of the Lehman Brothers bankruptcy and the European credit crisis, respectively; 5,000 particle APF + SS filters are run for the two stochastic volatility models. Latent state and parameter estimates are obtained fitting $k = 1, 2$ MSSV models. Evaluating the results, we obtained a *log Bayes* factor of 15.52 in favor of the two-state MSSV model, implying that there are periods of time where the substantial and

[13] $E(.|D_t)$ is the conditional expected value given the data up to time t.

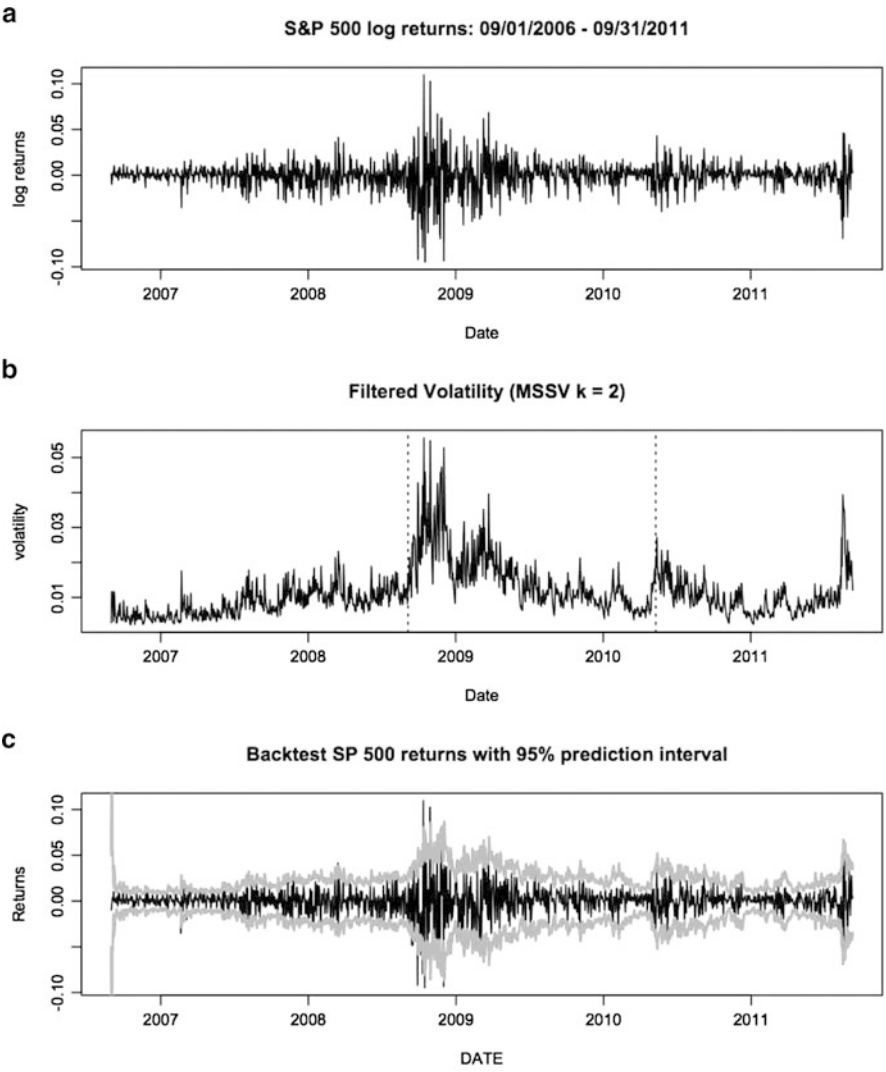

Fig. 2.17 Panel A shows the log returns for the S&P 500 stock index between 09/01/2006 and 08/31/2011. Panel B shows the volatility estimate obtained fitting an MSSV with $k = 2$. Panel C presents a graphical summary of the MSSV $k = 2$ estimates backtest. The *black line* is the true return process and the *dark gray lines* are the 95 % predictive intervals.

sustained volatility increases. Panel B of Fig. 2.17 shows the $k = 2$ MSSV volatility estimates. From the volatility estimates, it is somewhat clear when regime-shifting takes place, which in turn match periods of increased volatility. Dotted lines in the volatility estimate plots highlight an important date around the Lehman Brothers

Bankruptcy filing (September 7, 2008[14]) and a key date on the European credit crises (May 10, 2010[15]).

The sequential strategy used here is able to capture the structural changes in the volatility, by accurately identifying the moments of higher volatility through the discrete state prediction. The diagonal elements of the transition probability matrix for the discrete states are estimated to be high, with $E(p_{1,1}|D_t) = 0.981$ and $E(p_{2,2}|D_t) = 0.994$. This implies that the duration in each regime is quite long which is a fact also encountered by So et al. (1998) when analyzing the US S&P 500 series.

Further assessment of the parameters reveals, as expected, that there is no evidence of parameter degeneracy. Point estimates and 95 % credible intervals for the components of the Θ_{MSSV} are presented in Table 2.7. To test the accuracy of the parameter estimates we run a backtest of the data. Panel C of Fig. 2.17 presents the 95 % predictive intervals compared to the real returns. To test the accuracy of the predictions, we calculate the percentage of observations that lie outside the predictive interval. In this case we have 9.308 % of the observations outside the interval. This means that the estimates are underestimating the volatility. This is consistent with the observations of the simulation study that revealed that all filtering strategies had limitations with correctly estimating the latent process when there were sudden large increases.

Table 2.7 APF + SS fixed parameter estimates and 95 % credible intervals for the $k = 2$ MSSV models fitted to the short S&P 500 data set. $E(.|D_t)$ is the conditional expected value given the data up to time t.

Parameter	95 % credible interval	$E(.\|D_t)$
α_1	(−2.733, −1.892)	−2.313
α_2	(−2.289, −1.609)	−1.948
η	(0.734 , 0.814)	0.775
τ^2	(0.493, 0.656)	0.592
p_{11}	(0.961, 0.994)	0.982
p_{22}	(0.982, 0.998)	0.994

[14] The Federal Housing Finance Agency (FHFA) places Fannie Mae and Freddie Mac in government conservatorship. The U.S. Treasury Department announces three additional measures to complement the FHFAs decision: (1) preferred stock purchase agreements between the Treasury/FHFA and Fannie Mae and Freddie Mac to ensure the GSEs positive net worth; (2) a new secured lending facility which will be available to Fannie Mae, Freddie Mac, and the Federal Home Loan Banks; and (3) a temporary program to purchase GSE MBS.

[15] The Federal Housing Finance Agency (FHFA) places Fannie Mae and Freddie Mac in government conservatorship. The U.S. Treasury Department announces three additional measures to complement the FHFAs decision: (1) Preferred stock purchase agreements between the Treasury/FHFA and Fannie Mae and Freddie Mac to ensure the GSEs positive net worth; (2) a new secured lending facility which will be available to Fannie Mae, Freddie Mac, and the Federal Home Loan Banks; and (3) a temporary program to purchase GSE MBS.

2.4.2.2 Long Time Series

To further assess the applicability of the APF + SS filter we implemented the filter on a significantly longer data set. We use a 40-year or 10,094 observation data set to explore the performance of the filter of interest on longer time series, as the likelihood of parameter collapses increases with the length of the data. We implemented 5,000 and 10,000 particle APF + SS filters and evaluated the results. In both cases we observed that the filter performed well with no collapses or parameter degeneracy present in either setting which corroborates the wide applicability scope of the APF + SS filter.

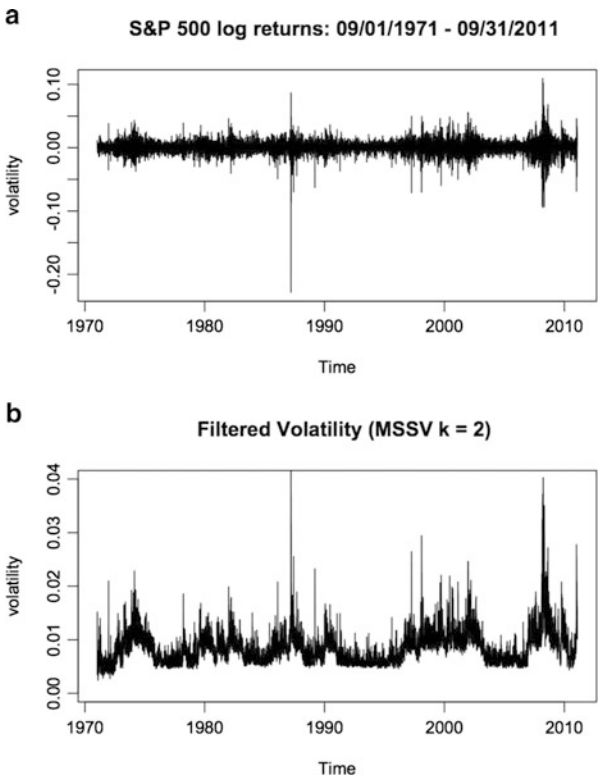

Fig. 2.18 Panel A shows the log returns for the S&P 500 stock index between 09/01/1971 and 08/31/2011. Panel B shows the volatility estimate obtained fitting an MSSV with $k = 2$.

Here, we present a summary of the 10,000 particle estimation, given that results are very similar regardless of the amount of particles used. Figure 2.18 panel A presents the log returns of the S&P 500 in the analyzed period. After fitting the $k = 1, 2$ MSSV models, a Bayes factor analysis revealed strong evidence supporting the two-state MSSV. Panel B of Fig. 2.18 shows the volatility estimates for this model. Again, it is interesting to observe how in the return and volatility plots the

periods of high variability match. Once more, this proves that the sequential learning characteristic of the filter allows to detect the structural changes in the data through the discrete state prediction.

A summary of the parameter estimates and 95 % credible intervals is presented in Table 2.8. Like in the short series, the diagonal elements in the transition probabilities are quite high, again being consistent with the So *et al.* (1998) findings. Furthermore, by allowing discrete shifts in the volatility level the posterior mean for η is no longer close to one which means that the persistence parameter is not being overestimated as discussed in Carvalho and Lopes (2007).

Table 2.8 APF + SS fixed parameter estimates and 95 % credible intervals for the $k = 2$ MSSV models fitted to the long S&P 500 data set. $E(.|D_t)$ is the conditional expected value given the data up to time t.

Parameter	95 % credible interval	$E(.\|D_t)$
α_1	$(-4.488, -4.151)$	-4.321
α_2	$(-4.013, -3.710)$	-3.863
η	(0.559, 0.592)	0.576
τ^2	(0.410, 0.435)	0.422
p_{11}	(0.997, 0.999)	0.998
p_{22}	(0.997, 0.999)	0.998

2.5 Conclusions

Our main contribution is to extend the APF and BF filters to accommodate sequential parameter learning via conditional sufficient statistics. We showed that, among a group of four filters, the APF + SS filter outperforms while the LW filter underperforms for the standard MSSV models. Our APF + SS filter also avoids or at least reduces the dependence of the PF on the somewhat arbitrariness of selecting a shrinkage/smoothness parameter. This is particularly important when dealing with variance parameters, such as the volatility of the volatility parameter in the SV and MSSV models.

The simulation study highlighted some of the important shortcomings in the LW filter. As expected, there is strong evidence of collapses in the parameter estimates and high Monte Carlo error. The only dimension in which this filter appeared to be efficient was in computational time. On the other hand, the APF + SS filter produces accurate parameter estimates that have the lowest Monte Carlo variability and show no evidence of particle degeneracy. Furthermore, for the two-state MSSV this filter is able to correctly track the regime changes. Nonetheless, given the complexity of the calculation of the conditional sufficient statistics, the APF + SS filter is not efficient in terms of computational time. As such we believe that compared to the other filters considered for analysis here the APF + SS the arguably the best.

Using real data examples we confirmed the applicability of the Bayesian filtering strategies presented here. We were able to fit and estimate the MSSV models with the APF + SS filter for both long and short return series. Results strongly suggested fitting two-state MSSV model in the three analyzed cases. For the IBOVESPA and S&P 500 analyses regime shifts matched key economic dates linked both to rises and decreases in market volatility.

Despite our concentration on stochastic volatility models, the APF + SS filter can be useful for many other statistical problems, such as (dynamic) factor models, space–time models, hierarchical models and several classes of time-series models, to name but a few. We believe that the flexibility of the APF + SS filter should be combined with careful (and potentially optimal) choice of resample-propagate proposal distributions (see Carvalho *et al.*, 2010) for efficient sequential learning in large and/or more complex dynamic systems.

One of the drawbacks of the SS estimation technique is that its application is highly dependent on fixed parameters admitting recursive conditional sufficient statistics. To overcome this a hybrid parameter learning technique can be implemented in situations where the vector of fixed parameters can be divided into two vector components, θ and η, where η, conditional on θ, admits recursive conditional sufficient statistics. More specifically, the prior for (θ, η) is

$$p(\theta, \eta) = p(\theta)p(\eta|x_0, \theta) = p(\theta)p(\eta|s_0, \theta) \tag{2.21}$$

and $s_t = \mathscr{S}(s_{t-1}, x_t, \theta, y_t)$ is a vector of recursive sufficient statistics, for $t = 1, \ldots, n$.

References

1. Bodart, V., Kholodilin, K., Shadman-Mehta, F.: Identifying and forecasting the turning points of the Belgian business cycle with regime-switching and logit models. Working Paper, Universite catholique de Louvain, 2005
2. Bruno, G., Otranto, E.: Models to date the business cycle: The Italian case. Economic Modelling, **25**,899–911, 2008
3. Carvalho, C.M., Johannes, M., Lopes, H.F., Polson, N.G.: Particle learning and smoothing. Statistical Science, **25**, 88–106, 2010
4. Carvalho, C.M., Lopes, H.F: Simulation-based sequential analysis of Markov switching stochastic volatility models. Computational Statistics & Data Analysis, **51**, 4526–4542, 2007
5. Diebold, F.X., Rudebusch, G.D.: Scoring the Leading Indicators, The Journal of Business, University of Chicago Press, **62**(3), 369–91, 1989
6. Doucet, A., de Freitas, J., and Gordon, N. (eds.): Sequential Monte Carlo Methods in Practice. Springer, New York, 2001
7. Doucet, A., Johansen,A. M.: A tutorial on particle filtering and smoothing: Fifteen years later. In: Crisan, D., Rozovsky, B. (eds.) Handbook of Nonlinear Filtering, Oxford University Press, 2010
8. Eraker, B., Johannes, M., Polson, N.: The impact of jumps in volatility and returns. Journal of Finance, **58**, 1269–1300, 2003
9. Fearnhead, P.: Markov chain Monte Carlo, sufficient statistics, and particle filters. Journal of Computational and Graphical Statistics, **11**, 848–862, 2002
10. E. Ghysels, E., Harvey, A. C., Renault, E.: Stochastic volatility. In: Maddala, G.S., Rao, C.R. (eds) Handbook of Statistics, pp. 119–191. Amsterdam: North-Holland, 1996

11. Gordon, N., Salmond, D., Smith, A.F.M: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEEE Proceedings, **F-140**, 107–113, 1993
12. Jacquier, E., Polson, N.G., Rossi, P.E.: Bayesian analysis of stochastic volatility models. Journal of Business & Economic Statistics, **12**, 371–389, 1994
13. Kim, S., Shephard, N., Chib, S.: Stochastic volatility: Likelihood inference and comparison with ARCH models. Review, Econom. Stud., **65**, 361–393, 1998
14. Kong, A., Liu, J.S., W. Wong, W: Sequential imputation and Bayesian missing data problems. Journal of the American Statistical Association, **89**, 590–599, 1994
15. Liu, J., West, M.: Combined parameter and state estimation in simulation-based filtering. In: Doucet, A, de Freitas, J., Gordon, N.J. (eds.), Sequential Monte Carlo Methods in Practice. Springer-Verlag, New York, 2001
16. Lopes, H.F., Carvalho, C.M.: Online Bayesian learning in dynamic models: An illustrative introduction to particle methods. In: West, M.,Damien, P., Dellaportas, P., Polson, N.G., Stephens, D.A. (eds.) Bayesian Dynamic Modelling, Bayesian Inference and Markov Chain Monte Carlo: In Honour of Adrian Smith, Oxford University Press, 2011
17. Lopes H.F., Tsay, R.S.: Particle filters and Bayesian inference in financial econometrics. Journal of Forecasting, **30**, 168–209, 2011
18. Olsson, J., Cappé, O., Douc, R., Moulines, E.: Sequential Monte Carlo smoothing with application to parameter estimation in nonlinear state space models. Bernoulli, **14**, 155–179, 2008
19. Otranto, E.: The stock and Watson model with Markov switching dynamics: an application to the Italian business cycle. Statistica Applicata, **13**, 413–429, 2001
20. Pitt, M.K., Shephard, N., Filtering via simulation: Auxiliary particle filters. Journal of the American Statistical Association, **94**, 590–599, 1999
21. Polson, N.G., Stroud, J.R., Muller, P.: Practical filtering with sequential parameter learning. Journal of the Royal Statistical Society, Series B, **70**, 413–428, 2008
22. So, M., Lam, K., Li, W.K.: A stochastic volatility model with Markov switching. Journal of Business & Economic Statistics, **16**, 244–253, 1998
23. Storvik, G.: Particle filters in state space models with the presence of unknown static parameters. IEEE: Trans. of Signal Processing, **50**, 281–289, 2002
24. West, M.: Approximating posterior distributions by mixture. Journal of the Royal Statistical Society. Series B, **55**, 409–422, 1993

Chapter 3
A Survey of Implicit Particle Filters for Data Assimilation

Alexandre J. Chorin, Matthias Morzfeld, and Xuemin Tu

3.1 Introduction

In many problems in science and engineering, e.g. in statistics, statistical signal processing, oceanography, meteorology, geomagnetics, econometrics, or finance, one wants to identify the state of a system from an uncertain model supplemented by a stream of noisy and incomplete data (see, e.g., [12, 34] for recent reviews in economics). The model is typically a Markovian state space model (often a discretization of a stochastic differential equation, SDE) and describes the state sequence $\{\mathbf{x}^n;\ n \in N\}$, where $\mathbf{x}^n$ is a real, m-dimensional vector. For simplicity, we assume that the noise is additive, so that the model equations are

$$\mathbf{x}^n = f^n(\mathbf{x}^{n-1}) + \mathbf{v}^{n-1}, \tag{3.1}$$

where f^n is an m-dimensional vector function, and $\{\mathbf{v}^{n-1},\ n \in N\}$ is a sequence of independent identical distributed (i.i.d.) m-dimensional random vectors which, in many applications, are Gaussian vectors with independent components. One can think of the $\mathbf{x}^n$ as values of a process $\mathbf{x}(t)$ evaluated at times $n\delta$, where δ is a fixed time increment. The probability density function (pdf) of the initial state $\mathbf{x}^0$ is assumed to be known.

A.J. Chorin
Department of Mathematics, University of California and Lawrence, Berkeley National Laboratory, Berkeley, CA, USA
e-mail: chorin@math.berkeley.edu

M. Morzfeld
Lawrence Berkeley National Laboratory, Berkeley, CA, USA
e-mail: mmo@math.lbl.gov

X. Tu (✉)
Department of Mathematics, University of Kansas, Lawrence, KS, USA
e-mail: xtu@math.ku.edu

Y. Zeng and S. Wu (eds.), *State-Space Models: Applications in Economics and Finance*, Statistics and Econometrics for Finance 1, DOI 10.1007/978-1-4614-7789-1_3,

The model (3.1) is supplemented by an observation (or measurement) equation, which relates observations $\{\mathbf{b}^n;\ n \in N\}$, where $\mathbf{b}^n$ is a real, k-dimensional vector and $k \leq m$, to the states $\mathbf{x}^n$; we assume the observation equation is

$$\mathbf{b}^n = h^n(\mathbf{x}^n) + \mathbf{w}^n, \tag{3.2}$$

where h^n is a k-dimensional vector function, and $\{\mathbf{w}^n,\ n \in N\}$ is a k-dimensional i.i.d. process, independent of $\mathbf{v}^n$. The model and the observation equations together constitute a hidden Markov state space model (HMM). To streamline our notation, we denote the state and observation sequences up to time n by $\mathbf{x}^{0:n} = \{\mathbf{x}^0, \cdots, \mathbf{x}^n\}$ and $\mathbf{b}^{1:n} = \{\mathbf{b}^1, \cdots, \mathbf{b}^n\}$, respectively.

Our goal is to estimate the state sequence $\mathbf{x}^{0:n}$, based on (3.1) and (3.2) and we propose to use the minimum mean square error estimator $E\left[\mathbf{x}^{0:n}|\mathbf{b}^{1:n}\right]$ (see, e.g., [5]). If f^n and h^n are linear functions and if, in addition, $\mathbf{v}^n$ and $\mathbf{w}^n$ are Gaussian random variables, this conditional mean can be computed by the Kalman filter (KF) [22, 27, 28]. The ensemble Kalman filter (EnKF) [19] uses the KF formalism but updates the covariance matrix using an "ensemble of particles," i.e. by Monte Carlo simulations of the model (3.1). Because EnKF uses this ensemble approach, it can give good results even with nonlinear models (3.1), provided the nonlinearity is not too strong. Variational data assimilation methods find the mode of the target pdf, i.e. the most likely state given the data, and often use linearizations and Gaussian approximations to streamline the computations (see, e.g., [2, 10, 11, 33, 48–50, 56] and Sect. 3.4 for more details on KF, EnKF, and variational data assimilation).

In nonlinear, non-Gaussian situations, one can approximate the conditional mean using sequential Monte Carlo methods, called particle filters. Particle filters follow replicas of the system (called particles), whose empirical distribution weakly approximates the pdf $p(\mathbf{x}^{0:n} \mid \mathbf{b}^{1:n})$ (called the target density), and approximate the conditional mean by the weighted sample mean [1, 14, 16]. A standard particle filter, also called the sequential importance sampling with resampling (SIR) filter, first generates a set of particles $\{\mathbf{x}_i^n\}$ from the model equation (3.1) and then weighs these particles by the observation equation (3.2) [16, 17, 23]. The empirical distribution of the weighted particles forms a weak approximation of the target density at the current time step. One then removes particles with small weights (which contribute very little to the approximation of the target density) by "resampling," (see [1, 16] and the references therein for efficient resampling algorithms). The SIR filter is easy to implement, however after several time steps, often only a few of the particles carry a significant weight, which means that the weak approximation of the target density is poor. A cure here is to increase the number of particles (so that at least some carry a significant weight); however, it has been shown that the number of particles required can grow catastrophically with the state dimension m [3, 47]. Various strategies have been proposed to ameliorate this difficulty, and most of them focus on finding a better way to generate the samples [13, 16, 45, 51, 52, 54].

In what follows, we explain how implicit particle filters [7–9, 39–41] tackle this problem. The basic idea is to first look for regions of high probability in the target density and then focus the particles onto these regions, so that only particles

with significant weights are generated and the number of particles required remains manageable even if the state dimension is large. The high probability regions are identified by particle-by-particle minimizations, and the samples are obtained by solving data-dependent algebraic equations. The solutions of these equations define a high probability sample from the target density. Related work can be found in [42, 53].

The remainder of this chapter is organized as follows. In Sect. 3.2 we present the mathematical formulation of implicit particle filters and highlight special cases of interests. Several implementations of implicit particle filters are discussed in Sect. 3.3. In Sect. 3.4, the relations with other data assimilation methods are studied. We present six examples in Sect. 3.5 to demonstrate the efficiency and broad applicability of the implicit particle filter. Conclusions are offered in Sect. 3.6.

3.2 Implicit Particle Filters

The implicit particle filter is a sequential Monte Carlo method for data assimilation that uses importance sampling. In importance sampling one wants to find a weak approximation of the pdf, f, of a continuous random variable (called the target pdf), by generating weighted samples from a known density f_0 (called the importance function), see, e.g., [5]. The weight of the sample X_j (obtained by sampling f_0),

$$w(X_j) = \frac{f(X_j)}{f_0(X_j)},$$

is the ratio of the target pdf and the importance function. The N weighted samples X_j, $j = 1,\ldots,N$, with their weights w_j normalized so that their sum equals 1, form a weak approximation of the target pdf f such that

$$E_f[g(x)] = \int_{-\infty}^{\infty} g(x)f(x)dx \approx \sum_{j=1}^{N} g(X_j)w(X_j),$$

for all sufficiently smooth, scalar functions g, where $E_f[g(x)]$ denotes the expected value of the function g with respect to the pdf f. The key to making importance sampling efficient is choosing a suitable importance function f_0, such that the weights vary little from one sample to the next. In data assimilation, the target density is the conditional pdf $p(\mathbf{x}^{0:n} \mid \mathbf{b}^{1:n})$. We now present the importance function generated by the implicit particle filter and describe why it produces samples with a small variance in the weights.

For simplicity of presentation, we assume that the model equation (3.1) is synchronized with the observations (3.2), i.e. observations $\mathbf{b}^n$ are available at every model step (it is not hard to drop this assumption, see Sect. 3.2.2). Using Bayes' rule and the Markovian property of the model, we obtain the recursive expression:

$$p(\mathbf{x}^{0:n+1} \mid \mathbf{b}^{1:n+1}) = p(\mathbf{x}^{0:n} \mid \mathbf{b}^{1:n})p(\mathbf{x}^{n+1} \mid \mathbf{x}^n)p(\mathbf{b}^{n+1} \mid \mathbf{x}^{n+1})/p(\mathbf{b}^{n+1} \mid \mathbf{b}^{1:n}). \quad (3.3)$$

At the current time $n+1$, the first term in (3.3) can be assumed to be known, because it is the result of our calculations at time n. The denominator is common to all particles and thus drops out in the importance sampling scheme (where the weights are normalized so that their sum equals 1).

Suppose that at time n, we have M samples $\mathbf{X}_j = \mathbf{X}_j^{0:n}$ with weights w_j^n, $j = 1,\ldots,M$, whose empirical distribution weakly approximates $p(\mathbf{x}^{0:n} \mid \mathbf{b}^{1:n})$. For each sample (particle), define a function F_j by

$$F_j(\mathbf{X}_j^{n+1}) = -\log\left(p(\mathbf{X}_j^{n+1} \mid \mathbf{X}_j^n)p(\mathbf{b}^{n+1} \mid \mathbf{X}_j^{n+1})\right), \tag{3.4}$$

where we obtain the first term from the model equation (3.1) and the second from the observation equation (3.2). Note that the arguments of the functions F_j are the state variables of the jth particle at time $n+1$. The previous state of the jth particle, $\mathbf{X}_j^n$ and the current observation $\mathbf{b}^{n+1}$ are merely parameters.

By definition of the F_j's, the high probability region of the target density corresponds to the region around the global minimum of F_j. Thus, searching for the high probability region in the target density is equivalent to minimizing the functions F_j. We first assume that the functions F_j are convex and carry out this minimization with standard techniques (e.g., Newton's method, quasi-Newton methods, gradient descent, see [20, 43]) and obtain samples within these regions by solving the data-dependent equations

$$F_j(\mathbf{X}_j^{n+1}) - \phi_j = \frac{1}{2}\xi_j^T\xi_j, \tag{3.5}$$

where

$$\phi_j = \min F_j(\mathbf{X}_j^{n+1}),$$

and where ξ_j is a realization of an easy-to-sample Gaussian reference variable $\xi \sim N(0,I)$, where $N(\mu,\Sigma)$ denotes a Gaussian pdf with mean μ and covariance matrix Σ, and I is the m-dimensional identity matrix. A Gaussian reference variable is chosen for simplicity of presentation and is by no means necessary (it may be suboptimal in some applications). More importantly, a Gaussian reference variable does not imply a Gaussianity or linearity assumption. All that is needed here is a reference variable with a high probability close to the origin, so that (3.5) maps the high probability region of the reference density to the high probability region of the target pdf.

The solutions of (3.5) define the samples $\mathbf{X}_j^{n+1}$; however, the solutions are not unique because (3.5) connects the m-dimensional samples $\mathbf{X}_j^{n+1}$ to the m-dimensional reference variable ξ_j. The samples we find thus depend on the map $\xi_j \to \mathbf{X}_j^{n+1}$ we choose to solve (3.5). To obtain a high probability sample $\mathbf{X}_j^{n+1}$, we choose maps that satisfy the following conditions (see [7] for detailed explanation): the map should be (1) one-to-one and onto with probability one (so that the whole sample space is covered); (2) smooth near the high-probability region of ξ (so that the weights do not vary unduly from particle to particle); and (3) there

should be an easy way to evaluate the Jacobians $J = \left|\det\left(\partial \mathbf{X}_j^{n+1}/\partial \xi_j\right)\right|$ (for efficient implementation).

We will present specific choices for these maps in Sect. 3.3, but for now assume that we can compute these Jacobians. The probability of the sample we obtain is $p(\mathbf{X}_j^{n+1}) = p(\xi_j)/J$, so that, using (3.5) and (3.4), the weight of the sample can be computed to be

$$w_j^{n+1} \propto w_j^n \exp(-\phi_j) J. \tag{3.6}$$

With these M samples $\mathbf{X}_j^{n+1}$ and the samples $\mathbf{X}_j = \mathbf{X}_j^{0:n}$ from the previous time step, we can form a sample $\hat{\mathbf{X}}_j = \mathbf{X}_j^{0:n+1}$ from $p(\mathbf{x}^{0:n+1} \mid \mathbf{b}^{1:n+1})$ with weight w_j^{n+1} as in (3.6). Once one has samples and weights, one can resample the pdf they define so as to remove some of the low probability particles and reset all the weights to $1/M$, see, e.g., [1].

If the functions F_j are not convex, one can use the degeneracy of (3.5) to replace these functions by convex functions F_j^0 in (3.5) in such a way that the focusing effect is maintained; the weights have to be recomputed so that there is no bias, see, e.g., [7].

Before focusing our attention on implementing the implicit particle filter (see Sect. 3.3), we give more details on the functional form of F_j for some cases of interest.

3.2.1 Linear Observation Function and Gaussian Noise

To illustrate the method on a simple example, we assume that the observation equation is linear, i.e. $h^n(\mathbf{x}) = H\mathbf{x}$, where H is a $k \times m$ matrix, and that the noise processes $\mathbf{w}^n$ and $\mathbf{v}^n$ in (3.1) and (3.2) are Gaussian with zero mean and known covariance, i.e. $\mathbf{v}^n \sim N(\mathbf{0}, G)$, $\mathbf{w}^n \sim N(\mathbf{0}, Q)$, where G is an $m \times m$, real, symmetric positive definite (SPD) matrix and Q is a $k \times k$ SPD matrix. The functions F_j in (3.4) can now be written as

$$F_j(\mathbf{X}_j^{n+1}) = \frac{1}{2}\left(\mathbf{X}_j^{n+1} - \mu_j\right)^T \Sigma^{-1}\left(\mathbf{X}_j^{n+1} - \mu_j\right) + \phi_j, \tag{3.7}$$

where

$$\Sigma^{-1} = G^{-1} + H^T Q^{-1} H, \tag{3.8}$$

$$K = HGH^T + Q, \tag{3.9}$$

$$\mu_j = \Sigma\left(G^{-1} f^n(\mathbf{X}_j^n) + H^T Q^{-1} \mathbf{b}^{n+1}\right), \tag{3.10}$$

$$\phi_j = \frac{1}{2}\left(\mathbf{b}^{n+1} - Hf^n(\mathbf{X}_j^n)\right)^T K^{-1}\left(\mathbf{b}^{n+1} - Hf^n(\mathbf{X}_j^n)\right). \tag{3.11}$$

It is clear that $\phi_j = \min F_j(\mathbf{X}_j^{n+1})$, so that (3.5) becomes

$$\left(\mathbf{X}_j^{n+1} - \mu_j\right)^T \Sigma^{-1}\left(\mathbf{X}_j^{n+1} - \mu_j\right) = \xi_j^T \xi_j. \tag{3.12}$$

We can solve (3.12) by computing the Cholesky factorization $\Sigma = LL^T$ and putting

$$X_j^{n+1} = \mu_j + L\xi_j. \tag{3.13}$$

The Jacobian $J = |\det\left(\frac{\partial \mathbf{x}}{\partial \xi}\right)| = |\det L|$ is constant (the same for all particles) and thus need not be computed. By (3.6), the weight for each particle is

$$w_j^{n+1} \propto w_j^n \exp(-\phi_j). \tag{3.14}$$

Moreover, a simple calculation shows that

$$w_j^{n+1} \propto w_j^n p(\mathbf{b}^{n+1} \mid \mathbf{X}_j^n).$$

The above weights are the same as those of a filter that uses the optimal importance function $\hat{q} = p(\mathbf{x}^{n+1}|\mathbf{x}^n, \mathbf{b}^{n+1})$ (see [17] and the references therein) . "Optimal" here refers to "having minimum variance in the weights per particle," i.e. for a fixed $\mathbf{X}_j^n$, the variance of $\mathbf{w}_j^{n+1}$ is zero. The implicit particle filter produces minimum variance weights in this sense if the observation equation (3.2) is linear and the observations are in sync with the model (see Sect. 3.4.2 for more details on the optimal importance function).

3.2.2 Sparse Observations

The assumption that the observations are available at every model step can be relaxed. Let $r \geq 1$ be the number of model steps between observations (it is an easy exercise to adjust F_j for the case when the number of model steps between observations is not constant). The recursive formula (3.3) becomes

$$\begin{aligned} p(\mathbf{x}^{0:r(n+1)} \mid \mathbf{b}^{1:n+1}) = p(\mathbf{x}^{0:rn} \mid \mathbf{b}^{1:n}) p(\mathbf{x}^{rn+1} \mid \mathbf{x}^{rn}) \cdots p(\mathbf{x}^{r(n+1)} \mid \mathbf{x}^{r(n+1)-1}) \\ \times p(\mathbf{b}^{n+1} \mid \mathbf{x}^{r(n+1)}) / p(\mathbf{b}^{n+1} \mid \mathbf{b}^{1:n}). \end{aligned}$$

Again, suppose we have M weighted samples $\mathbf{X}_j = \mathbf{X}_j^{0:rn}$, $j = 1, \ldots, M$, from $p(\mathbf{x}^{0:rn} \mid \mathbf{b}^{1:n})$ and, for each sample, we define a function F_j by

$$\begin{aligned} F_j(\mathbf{X}_j^{rn+1}, \ldots, \mathbf{X}_j^{r(n+1)}) = -\log(p(\mathbf{X}_j^{rn+1} \mid \mathbf{X}_j^{rn}) \cdots p(\mathbf{X}_j^{r(n+1)} \mid \mathbf{X}_j^{r(n+1)-1}) \\ \times p(\mathbf{b}^{n+1} \mid \mathbf{X}_j^{r(n+1)})). \end{aligned}$$

With this F_j, one can follow the steps starting with (3.5) to obtain a sample from $p(\mathbf{x}^{0:r(n+1)} \mid \mathbf{b}^{1:n+1})$. Note that the functions F_j depend on rm variables (the components of $\mathbf{X}_j^{rn+1:r(n+1)}$), so that we need to choose an rm-dimensional reference density. However, the general procedure for generating samples does not change when the observations are sparsely available in time.

3.2.3 Models with Partial Noise

Following [40], we consider the case of "partial noise," i.e. the model noise, $\mathbf{v}^n \sim N(0,G)$, is Gaussian with singular covariance matrix G. Such models appear frequently, for example in the discretization of a stochastic partial differential equation (SPDE) driven by spatially smooth noise (see, e.g., [25, 35, 40]). Another class of models with partial noise are stochastic dynamical equations supplemented by conservation laws. There is typically zero uncertainty in the conservation laws (e.g., the conservation of mass), so that the model is subject to partial noise [33]. This situation is similar to that of second-order (in time) SDEs, that appear, for example, in robotics. The second-order equation is often converted into a set of first-order equations, some of which are trivial (e.g., $u'' = f$ is converted into $v = u', v' = f$) and it is unphysical to inject noise into these trivial equations.

We use a linear coordinate transformation to diagonalize the state covariance matrix G [44] to obtain a canonical form of a model with partial noise from (3.1) and (3.2):

$$\hat{\mathbf{x}}^{n+1} = \hat{f}(\hat{\mathbf{x}}^n, \hat{\mathbf{y}}^n) + \hat{\mathbf{v}}^{n+1}, \quad \hat{\mathbf{v}}^{n+1} \sim N(0,\hat{G}), \tag{3.15}$$

$$\hat{\mathbf{y}}^{n+1} = g(\hat{\mathbf{x}}^n, \hat{\mathbf{y}}^n), \tag{3.16}$$

$$\mathbf{b}^{n+1} = \hat{h}(\hat{\mathbf{x}}^n, \hat{\mathbf{y}}^n) + \hat{\mathbf{w}}^n. \tag{3.17}$$

Here $\hat{\mathbf{x}}^n$ is a p-dimensional column vector, $p < m$ is the rank of the state covariance matrix G, and $\hat{f}$ and $\hat{h}$ a p-dimensional, respectively, k-dimensional vector functions; $\hat{G}$ is a non-singular, diagonal $p \times p$ matrix, $\hat{\mathbf{y}}^n$ is a $(m-p)$-dimensional vector, and g is a $(m-p)$-dimensional vector function. For ease of notation, we drop the hats and, for convenience, we refer to the set of variables $\mathbf{x}^n$ and $\mathbf{y}^n$ as the "forced" and "unforced variables," respectively.

The key to filtering a model with partial noise is observing that the unforced variables at time $n+1$, given the state at time n, are not random. To be sure, $\mathbf{y}^n$ is random for any n due to the nonlinear coupling $g(\mathbf{x}^n, \mathbf{y}^n)$, but the conditional pdf $p(\mathbf{y}^{n+1} \mid \mathbf{x}^n, \mathbf{y}^n)$ is the delta-distribution. For a given initial state $\mathbf{x}^0$, $\mathbf{y}^0$, the target density is

$$\begin{aligned} p(\mathbf{x}^{0:n+1}, \mathbf{y}^{0:n+1} \mid \mathbf{b}^{1:n+1}) \propto\ & p(\mathbf{x}^{0:n}, \mathbf{y}^{0:n} \mid \mathbf{b}^{1:n}) \\ & \times p(\mathbf{b}^{n+1} \mid \mathbf{x}^{n+1}, \mathbf{y}^{n+1}) p(\mathbf{x}^{n+1} \mid \mathbf{x}^n, \mathbf{y}^n) \end{aligned}$$

and the corresponding functions F_j as in (3.4) for models with partial noise are defined by

$$F_j(\mathbf{X}_j^{n+1}) = -\log\left(p(\mathbf{b}^{n+1} \mid \mathbf{X}_j^{n+1}, \mathbf{y}_j^{n+1}) p(\mathbf{X}_j^{n+1} \mid \mathbf{X}_j^n, \mathbf{y}_j^n)\right).$$

With this F_j, we can use the implicit particle filter as described above.

The difference in filtering models with partial noise is that $\mathbf{y}_j^{n+1}$ is fixed for each particle, because its previous state, $(\mathbf{X}_j^n, \mathbf{y}_j^n)$, is known, and because there is no noise

in the equation for the unforced variables $\mathbf{y}^n$. That means that the filter only updates the forced variables $\mathbf{X}_j^{n+1}$ when the observations $\mathbf{b}^{n+1}$ become available. The unforced variables $\mathbf{y}_j^{n+1}$ are moved forward in time using the model, as they should be, since there is no uncertainty in $\mathbf{y}^{n+1}$ given $\mathbf{x}^n, \mathbf{y}^n$. Because the functions F_j depend on the forced variables only, the implicit particle filter reduces in dimension from m to p (the rank of the state covariance matrix G). This fact makes the implicit particle filter particularly effective for models with partial noise, because other filtering techniques, e.g., SIR, struggle to make direct use of the structure of the model.

3.2.4 Combined State and Parameter Estimation

Next, we consider models with unknown parameters, say $\theta \in R^l$, so that the model equation (3.1) becomes

$$\mathbf{x}^n = f^n(\mathbf{x}^{n-1}, \theta) + \mathbf{v}^{n-1}.$$

The goal is to estimate both the states and the parameters, i.e. compute the conditional mean $E\left[\mathbf{x}^{0:n}, \theta | \mathbf{b}^{1:n}\right]$. We approximate the conditional mean by the sample mean, using weighted samples from $p(\mathbf{x}^{0:n}, \theta | \mathbf{b}^{1:n})$. A relatively simple way of estimating the parameters is to append a state equation for the parameters of the form

$$\theta^n = g^n(\mathbf{x}^{n-1}, \theta^{n-1}) + \mathbf{v}_\theta^{n-1},$$

where g^n is an l-dimensional vector function and v_θ^n is an l-dimensional i.i.d. random process. Defining an "extended" state $\hat{\mathbf{x}} = (\mathbf{x}, \theta)$, the implicit particle filter as described above can be applied to estimate $\hat{\mathbf{x}}^{0:n}$ given the data $\mathbf{b}^{1:n}$. The difficulty here lies in how to choose the dynamics g^n of the parameters θ. In particular, this approach is questionable if the parameters are known to be constant in time.

Alternatively, one can use the Markov property of the model and Bayes' rule to derive the recursive formula

$$p(\mathbf{x}^{0:n+1}, \theta \mid \mathbf{b}^{1:n+1}) = p(\mathbf{x}^{0:n}, \theta \mid \mathbf{b}^{1:n}) p(\mathbf{x}^{n+1} \mid \mathbf{x}^n, \theta) p(\mathbf{b}^{n+1} \mid \mathbf{x}^{n+1}) / p(\mathbf{b}^{n+1} \mid \mathbf{b}^{1:n}).$$

Following the now familiar steps and assuming that we have M samples from $p(\mathbf{x}^{0:n}, \theta \mid \mathbf{b}^{1:n})$, we can define the functions F_j by

$$F_j(\mathbf{X}_j^{n+1}, \theta_j) = -\log\left(p(\mathbf{X}_j^{n+1} \mid \mathbf{X}_j^n, \theta_j) p(\mathbf{b}^{n+1} \mid \mathbf{X}_j^{n+1})\right).$$

With these F_j, we can again apply the implicit particle filter as described above. The details and numerical tests for this method applied to ecological models can be found in [55].

3.3 Implementations of the Implicit Particle Filter

The setup for the implicit particle filter as presented in the previous section is rather general, i.e. we have a lot of freedom in how we execute the various steps of the method. The crucial steps are (1) to find the minima of the functions F_j; and (2) choose a map that solves the implicit equation (3.5). In practice, any one of the standard numerical optimization algorithms, e.g. Newton's method, quasi-Newton methods, trust-region methods, or gradient descent, can be used for the minimizations. However the applicability and efficiency of the minimization algorithms are problem dependent. For example, it may be very difficult to compute the Hessians of the functions F_j (for which the derivatives of the model equations are needed), so that Newton's method is not applicable but a quasi-Newton method can be used. If the state dimension is large, and memory limitations become an issue, then a gradient descent method may be the method of choice for minimization of F_j. We explain the minimization algorithms we use in the examples in Sect. 3.5. In the present section, we present two efficient ways of solving the implicit equation (3.5).

3.3.1 Solution of the Implicit Equation via Quadratic Approximation

Inspired by the simplicity of the case for which linear observations are available at each model step, one can try solving a quadratic equation, rather than the implicit equation (3.5). This idea was presented in [7] and is related to the quadratic expansion construction in [17] (see Sect. 3.4.2 for more details) and the Laplace approximation [29] (where the target pdf is approximated by a Gaussian). To find a suitable quadratic equation, expand F_j to second-order accuracy around its minimum:

$$F_j^0 = \phi_j + \frac{1}{2}(\mathbf{X}_j^{n+1} - \mu_j)^T H_j (\mathbf{X}_j^{n+1} - \mu_j),$$

where H_j is the Hessian of F_j, evaluated at the minimizer $\mu_j = \operatorname{argmin} F_j$. With this F_j^0, define the equation

$$F_j^0(\mathbf{X}_j^{n+1}) - \phi_j = \frac{1}{2}\xi_j^T \xi_j, \tag{3.18}$$

which can be solved efficiently using a Cholesky decomposition of the Hessian H_j. Let L_j be a Cholesky factor of $H_j = L_j L_j^T$. It is easy to verify that

$$\mathbf{X}_j^{n+1} = \mu_j + L_j \xi^j,$$

solves (3.18) and that the Jacobian, $J = \left|\det(L_j)\right|$, of this map is easy to calculate, since it is the product of the diagonal elements of L_j. To avoid introducing any bias, one needs to account for the error we made by solving (3.18) rather than the true equation (3.5) in the weights. With this importance function, we obtain the weights

$$w_j^{n+1} = w_j^n \frac{\exp(-F_j(\mathbf{X}_j^{n+1}))}{\exp(-\xi_j^T \xi_j/2)} J,$$
$$= w_j^n \exp(-\phi_j) \left|\det(L_j)\right| \exp(F_j^0(\mathbf{X}_j^{n+1}) - F_j(\mathbf{X}_j^{n+1})). \tag{3.19}$$

This method of sampling has a geometric interpretation: the target pdf is approximated locally by a Gaussian centered at the mode of the target pdf, and with a covariance matrix that depends on the curvature of the target pdf at the minimum (i.e., the importance function is obtained from the Laplace approximation). These samples are weighted by $\exp(F_j^0(\mathbf{X}_j^{n+1}) - F_j(\mathbf{X}_j^{n+1}))$ to account for the error we make by solving the quadratic equation (3.18) rather than the true equation (3.5). If a Newton or quasi-Newton method is used for the minimization of F_j, then H_j, and often even L_j, are already available and this sampling method is easy to code and numerically efficient.

However, if the Gaussian approximation is not valid, for example, because the skewness in the target density is significant, the variance of the weights is increased, which should be avoided. In such cases, exact solution of (3.5) is advisable and we present an efficient method for doing so in the next subsection.

3.3.2 Solution of the Implicit Equation via Random Maps

Here we review the approach presented in [39] which solves (3.5) by the random change of variables (random map)

$$\mathbf{X}_j^{n+1} = \mu_j + \lambda_j L_j^T \eta_j, \tag{3.20}$$

where λ_j is a scalar and $\eta_j = \xi_j / \sqrt{\xi_j^T \xi_j}$ is uniformly distributed on the unit m-sphere (or rm-sphere if the observations are sparse in time), and where $\mu_j = \operatorname{argmin} F_j$. The square matrix L_j contains all prior information we have about F_j and is deterministic and invertible. We will discuss the choice of L_j in more detail below.

By substitution of (3.20) into (3.5), we obtain a single algebraic equation in a single variable λ_j:

$$F_j(\mu_j + \lambda_j L_j^T \eta_j) - \phi_j = \frac{1}{2} \xi_j^T \xi_j.$$

The solution of the above equations defines $\mathbf{X}_j^{n+1}$ through (3.20). The geometric interpretation of this approach is that we choose a direction η_j at random, and then solve for λ_j, which tells us how far we need to search in this direction to hit the level set of F_j that is defined by the sample from the reference variable ξ.

To compute the weights in (3.6), we need to compute the Jacobian of the random map (3.20), which is:

$$J = 2|\det L_j|\, \rho_j^{1-m/2} \left| \lambda_j^{m-1} \frac{\partial \lambda_j}{\partial \rho_j} \right|,$$

where $\rho_j = \xi_j^T \xi_j$, and m is the dimension of the state space (if the observations are sparse, m in the above formula is to be replaced by rm). We refer to [39] for the details of this calculation; however, note that the Jacobian is easy to evaluate, since the scalar derivative $\partial\lambda_j/\partial\rho_j$ can be computed efficiently either by using finite differences or by evaluating

$$\frac{\partial\lambda_j}{\partial\rho_j} = \frac{1}{2(\nabla F_j)L^T\eta_j},$$

where ∇F_j is the gradient of F_j. The weight of the sample we obtained by solving (3.5) with the random map (3.20) is thus

$$w_j^n \propto w_j^{n-1} \exp(-\phi_j) |\det L_j| \, \rho_j^{1-m/2} \left| \lambda_j^{m-1} \frac{\partial\lambda_j}{\partial\rho_j} \right|. \tag{3.21}$$

We now discuss the choices of the matrix L_j. Suppose we apply our random map method to the special case we discussed in Sect. 3.2.1, i.e. the observation equation (3.2) is linear and in-sync with the model. With $L_j = I$, we find that

$$\lambda_j = \frac{\sqrt{\rho_j}}{\sqrt{\eta_j^T \Sigma^{-1} \eta_j}}$$

where Σ is given in (3.8). The weights of the particles become

$$w_j^{n+1} \propto w_j^n \exp(-\phi_j)(\eta_j^T \Sigma^{-1} \eta_j)^{-m/2},$$

and, since Σ is symmetric, are bounded above and below by the eigenvalues of Σ. The Jacobian J can vary dramatically from one sample to another, especially if the largest and smallest eigenvalues of Σ_j are separated by a large gap. If we choose L_j such that $\Sigma = L_j^T L_j$, we find $\lambda_j = \sqrt{\rho_j}$ and $J = |\det L_j|$. This Jacobian is constant and need not be computed, and with this choice of L_j, we sample the optimal importance function, by solving (3.5) with the random map (3.20), see [39].

In the general case, we can use the information on the curvature of F_j we have in its Hessian H_j, by choosing L_j to be a Cholesky factor of this Hessian, i.e. $H_j = L_j^T L_j$. This choice should speed up the solution of (3.5), especially if F_j is quadratic or nearly so. This choice of L_j also suggests a "good" initialization for the numerical computation of the parameter λ_j in the random map. One can expect λ to be on the order of $\sqrt{\rho}$ and so that the iterative solution of (3.5) is initialized with $\lambda_j^0 = \sqrt{\rho_j}$.

3.4 Comparison with Other Sequential Monte Carlo Schemes

We wish to compare the implicit particle filter with other data assimilation methods and point out differences and similarities between these methods and the implicit particle filter.

3.4.1 Comparison with the SIR Filter

The SIR filter [16] uses the model (3.1), i.e. $p(\mathbf{x}^{n+1}|\mathbf{x}^n)$, as the importance function, so that the weights are

$$w_{j,\mathrm{SIR}}^{n+1} \propto w_{j,\mathrm{SIR}}^{n} p(\mathbf{b}^{n+1}|\mathbf{x}^{n+1}). \tag{3.22}$$

The SIR filter can be formulated as an implicit particle filter with a different choice of ϕ_j in (3.5). Recall that for the implicit particle filter $\phi_j = \min F_j$ in (3.5). If we replace ϕ_j by

$$\phi_{j,\mathrm{SIR}} = -\log(p(\mathbf{b}^{n+1}|\mathbf{X}_j^{n+1})), \tag{3.23}$$

then (3.5) becomes

$$-\log(p(\mathbf{X}_j^{n+1}|\mathbf{X}_j^n)) = \frac{1}{2}\xi_j^T \xi_j.$$

For Gaussian model noise with covariance matrix Q, we thus have

$$(\mathbf{X}_j^{n+1} - \mathbf{X}_j^n)^T Q^{-1} (\mathbf{X}_j^{n+1} - \mathbf{X}_j^n) = \xi_j^T \xi_j, \tag{3.24}$$

which can be solved by

$$\mathbf{X}_j^{n+1} = \mathbf{X}_j^n + L^T \xi_j, \tag{3.25}$$

where L is a Cholesky factor of $Q = LL^T$. Note that the Jacobian of this map is constant for all particles (it is the determinant of L), and thus need not be determined for computation of the weights. Moreover, computing $\mathbf{X}_j^{n+1}$ by (3.25) is equivalent to running the model forward for one time step, so that this implicit particle filter uses the same importance function as the SIR filter. The weights of the particles of this implicit particle filter, given by (3.6), are therefore also the same as the weights of the SIR filter in (3.22).

The SIR filter is thus an implicit particle filter with ϕ_j in (3.5) replaced by $\phi_{j,\mathrm{SIR}}$ in (3.23). This observation illustrates why the SIR filter requires significantly more particles than the implicit particle filter (with $\phi_j = \min F_j$): choosing ϕ_j to be the minimum of F_j in (3.5) maps the high probability region of the reference variable ξ to the neighborhood of the minimum of F_j, which corresponds to the high probability region of the target pdf. Choosing ϕ_j as in (3.23), on the other hand, maps the high probability region of the reference variable to the high probability region of the model (3.1). The overlap of the high probability region of the model with the high probability region of the target pdf can be very small, and in these cases, the SIR filter requires a large number of particles to provide accurate state estimates.

3.4.2 Comparison with Optimal Importance Function Filters

A crucial step in designing an efficient sequential Monte Carlo method is "a good choice" of the importance function. In the context of data assimilation, an "optimal" importance function can be found in [17] and the references therein:

$$\hat{q} = p(\mathbf{X}_j^{n+1}|\mathbf{b}^{n+1}, \mathbf{X}_j^n). \tag{3.26}$$

Here, "optimal" means that the variance of the weights of a given particle is zero (but not the variance of the weights of all the particles), and a particle filter which uses the optimal importance function is often called an optimal importance function filter. The weights of the optimal particles can be shown to be

$$\hat{w}_j^{n+1} \propto \hat{w}_j^n p(\mathbf{b}^{n+1} \mid \mathbf{X}_j^n).$$

If observations are available at every model step and if, in addition, the model and observation noise are Gaussian and the observation function h^n in (3.2) is linear, then the optimal importance function $\hat{q}$ is Gaussian with mean μ and covariance Σ as in (3.10) and (3.8) [17]. It was shown in Sect. 3.2.1 that in this case the implicit particle filter uses exactly this density as the importance function and that its weights are proportional to $p(\mathbf{b}^{n+1} \mid \mathbf{X}_j^n)$. Thus, for this special case, the implicit particle filter samples the optimal importance function and represents a convenient implementation of the optimal importance function filter.

In the general case, the optimal importance function is not readily available. One can rewrite (3.26) as

$$\hat{q} = \frac{p(\mathbf{b}^{n+1}|\mathbf{X}_j^{n+1})p(\mathbf{X}_j^{n+1}|\mathbf{X}_j^n)}{p(\mathbf{b}^{n+1}|\mathbf{X}_j^n)}, \tag{3.27}$$

and try to compute the denominator, e.g., by Monte Carlo using

$$p(\mathbf{b}^{n+1}|\mathbf{X}_j^n) = \int p(\mathbf{b}^{n+1}|\mathbf{X}_j^{n+1})p(\mathbf{X}_j^{n+1}|\mathbf{X}_j^n)d\mathbf{X}_j^{n+1}.$$

which is often hard to do. However, even if $p(\mathbf{b}^{n+1}|\mathbf{X}_j^n)$ is available, sampling directly from the optimal importance function may be hard. In this case, one can define the function

$$l_j = \log p(\mathbf{X}_j^{n+1}|\mathbf{b}^{n+1}, \mathbf{X}_j^n),$$

find its maximum $\lambda_j = \max l_j$, and expand l_j around its maximum:

$$l_j = \lambda_j + \frac{1}{2}\left(\mathbf{X}_j^{n+1} - \gamma_j\right)^T H_j \left(\mathbf{X}_j^{n+1} - \gamma_j\right),$$

where $\gamma_j = \text{argmax}\, l_j$ and H_j is the Hessian of l_j, evaluated at the maximum. The quadratic expansion suggests the (suboptimal) Gaussian importance function

$$q = N(\gamma_j, H_j^{-1}).$$

This approach has many similarities to the implicit particle filter when F_j in (3.5) is approximated by its quadratic expansion, also see Sect. 3.3.1. Specifically, both methods require the solution of an optimization problem (to search for the high probability regions in the target pdf), and sampling from a multivariate Gaussian density. However, the implicit particle filter avoids the difficulties which arise in the optimal importance function filter from the need to compute $p(\mathbf{b}^{n+1}|\mathbf{X}_j^n)$ in (3.27). Since the computation of this term can be expensive, the implicit particle filter seems to be more efficient and easier to implement.

3.4.3 Comparison with the Kalman Filter and with Variational Data Assimilation Methods

The KF is, strictly speaking, only applicable to linear systems (f and h are linear in (3.1) and (3.2)), driven by Gaussian noise (both $\mathbf{v}^n$ and $\mathbf{w}^n$ in (3.1) and (3.2) are Gaussian) [27, 28]. In this special case, the KF is widely used and efficient implementations are available for large, linear models. The implicit particle filter (with one particle) implements the KF for linear dynamics and Gaussian noise, because (3.8) and (3.10) become, upon rearrangement of the terms, the KF formulas.

For nonlinear, non-Gaussian HMM models, the extended Kalman filter (EKF) uses a linearization of the model and observation equation along with the standard KF formalism [26]. The EnKF implements the KF step using a covariance matrix that is approximated by Monte Carlo, i.e. by the sample covariance of many model runs. This step avoids the often costly computation of the covariance matrix in the KF formalism, and the EnKF can outperform the KF in linear systems with a very large state dimension [19]. Moreover, the EnKF injects the nonlinearity of the model into the KF formalism through the sample covariance matrix but relies on a linearization of the observation equation. For this reason, both EKF and EnKF can give good results if the nonlinearity is not too strong and if the number of model steps between observations is not too large. The implicit particle filter, on the other hand, tackles the full nonlinear problem and can outperform EnKF in nonlinear problems [40].

Variational data assimilation finds the most likely state given the data by finding the mode of the target pdf [2, 10, 11, 33, 48–50, 56]. This mode can be found by minimizing a suitable cost function which is very similar to the functions F_j used in the implicit particle filter. Specifically, the cost function in weak constraint 4D-Var is the function F_j, with $\mathbf{X}_j^n$ being a variable (in the context of the filter, $\mathbf{X}_j^n$ is a parameter), and with an additional quadratic term, that corresponds to a Gaussian approximation of prior information on the state $\mathbf{X}_j^n$ (see, e.g., [50]). Thus, the computational cost of the implicit particle filter is roughly the cost of a variational method, times the number of particles required (which should be a relatively small number), since the sampling can be carried out very efficiently once the minima of the F_j are obtained (see [41] for a detailed comparison of the implicit particle filter

with variational data assimilation). It is also important to note that the minimization of the functions F_j and the sampling for the different particles can be carried out in parallel.

The main benefits of investing the additional effort and using the implicit particle filter rather than a variational method are: (1) a variational method computes the maximum a posteriori estimate (MAP), while the implicit particle filter approximates the minimum mean square error estimate (MMSE); in many situations, e.g., if the skewness in the target density is significant, the MMSE is a better estimator than the MAP[5]; (2) the implicit particle filter provides a quantitative measure of the uncertainty of its state estimate (e.g., sample covariance or higher moments), while variational methods only provide a state trajectory, but no error bounds; and (3) the implicit particle filter is a sequential method and thus it is relatively easy to assimilate more observations as they become available, while there are theoretical and practical issues with the sequential continuation of variational methods.

3.5 Applications

We demonstrate the applicability and efficiency of the implicit particle filter on six examples and provide details about the implementation in each of the examples. The first two examples show that the implicit filter can outperform other methods even in relatively simple problems. In examples 3 and 4 we demonstrate the applicability of the implicit particle filter to models that exhibit chaotic behaviors. We then consider data assimilation for a model of the geomagnetic field coupled to the core velocity. Finally, we consider an ecological model and use the implicit particle filter to assimilate ocean color data obtained by NASA's SeaWiFS satellite.

3.5.1 A Simple Example

This first example, taken from [47], corresponds to the special case discussed in Sect. 3.2.1: the functions f^n in (3.1) are zero and h^n in (3.2) is the identity, i.e. the model and data equations become

$$\mathbf{x}^{n+1} = \mathbf{v}^n, \quad \mathbf{b}^{n+1} = \mathbf{x}^{n+1} + \mathbf{w}^{n+1},$$

where $\mathbf{v}^n$ and $\mathbf{w}^n$ are m-dimensional vectors whose elements are independent standard normal variates. If the initial conditions are known precisely (i.e., if we start all particles from the same state), then, by (3.14), the weights of the particles produced by the implicit filter are constant, since ϕ_j in (3.11) is constant. The weights of the SIR filter, on the other hand, are uneven and, if the dimension m exceeds 100, one observes frequently that a single SIR particle hogs all the probability [3, 47], so that this version of SIR fails.

3.5.2 Stochastic Volatility Model

We consider the stochastic volatility model

$$x^{n+1} = ax^n + v^n, \quad b^{n+1} = w^{n+1}\sigma_b \exp(x^{n+1}/2),$$

where, as in [45], $a = 0.9702$, $\sigma_b = 0.5992$, and v^n and w^n are Gaussian random variables with mean 0 and standard deviation $\sigma_v = 0.178$ and $\sigma_w = 1$, respectively. For this (non-Gaussian) model, the functions F_j become

$$F_j(X_j^{n+1}) = \frac{\left(X_j^{n+1} - aX_j^n\right)^2}{2\sigma_v^2} + \frac{\left(b^{n+1}\right)^2}{2\sigma_b^2} \exp(-X_j^{n+1}).$$

For the implicit particle filter we minimize these functions using Newton's method and then use the quadratic approximation as in Sect. 3.3.1 to generate the particles with weights given by (3.19). Results of a test-run with the implicit filter as well as with an SIR filter and the (adapted) auxiliary particle filters in (see Sect. 3.3.4 in [45]) are shown in Fig. 3.1. We observe that the implicit particle filter gives good results with a small number of particles ($\sim$100); the SIR and auxiliary particle filters are less accurate even if we increase the number of particles significantly (to about 100,000 for SIR and 5,000 for the auxiliary particle filter).

3.5.3 The Stochastic Lorenz Attractor

This example is taken from [39]. We follow [6, 37, 38] and consider the stochastic Lorenz attractor [36]

$$dx = \sigma(y - x)dt + g_1 dW_1, \tag{3.28}$$
$$dy = (x(\rho - z) - y)dt + g_2 dW_2, \tag{3.29}$$
$$dz = (xy - \eta z)dt + g_3 dW_3, \tag{3.30}$$

with the standard parameters $\sigma = 10$, $\rho = 28$, $\eta = 8/3$, and initial conditions $x(0) = -5.91652$, $y(0) = -5.52332$, $z(0) = 24.5723$. The noise is chosen equally strong for all variables, so that $g_1 = g_2 = g_3 = g = \sqrt{2}$. We discretized the continuous equations by the Klauder–Petersen (KP) scheme [30]

$$\mathbf{x}^{n+1,*} = \mathbf{x}^n + \delta f(x^n) + g\mathbf{v}_1,$$
$$\mathbf{x}^{n+1} = \mathbf{x}^n + \frac{\delta}{2}\left(f(x^n) + f(x^{n+1,*})\right) + g\mathbf{v}_2,$$

where δ is the time step, $\mathbf{v}_1, \mathbf{v}_2 \sim N(0, \delta I)$, and where $f(\cdot)$ can be read off the Lorenz attractor (3.28)–(3.30). We are content here with an approximation with time step $\delta = 0.01$ (see [39] for more details on the discretization). Observations of all three

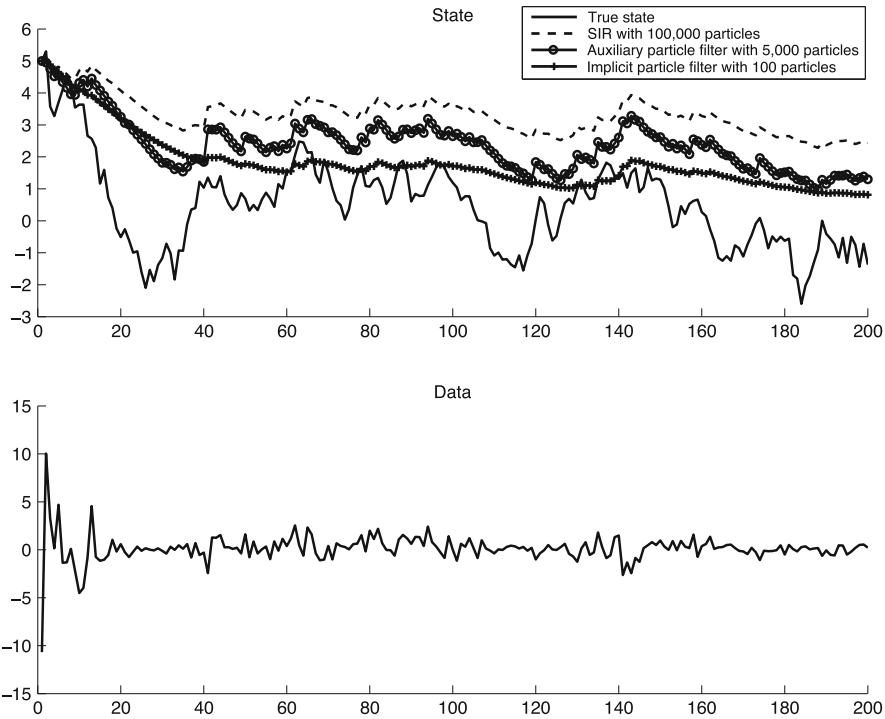

Fig. 3.1 Filtering results for the stochastic volatility model: an SIR filter with 100,000 particles and the auxiliary particle filter with 5,000 particles are less accurate than the implicit particle filter with 100 particles.

state variables, corrupted by noise with variance 0.1, became available every 0.48 dimensionless time units (every 48 steps). This is a hard data assimilation problem and some filters miss transitions from one wing of the Lorenz butterfly to the other [38].

The minimization of the functions F_j was done using Newton's method, initialized by a free model run without noise (a larger gap between observations can cause problems here; however, a more sophisticated initialization and a more robust minimization provide a cure, see [41]). Note that the argument of the F_j's are the state variables $\mathbf{X}_j^n$, as well as the intermediate model steps $\mathbf{X}_j^{n,*}$. The problem is thus of dimension 288: 3 dimensions for the Lorenz attractor, times 2 for the intermediate step x^* of the KP scheme, times 48 for the gap between observations. If the variance matrix of the reference variable ξ is the identity matrix I, we are expressing a vector variable of small variance as a function of a unit reference variable, and this produces very small Jacobians J which can lead to underflow. One solution is to rescale ξ which, after all, is arbitrary. What we did instead is keep track of the logarithms of the weights rather than the weights themselves wherever we could; this solved the problem. At each assimilation step, we thus sampled a 288

dimensional standard normal variate ξ_j (the reference variable) and computed the random direction $\eta_j = \xi_j / \sqrt{\xi_j^T \xi_j}$ to be used in the random map (3.20). Because we used Newton's method for the minimization of F_j, the Cholesky factor L_j of the Hessian evaluated at the minimum was available and we used it in (3.20). Substitution of the map (3.20) into the algebraic equation (3.5) gave the required equation for λ_j, which we solved by Netwon's method. The iteration was initialized by choosing $\lambda_j^0 = \sqrt{\rho_j}$ and typically converged within four to six steps. Finally, we computed the weight of the particle using (3.21) and the numerical derivative $\partial\lambda/\partial\rho$, with a perturbation $\Delta\lambda = 10^{-5}\sqrt{\rho}$. We repeated this process for each particle and resample with "algorithm 2" in [16]. We decided to resample at every time an observation became available.

To compare the implicit particle filter with an SIR filter, we ran 1,000 twin experiments. That is, we ran the model for 960 time steps and produced artificial observations (corrupted by the assumed noise). This model run was the reference we wished to reconstruct using the SIR and the implicit particle filters. For each experiment, the error at time $t^N = 9.6$ is measured by

$$e = ||\mathbf{x}_{ref}^N - \mathbf{x}^N||,$$

where the norm is Euclidean, $\mathbf{x}_{ref}^N$ is the reference state, and $\mathbf{x}^N$ is the reconstruction by a filter. We computed this error for each twin experiment, and, after running 1,000 twin experiments, we computed the mean value of the error norms (mean error, for short) and the mean of the variance of the error norms (mean variance of the error, for short). The mean of the error norm is a better estimate for the errors than the mean error because it does not allow for cancelations. The mean variance of the error is not the variance of the mean, it is a fair estimate of the error in each individual run. Our results are in Table 3.1.

Table 3.1 Filtering results for the Lorenz attractor.

# of Particles	Mean error/mean variance of the error	
	Implicit particle filter	SIR
5	0.2192/0.3457	-/-
10	0.2317/0.4905	0.9964/1.9970
20	0.1927/0.1646	0.5352/0.7661
50	-/-	0.4271/0.5445
100	-/-	0.2336/0.1229

The implicit particle filter yielded good results with 20 particles, while an SIR filter required about 100 particles for comparable accuracy (the results obtained for the SIR filter are in agreement with those previously reported in [6, 38]). The reason for the large difference in the number of particles required for these two filters is as follows. The large gap between observations implies that the SIR importance function and the target density become nearly mutually singular. The "unguided" SIR particles are therefore very likely to become unlikely, and only very few of them

carry a significant weight. The particles of the implicit particle filter, on the other hand, are guided towards the high probability regions because they are generated by solving (3.5), which incorporates information from the data. The larger number of particles required by the SIR filter thus indicates that the "focusing" of the particles towards the high probability regions of the target pdf was indeed achieved by the implicit particle filter.

The computational cost of these filters is comparable in this example. The implicit particle filter requires fewer particles, but the computations for each particle are more expensive when compared with the SIR filter.

3.5.4 The Stochastic Kuramoto–Sivashinsky Equation

We follow [39] and consider data assimilation for the stochastic Kuramoto–Sivashinsky (SKS) equation

$$u_t + uu_x + u_{xx} + \nu u_{xxxx} = g\, W(x,t)$$

where $\nu > 0$ is the viscosity, g is a scalar, and $W(x,t)$ is a space–time white noise process. The SKS equation is a chaotic SPDE that models laminar flames or reaction–diffusion systems [32, 46] and recently has been used as a large dimensional test-problem for data assimilation algorithms [6, 24].

We consider the m-dimensional Itô–Galerkin approximation of the SKS equation

$$dU = (\mathscr{L}(U) + \mathscr{N}(U))dt + g\, dW_t^m,$$

where U is a finite dimensional column vector whose components are the Fourier coefficients of the solution and where dW_t^m is a truncated cylindrical Brownian motion (BM) [35], obtained from the projection of the noise process $W(x,t)$ into the Fourier modes. Assuming that the initial conditions $u(x,0)$ are odd with $\tilde{U}_0(0) = 0$ and that dW_t^m is imaginary, all Fourier coefficients $U_k(t)$ are imaginary for all $t \geq 0$. Writing $U_k = i\hat{U}_k$ and subsequently dropping the hat gives

$$\mathscr{L}(U) = \mathrm{diag}(\omega_k^2 - \nu\omega_k^4)U,$$
$$\{\mathscr{N}(U)\}_k = -\frac{\omega_k}{2} \sum_{k'=-m}^{m} U_{k'} U_{k-k'},$$

where $\omega_k = 2\pi k/L, k = 1,\ldots,m$ and $\{\mathscr{N}(U)\}_k$ denotes the kth element of the vector $\mathscr{N}(U)$. We choose a period $L = 16\pi$ and a viscosity $\nu = 0.251$, to obtain SKS equations with 31 linearly unstable modes. This setup is similar to the SKS equation considered in [24]. With these parameter values there is no steady state as in [6]. We chose zero initial conditions $U(0) = 0$, so that the solution evolves solely due to the effects of the noise. To approximate the SKS equation, we keep $m = 512$ of the Fourier coefficients and use the exponential Euler scheme [25], with time step $\delta = 2^{-12}$ for time discretization (see [39] for details).

We are solving the SKS equations in Fourier variables, but we choose to observe in physical space (as is maybe physically reasonable). Specifically, we observe the solution $u(x,t)$ at $m/2$ equidistant locations and at every model step through the nonlinear observation operator $h(x) = x + x^3$. The solution of the algebraic equation (3.5) is easiest when the functions F_j is nearly diagonal, i.e., when its linearizations around a current state are nearly diagonal matrices; this requires in particular that the variables that are observed coincide with the variables that are evolved by the dynamics. Observing in physical space while computing in Fourier space creates the opposite situation, in which each observation is related to the variables one computes by a dense matrix. This problem can be overcome using the random map algorithm presented in Sect. 3.3.2.

The minimization of F_j was done using Newton's method, initialized by a free model run without noise. The Cholesky factor of the Hessian of F_j at the minimum was used as the matrix L_j in (3.20), and (3.5) was solved using this random map and a Newton iteration on λ_j. To test the implicit particle filter we ran twin experiments as in Sect. 3.5.3. The error at time t^n is defined as

$$e^n = ||U^n_{ref} - U^n_F||$$

where the norm is the Euclidean norm U^n_{ref} denotes the set of Fourier coefficients of the reference run, and U^n_F denotes the reconstruction by the filter, both at the fixed time t^n. Results of 500 twin experiments are shown in Fig. 3.2.

We observe from Fig. 3.2 that the implicit particle filter requires far fewer particles than the SIR filter. Again, the example confirms that the implicit particle filter focuses its particles on the high probability regions of the target pdf. The focusing effect is more pronounced in the SKS equation than in the Lorenz attractor (see Sect. 3.5.3), because the dimension of the state space of the SKS equation is 512, and therefore much larger than the dimension of the Lorenz attractor (dimension 3).

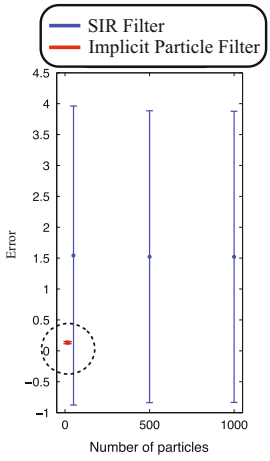

Fig. 3.2 Filtering results for the SKS equation: the error statistics are shown as a function of the number of particles for SIR (*blue*) and implicit particle filter (*red*). The error bars represent the mean of the errors and mean of the standard deviations of the errors.

3.5.5 Application to Geomagnetic Data Assimilation

The following example is taken from [40]. We wish to apply the implicit particle filter to a test problem in geomagnetic data assimilation, defined by two SPDEs

$$\begin{aligned}\partial_t u + u\partial_x u &= b\partial_x b + \nu\partial_x^2 u + g_u \partial_t W(x,t),\\ \partial_t b + u\partial_x b &= b\partial_x u + \partial_x^2 b + g_b \partial_t W(x,t),\end{aligned}$$

where $\nu = 10^{-3}$, $g_u = 0.01$, $g_b = 1$ are scalars and where W is a spatially smooth stochastic process [21, 40]. We consider the above equations on the strip $0 \le t \le 0.2$, $-1 \le x \le 1$ and with given boundary and initial conditions. Physically, u represents the velocity field at the core, and b represents the magnetic field of the earth. The model is essentially the model proposed in [21], but with additive noise

$$W(x,t) = \sum_{k=0}^{\infty} \alpha_k \sin(k\pi x) w_k^1(t) + \eta_k \cos(k\pi/2x) w_k^2(t),$$

where w_k^1, w_k^2 are independent BMs and

$$\alpha_k = \eta_k = \begin{cases} 1, \text{ if } k \le 10,\\ 0, \text{ if } k > 10.\end{cases} \tag{3.31}$$

This simple noise model represents a spatially smooth noise which decreases in amplitude near the boundaries. The continuous equations are discretized using Legendre spectral elements in space, and an implicit–explicit first-order scheme in time (see [4, 15, 31, 40]). We are content with an approximation that uses one Legendre element of order 300 for u and one for b, and a time step $\delta = 0.002$.

The data are the values of the magnetic field b, measured at 200 equally spaced locations in $[-1,1]$ and corrupted by noise:

$$z^l = Hb^{q(l)} + sV^l,$$

where $s = 0.001$ and H is a $k \times m$-matrix that maps the numerical approximation b to the locations where data is collected. We consider data that are available every ten model steps.

For our choice of α_k, η_k in (3.31), the state covariance matrices of the discrete equations are singular, i.e. the model is subject to partial noise (see Sect. 3.2.3). Upon linear coordinate transformation, that diagonalizes the state covariance matrix, we obtain a model of the form (3.15)–(3.17). Because the second derivatives of the functions F_j are hard to calculate, we use a simple gradient descent algorithm with line search to carry out the minimization. As in previous examples, the minimization is initialized by a free model run. Since no information on the curvature of F_j is available, we set L_j in the random map (3.20) to the identity matrix. Equation (3.5) is then solved by a Newton iteration, initialized with $\lambda_j = 0$ (i.e., we start close to the minimum of F_j).

To assess the performance of the implicit particle filter, we ran 100 twin experiments. For each twin experiment, we calculate and store the error at $t = T = 0.2$ in the velocity, $e_u = ||u(x,T) - u_{Filter}(x,T)||$, and in the magnetic field, $e_b = ||b(x,T) - b_{Filter}(x,T)||$. After running the 100 twin experiments, we calculate the mean of the error norms and the variance of the error norms and scale the results by the mean of the norm of u and b, respectively. Figure 3.3 shows the results.

It is evident from this figure that the implicit particle filter requires few particles to yield accurate state estimates with less than 1 % error in the observed magnetic field b and less than 15 % error in the unobserved velocity u. The SIR filter with 1,000 particles gives significantly larger errors (about 10 % in the observed variables b and 20 % in the unobserved variable u) as well as much larger variances in the errors. The EnKF requires about 500 particles to achieve the accuracy of the implicit particle filter with only 4 particles. These examples thus provide further numerical evidence that the implicit particle filter can achieve the desired focusing effect and, as a consequence, is applicable to large dimensional data assimilation problems.

3.5.6 *Assimilation of Ocean Color Data from NASA's SeaWiFS Satellite*

We apply the implicit particle filter in its iterative implementation [9] (which is not discussed in this review) to a prototypical marine ecosystem model described in [18]. The model involves four state variables: phytoplankton P (microscopic

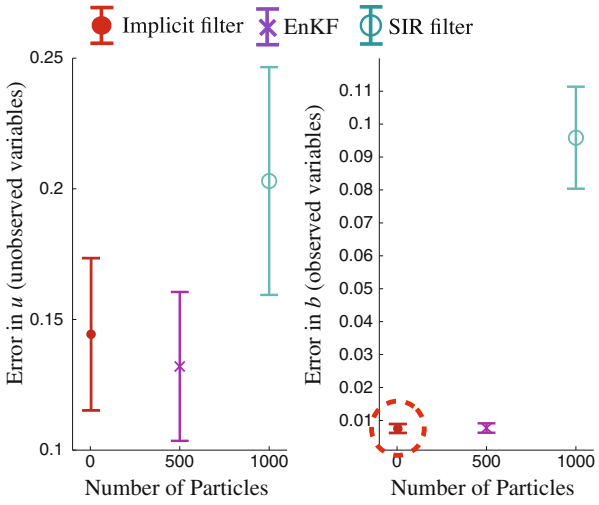

Fig. 3.3 Filtering results for the geomagnetic test problem. The errors of the implicit particle filter (*red*), EnKF (*purple*), and SIR filter (*green*) are plotted as a function of the number of particles. The error bars represent the mean of the errors and mean of the standard deviations of the errors.

plants), zooplankton Z (microscopic animals), nutrients N (dissolved inorganics), and detritus D (particulate organic nonliving matter). At the initial time $t = 0$ we have $P(0) = 0.125$, $Z(0) = 0.00708$, $N(0) = 0.764$, and $D(0) = 0.136$. The system is described by the nonlinear ordinary differential equations

$$\begin{aligned}
\frac{dP}{dt} &= \frac{N}{0.2+N}\gamma P - 0.1P - 0.6\frac{P}{0.1+P} + \mathcal{N}(0,\sigma_P^2),\\
\frac{dZ}{dt} &= 0.18\frac{P}{0.1+P}Z - 0.1Z + \mathcal{N}(0,\sigma_Z^2)\\
\frac{dN}{dt} &= 0.1D + 0.25\frac{P}{0.1+P}Z - \gamma P\frac{N}{0.2+N} + 0.05Z + \mathcal{N}(0\sigma_N^2)\\
\frac{dD}{dt} &= -0.1D + 0.1P + 0.18\frac{P}{0.1+P}Z + 0.05Z + \mathcal{N}(0,\sigma_D^2).
\end{aligned}$$

The variances of the noise terms are $\sigma_P = P(0)$, $\sigma_Z = 0.01Z(0)$, $\sigma_N = 0.01N(0)$, and $\sigma_D = 0.01D(0)$. We discretize the above equations with the stochastic Euler method (see [31]) with time step $\delta = 1$ day. The growth rate at time step t, γ_t, follows the recursion

$$\gamma_t = 0.14 + 3\Delta\gamma_t, \quad \text{where } \Delta\gamma_t = 0.9\Delta\gamma_{t-1} + \mathcal{N}(0,\sigma_\gamma^2),$$

with $\sigma_\gamma = 0.01$. The observations were obtained from NASA's SeaWiFS satellite ocean color images and provide a time series (190 data points from late 1997 to mid-2002) for the phytoplankton P by

$$\log b(t) = \log P(t) + \mathcal{N}(0,\sigma_b^2),$$

where $\sigma_b = 0.3$. We apply the implicit particle filter and the standard SIR filter to find a trajectory of the system consistent with the data. We observe that the implicit particle filter with only ten particles does better than the SIR filter with ten particles, in the sense that the filtered output matches the data better (see Fig. 3.4). In fact the SIR filter requires about 100 particles to achieve the accuracy of the implicit particle filter with only 10 particles. This example thus provides further numerical evidence that the implicit particle filter can provide accurate state estimates with only a few particles. Moreover, the example shows that the implicit particle filter can work well with real data.

3.6 Conclusion

One of the barriers to the successful application of sequential Monte Carlo methods is sample impoverishment, i.e. the fact that the number of samples (particles) required can grow dramatically with the state dimension. The implicit particle filter is an attempt to overcome this problem. The main idea is to focus the particles so that they remain within the high probability regions of the target pdf. We have described the mathematical background of this idea in detail and have shown that the regions

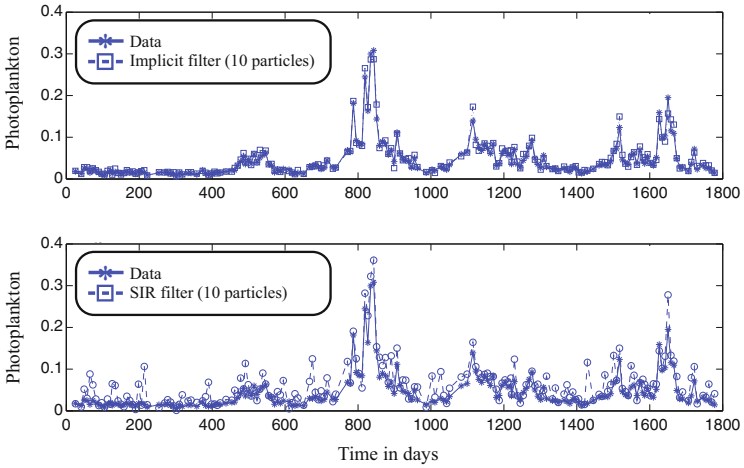

Fig. 3.4 The concentration of phytoplankton as a function of time. *Top*: data and reconstruction by the implicit particle filter with ten particles. *Bottom*: data and reconstruction by the SIR filter with ten particles.

of high probability can be identified by particle-by-particle minimization. Samples within these regions are obtained by solving data-dependent algebraic equations. We presented two effective algorithms for solving these equations and discussed the advantages of various numerical minimization algorithms in examples. We have considered special cases of interest (e.g., partial noise), in both theory and in examples, and made connections with several other data assimilation methods (SIR, EnKF, and variational methods). Six numerical examples have been given to illustrate the theory and demonstrate the broad applicability and efficiency of the implicit particle filter. The examples indicate that the implicit particle filter indeed achieves the desired focusing effect and that this focusing effect keeps the number of particles manageable even if the dimension of the state space is large.

Acknowledgments We would like to thank our collaborators Professor Robert Miller, Professor Yvette Spitz, and Dr. Brad Weir, at Oregon State University, for their comments and for very helpful discussions. This work was supported in part by the Director, Office of Science, Computational and Technology Research, U.S. Department of Energy under Contract No. DE-AC02-05CH11231, and by the National Science Foundation under grants DMS-1217065 and OCE-0934298.

References

1. Arulampalam M.S., Maskell S., Gordon N., Clapp T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, *IEEE Trans. Signal Process.* 10, 197–208.
2. Bennet A.F., Leslie L.M., Hagelberg C.R., and Powers P.E. (1993). A Cyclone prediction using a barotropic model initialized by a general inverse method, *Mon. Weather Rev.* 121, 1714–1728.

3. Bickel P., Li B., and Bengtsson T. (2008). Sharp failure rates for the bootstrap particle filter in high dimensions, *IMS Collections: Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, 318–329.
4. Canuto C., Hussaini M. Y., Quarteroni A., and Zang T. A. (2006). *Spectral Methods. Fundamentals in Single Domains* Berlin (Germany): Springer.
5. Chorin A.J., and Hald O.H. (2009). *Stochastic Tools in Mathematics and Science, Second Edition.* NY: Springer.
6. Chorin A.J., and Krause P. (2004). Dimensional reduction for a Bayesian filter, *PNAS*, 101, 15013–15017.
7. Chorin A.J., Morzfeld M., and Tu X. (2010). Implicit particle filters for data assimilation, *Commun. Appl. Math. Comput. Sci.*, 5(2), 221–240.
8. Chorin A.J. and Tu X. (2009). Implicit sampling for particle filters, *Proc. Nat. Acad. Sc. USA*, 106, 17249–17254.
9. Chorin A.J. and Tu X. (2012), An iterative implementation of the implicit nonlinear filter, *M2AN* 46, 535–543.
10. Courtier P., Thepaut J.N., and Hollingsworth A. (1994). A strategy for operational implementation of 4D-Var, using an incremental approach, *Q. J. R. Meteorol. Soc.* 120, 1367–1387.
11. Courtier P. (1997). Dual formulation of four-dimensional variational assimilation, *Q. J. R. Meteorol. Soc.* 123, 2449–2461.
12. Creal D. (2012). A survey of sequential Monte Carlo methods for economics and finance, *Economet. Rev.* 31(3), 245–296.
13. Del Moral P. (1998). Measure-valued processes and interacting particle systems. Application to nonlinear filtering problems, *Annals of Applied Probability* 8(2), 438–495.
14. Del Moral P. (2004). *Feynman-Kac Formulae.* NY: Springer.
15. Deville M. O., Fischer P. F., and Mund E. H. (2006). *Higher-Order Methods for Incompressible Flow* Oxford UK: Cambridge University Press.
16. Doucet A., de Freitas N. and Gordon N. (eds) (2001). *Sequential Monte Carlo Methods in Practice.* NY: Springer.
17. Doucet A., Godsill S., and Andrieu C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering, *Statistics and Computing* 50, 174–188.
18. Dowd M. (2006), A sequential Monte Carlo approach for marine ecological prediction, *Environmetrics* 17, 435–455.
19. Evensen G. (2007). *Data Assimilation.* NY: Springer.
20. Fletcher R. (1987). *Practical Methods of Optimization.* NY: Wiley.
21. Fournier A., Eymin C., and Alboussiere T. (2007). A case for variational geomagnetic data assimilation: insights from a one-dimensional, nonlinear, and sparsely observed MHD system, *Nonlinear Proc. Geoph.*, 14, 163–180.
22. Gelb A. (1974). *Applied optimal estimation.* MIT Press Cambridge.
23. Gordon N.J., Salmon D.J., and Smith A.F.M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *IEEE Proceedings F on Radar and Signal Processing* 140, 107–113.
24. Jardak M., Navon I.M., and Zupanski M. (2009). Comparison of sequential data assimilation methods for the Kuramoto-Sivashinsky equation, *Int. J. Numer. Methods Fluids*, 62, 374–402.
25. Jentzen A., and Kloeden P.E. (2009). Overcoming the order barrier in the numerical approximation of stochastic partial differential equations with additive space-time noise, *Proc. R. Soc. A*, 465,649–667.
26. Julier S.J., and Uhlmann J.K. (1997). A new extension of the Kalman filter to nonlinear systems, *International Symposium on Aerospace/Defense Sensing, Simulation and Controls*, 3.
27. Kalman R.E. (1960). A New Approach to Linear Filtering and Prediction Theory, *Trans. ASME, Ser. D* (Journal of Basic Engineering) 82, 35–48.
28. Kalman R.E. and Bucy R.S. (1961). New Results in Linear Filtering and Prediction Theory, *Trans. ASME, Ser. D* (Journal of Basic Engineering) 83, 95–108.
29. Kass R.E., Tierny, L., and Kadane, J.B. (1990). The validity of posterior expansions based on Laplace's method, *Bayesian and Likelihood methods in Statistics and Econometrics*.

30. Klauder J., and Petersen W. (1985). Numerical integration of multiplicative-noise stochastic differential equations, *SIAM J. Num. Anal.*, 22, 1153–1166.
31. Kloeden P.E., and Platen E. (1999). *Numerical solution of stochastic differential equations.* NY: Springer.
32. Kuramoto Y., and Tsuzuki T. (1975). On the formation of dissipative structures in reaction-diffusion systems, *Progr. Theoret. Phys.*, 54, 687–699.
33. Kurapov A. L., Egbert G. D., Allen J. S., and Miller R. N. (2007). Representer-based variational data assimilation in a nonlinear model of nearshore circulation, *J. Geophy. Res.* 112, C11019.
34. Lopes H.F., and Tsay R.S. (2011). Particle filters and Bayesian inference in financial econometrics, *J. Forecast.* 30, 168–209.
35. Lord G.J., and Rougemont J. (2004). A numerical scheme for stochastic PDEs with Gevrey regularity, *IMA Journal of Numerical Analysis*, 24, 587–604.
36. Lorenz E.N. (1963). Deterministic nonperiodic flow, *J. Atmos. Sci.*, 20, 131–141.
37. Miller R., Ghil M., and Gauthiez F. (1994). Advanced data assimilation in strongly nonlinear dynamical systems, *J. Atmos. Sci.*, 51, 1037–1056.
38. Miller R., Carter E., and Blue S., (1999). Data assimilation into nonlinear stochastic systems, *Tellus*, 51A, 167–194.
39. Morzfeld M., Tu X., Atkins E., and Chorin A.J. (2012). A random map implementation of implicit filters, *J. Comput. Phys.*, 231, 2049–2066.
40. Morzfeld M. and Chorin A.J. (2012). Implicit particle filtering for models with partial noise, and an application to geomagnetic data assimilation, *Nonlinear Proc. Geophys.*, 19, 365–382.
41. Atkins A., Morzfeld M., and Chorin A. J. (2013). Implicit particle and their connection with variational data assimilation, Monthly weather review, *submitted for publication*, 141(6), 1786–1803.
42. Moselhy T. and Marzouk, Y. (2012). Bayesian inference with optimal maps, *J. Comput. Phys.*, 231(23), 7815–7850.
43. Nocedal J., and Wright S.T. (2006). *Numerical Optimization (Second Edition).* NY: Springer.
44. Parlett B.N. (1998). *The symmetric eigenvalue problem.* Classics in Applied Mathematics, Vol. 20, Society for Industrial and Applied Mathematics, Philadelphia.
45. Pitt M.K., and Shephard N. (1999). Filtering via simulation: auxiliary particle filters, *JASA* 94(446), 590–599.
46. Sivashinsky G. (1977). Nonlinear analysis of hydrodynamic instability in laminar flames, Part I. Derivation of basic equations, *Acta Astronaut.*, 4, 1177–1206.
47. Snyder C., Bengtsson T., Bickel P., and Anderson J. (2008). Obstacles to high-dimensional particle filtering, *Mon. Wea. Rev.*, 136, 4629–4640.
48. Talagrand O., and Courtier P. (1987). Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory, *Q. J. R. Meteorol. Soc.* 113, 1311–1328.
49. Talagrand O. (1997). Assimilation of Observations, an Introduction, *J. R. Meteorol. Soc. of Japan* 75(1), 191–209.
50. Tremolet Y. (2006). Accounting for an imperfect mode in 4D-Var, *Q. J. R. Meteorol. Soc.* 621(132), 2483–2504.
51. van Leeuwen P.J. (2009). Particle filtering in geophysical systems, *Mon. Weather Rev.* 137, 4089–4114.
52. van Leeuwen P.J. (2010). Nonlinear data assimilation in geosciences: an extremely efficient particle filter, *Q. J. R. Meteorol. Soc.* 136(653), 1991–1999.
53. Vanden Eijnden, E. and Weare, J. (2012). Data assimilation in the low noise accurate observation regime, with application to the Kuroshio current, accepted for publication in *Mon. Weather Rev.*.
54. Weare J. (2009). Particle filtering with path sampling and an application to a bimodal ocean current model, *J. Comp. Phys.* 12(228), 4312–4331.
55. Weir B., Spitz Y.H., Miller R.N. (2013). Implicit estimation of ecological model parameters, accepted for publication in B. Math. Biol., 75(2), 223–257.
56. Zupanski D. (1997). A General Weak Constraint Applicable to Operational 4D-VAR Data Assimilation systems, *Q. J. R. Meteorol. Soc.* 125, 2274–2292.

Part II
Linear State-Space Models in Macroeconomics and Finance

Chapter 4
Model Uncertainty, State Uncertainty, and State-Space Models

Yulei Luo, Jun Nie, and Eric R. Young

4.1 Introduction

State-space models have been broadly applied to study macroeconomic and financial problems. For example, they have been applied to model unobserved trends, to model transition from one economic structure to another, to forecasting models, to study wage-rate behaviors, to estimate expected inflation, and to model time-varying monetary reaction functions.

A state-space model typically consists of two equations, a measurement equation which links the observed variables to unobserved state variables and a transition equation which describes the dynamics of the state variables. The Kalman filter, which provides a recursive way to compute the estimator of the unobserved component based on the observed variables, is a useful tool to analyze state-space models.

In this chapter, we show that a classic linear-quadratic-Gaussian (LQG) macroeconomic framework which incorporates two new assumptions can still be analytically solved and explicitly mapped to a state-space representation.[1] The two assumptions we consider are model uncertainty due to concerns for model misspecification (*robustness*, RB) and state uncertainty (SU) due to limited information con-

Y. Luo
Department of Economics, The University of Hong Kong, Pokfulam, Hong Kong
e-mail: yluo@econ.hku.hk

J. Nie (✉)
Federal Reserve Bank of Kansas City, 1 Memorial Drive, Kansas City, MO 64196, USA
e-mail: jun.nie@kc.frb.org

E.R. Young
Department of Economics, University of Virginia, Charlottesville, VA 22904, USA
e-mail: ey2d@virginia.edu

[1] Note that here "linear" means that the state transition equation is linear, "quadratic" means that the objective function is quadratic, and "Gaussian" means that the exogenous innovation is Gaussian.

Y. Zeng and S. Wu (eds.), *State-Space Models: Applications in Economics and Finance*, Statistics and Econometrics for Finance 1, DOI 10.1007/978-1-4614-7789-1_4,

straints (*rational inattention*, RI). We show that the state-space representation of the observable and unobservable can be used to quantify the key parameters by simulating the model. We provide examples on how this framework can be used to study a range of interesting questions in macroeconomics and international economics.

The remainder of the chapter is organized as follows. Section 4.2 presents the general framework. Section 4.3 shows how to introduce the model uncertainty and state uncertainty to this framework. Section 4.4 provides several applications how to apply this framework to address a range of macroeconomic and international questions. In addition, it shows how this framework has a state-space representation. And this state-space representation can be used to quantify the key parameters in different models. Section 4.5 concludes.

4.2 Linear-Quadratic-Gaussian State-Space Models

The LQG framework has been widely used in macroeconomics. This specification leads to the optimal linear regulator problem, for which the Bellman equation can be solved easily using matrix algebra. The general setup is as follows. The objective function has a quadratic form,

$$\max_{\{x_t\}} E_0 \left[\sum_{t=0}^{\infty} \eta^t f(x_t) \right] \tag{4.1}$$

and the maximization is subjected to a linear constraint

$$g(x_t, y_t, y_{t+1}) = 0, \text{ for all } t \tag{4.2}$$

where $g(\cdot)$ is a linear function, x_t is the vector of control variables, and y_t is the vector of state variables.

Example 4.1 (A Small-Open Economy Version of Hall's Permanent Income Model). Let $x_t = \{c_t, b_{t+1}\}$, $y_t = \{b_t, y_t\}$, $f(x_t) = -\frac{1}{2}(\bar{c} - c_t)^2$, $g(x_t, y_t, y_{t+1}) = Rb_t + y_t - c_t - b_{t+1}$, where $\bar{c}$ is the bliss point, c_t is consumption, R is the exogenous and constant gross world interest rate, b_t is the amount of the risk-free foreign bond held at the beginning of period t, and y_t is net income in period t and is defined as output minus investment and government spending. Then this becomes a small-open economy version of Hall's permanent income model in which a representative agent chooses the consumption to maximize his/her utility subject to the exogenous endowments. As the representative agent can borrow from the rest of the world at a risk-free interest rate, the resource constraint need not bind every period. If we remove this assumption, the model goes back to the permanent income model studied in Hall (1978).[2]

[2] We take a small-open economy version of Hall's model as we'll use it to address some small-open economy issues in later sectors.

Example 4.2 (Barro's Tax-Smoothing Model). Barro (1979) proposed a simple rational expectations (RE) tax-smoothing model with only noncontingent debt in which the government spreads the burden of raising distortionary income taxes over time in order to minimize their welfare losses to address these questions.[3] This tax-smoothing hypothesis has been widely used (to address various fiscal policies) and tested. The model also falls well into this linear-quadratic framework.[4] Specifically, let $x_t = \{\tau_t, B_{t+1}\}$, $y_t = \{Y_t, G_t\}$, $f(x_t) = -\frac{1}{2}\tau_t^2$, $g(x_t, y_t, y_{t+1}) = RB_t + G_t - \tau_t Y_t - B_{t+1}$, where $E_0[\cdot]$ is the government's expectation conditional on its available and processed information set at time 0, η is the government's subjective discount factor, τ_t is the tax rate, B_t is the amount of government debt, G_t is government spending, Y_t is real GDP, and R is the gross interest rate. Here we assume that the welfare costs of taxation are proportional to the square of the tax rate.[5]

In general, the number of the state variables in these models can be more than one. But in order to facilitate the introduction of robustness we reduce the above multivariate model with a general exogenous process to a univariate model with iid innovations that can be solved in closed-form. Specifically, following Luo and Young (2010) and Luo, Nie, and Young (2011a), we rewrite the model described by (4.1) and (4.2) as

$$\max_{\{z_t, s_{t+1}\}_{t=0}^{\infty}} \left\{ E_0 \left[\sum_{t=0}^{\infty} \eta^t f(z_t) \right] \right\} \tag{4.3}$$

subject to

$$s_{t+1} = Rs_t - z_t + \zeta_{t+1}, \tag{4.4}$$

where both z_t and s_t are single variables, and ζ_{t+1} is the Gaussian innovation to the state transition equation with mean 0 and variance ω_ζ^2.

For instance, for Example 4.1, the mapping is

$$\begin{aligned} z_t &= c_t, \\ s_t &= b_t + \frac{1}{R} \sum_{j=0}^{\infty} R^{-j} E_t [y_{t+j}], \\ \zeta_{t+1} &= \frac{1}{R} \sum_{j=t+1}^{\infty} \left(\frac{1}{R} \right)^{j-(t+1)} (E_{t+1} - E_t)[y_j]. \end{aligned}$$

[3] It is worth noting that the tax-smoothing hypothesis (TSH) model is an analogy with the permanent income hypothesis (PIH) model in which consumers smooth consumption over time; tax rates respond to permanent changes in the public budgetary burden rather than transitory ones.

[4] For example, see Huang and Lin (1993), Ghosh (1995)

[5] Following Barro (1979), Bohn (1990), and Huang and Lin (1993), we only need to impose the restriction, $f'(\tau) > 0$ and $f''(\tau) > 0$, on the loss function, $f(\tau)$.

And for Example 4.2, the mapping is

$$\begin{aligned} z_t &= \tau_t, \\ s_t &= E_t\left[b_t + \frac{1}{(1+n)\widetilde{R}}\sum_{j=0}^{\infty}\left(\frac{1}{\widetilde{R}}\right)^j g_{t+j}\right], \\ \zeta_{t+1} &= \sum_{j=0}^{\infty}\left(\frac{1}{\widetilde{R}}\right)^{j+1}(E_{t+1}-E_t)\left[g_{t+1+j}\right], \end{aligned}$$

where $\widetilde{R} = R/(1+n)$ is the effective interest rate faced by the government, n is the GDP growth rate, b_t and g_t are government debt and government spending as a ratio of GDP.[6]

Finally, the recursive representation of the above problem is as follows:

$$v(s_t) = \max_{z_t}\{f(z_t) + \eta E_t[v(s_{t+1})]\} \tag{4.5}$$

subject to:

$$s_{t+1} = Rs_t - z_t + \zeta_{t+1}, \tag{4.6}$$

given s_0.

4.3 Incorporating Model Uncertainty and State Uncertainty

In this section we show how to incorporate model uncertainty and state uncertainty into the framework presented in the previous section.

4.3.1 Introducing Model Uncertainty

We focus on the model uncertainty due to a concern for model misspecification (robustness). Hansen and Sargent (1995, 2007a) first introduce robustness (a concern for model misspecification) into economic models. In robust control problems, agents are concerned about the possibility that their model is misspecified in a manner that is difficult to detect statistically; consequently, they choose their decisions as if the subjective distribution over shocks was chosen by a malevolent nature in order to minimize their expected utility (that is, the solution to a robust decision-maker's problem is the equilibrium of a max–min game between the decision-maker and nature). Specifically, a robustness version of the model represented by (4.5) and (4.6) is

$$v(s_t) = \max_{z_t}\min_{v_t}\left\{f(z_t) + \eta\left[\vartheta v_t^2 + E_t[v(s_{t+1})]\right]\right\} \tag{4.7}$$

[6] n is assumed to be constant.

subject to the distorted transition equation (i.e., the worst-case model):

$$s_{t+1} = Rs_t - z_t + \zeta_{t+1} + \omega_\zeta v_t, \tag{4.8}$$

where v_t distorts the mean of the innovation and $\vartheta > 0$ controls how bad the error can be.[7]

4.3.2 Introducing State Uncertainty

The model presented in Sect. 4.3.1 is extended to incorporate state uncertainty in this subsection. It will be seen that state uncertainty will further amplify the effect due to model uncertainty.[8] We consider the model with imperfect state observation (*state uncertainty*) due to finite information-processing capacity (rational inattention or RI). Sims (2003) first introduced RI into economics and argued that it is a plausible method for introducing sluggishness, randomness, and delay into economic models. In his formulation agents have finite Shannon channel capacity, limiting their ability to process signals about the true state of the world. As a result, an impulse to the economy induces only gradual responses by individuals, as their limited capacity requires many periods to discover just how much the state has moved.

Under RI, consumers in the economy face both the usual flow budget constraint and information-processing constraint due to finite Shannon capacity first introduced by Sims (2003). As argued by Sims (2003, 2006), individuals with finite channel capacity cannot observe the state variables perfectly; consequently, they react to exogenous shocks incompletely and gradually. They need to choose the posterior distribution of the true state after observing the corresponding signal. This choice is in addition to the usual consumption choice that agents make in their utility maximization problem.[9]

Following Sims (2003), the consumer's information-processing constraint can be characterized by the following inequality:

$$\mathscr{H}(s_{t+1}|\mathscr{I}_t) - \mathscr{H}(s_{t+1}|\mathscr{I}_{t+1}) \leq \kappa, \tag{4.9}$$

where κ is the consumer's channel capacity, $\mathscr{H}(s_{t+1}|\mathscr{I}_t)$ denotes the entropy of the state prior to observing the new signal at $t+1$, and $\mathscr{H}(s_{t+1}|\mathscr{I}_{t+1})$ is the en-

[7] Formally, this setup is a game between the decision-maker and a malevolent nature that chooses the distortion process v_t. $\vartheta \geq 0$ is a penalty parameter that restricts attention to a limited class of distortion processes; it can be mapped into an entropy condition that implies agents choose rules that are robust against processes which are close to the trusted one. In a later section we will apply an error detection approach to calibrate ϑ.

[8] This will be clearer when we go to the applications in later sections.

[9] More generally, agents choose the joint distribution of consumption and current permanent income subject to restrictions about the transition from prior (the distribution before the current signal) to posterior (the distribution after the current signal). The budget constraint implies a link between the distribution of consumption and the distribution of next period permanent income.

tropy after observing the new signal.[10] The concept of *entropy* is from information theory, and it characterizes the uncertainty in a random variable. The right-hand side of (4.9), being the reduction in entropy, measures the amount of information in the new signal received at $t+1$. Hence, as a whole, (4.9) means that the reduction in the uncertainty about the state variable gained from observing a new signal is bounded from above by κ. Since the ex post distribution of s_t is a normal distribution, $N\left(\widehat{s}_t, \sigma_t^2\right)$, (4.9) can be reduced to

$$\log|\psi_t^2| - \log|\sigma_{t+1}^2| \leq 2\kappa \tag{4.10}$$

where $\widehat{s}_t$ is the conditional mean of the true state, and $\sigma_{t+1}^2 = \text{var}[s_{t+1}|\mathscr{I}_{t+1}]$ and $\psi_t^2 = \text{var}[s_{t+1}|\mathscr{I}_t]$ are the posterior variance and prior variance of the state variable, respectively. To obtain (4.10), we use the fact that the entropy of a Gaussian random variable is equal to half of its logarithm variance plus a constant term.

It is straightforward to show that in the univariate case (4.10) has a unique steady state σ^2.[11] In that steady state the consumer behaves as if observing a noisy measurement which is $s_{t+1}^* = s_{t+1} + \xi_{t+1}$, where ξ_{t+1} is the endogenous noise and its variance $\alpha_t^2 = \text{var}[\xi_{t+1}|\mathscr{I}_t]$ is determined by the usual updating formula of the variance of a Gaussian distribution based on a linear observation:

$$\sigma_{t+1}^2 = \psi_t^2 - \psi_t^2\left(\psi_t^2 + \alpha_t^2\right)^{-1}\psi_t^2. \tag{4.11}$$

Note that in the steady state $\sigma^2 = \psi^2 - \psi^2\left(\psi^2 + \alpha^2\right)^{-1}\psi^2$, which can be solved as $\alpha^2 = \left[\left(\sigma^2\right)^{-1} - \left(\psi^2\right)^{-1}\right]^{-1}$. Note that (4.11) implies that in the steady state $\sigma^2 = \frac{\omega_\xi^2}{\exp(2\kappa)-R^2}$ and $\alpha^2 = \text{var}[\xi_{t+1}] = \frac{\left[\omega_\xi^2 + R^2\sigma^2\right]\sigma^2}{\omega_\xi^2 + \left(R^2-1\right)\sigma^2}$.

We now incorporate state uncertainty due to RI into the RB model proposed in the last section. There are two different ways to do it. The simpler way is to assume that the consumer only has doubts about the process for the shock to permanent income ζ_{t+1}, but trusts his or her regular Kalman filter hitting the endogenous noise (ξ_{t+1}) and updating the estimated state. In the next subsection, we will relax the assumption that the consumer trusts the Kalman filter equation which generates an additional dimension along which the agents in the economy desire robustness.

The RB–RI model is formulated as

$$\widehat{v}(\widehat{s}_t) = \max_{z_t}\min_{v_t}\left\{f(z_t) + \eta E_t\left[\vartheta v_t^2 + \widehat{v}(\widehat{s}_{t+1})\right]\right\}, \tag{4.12}$$

subject to the (budget) constraint

$$s_{t+1} = Rs_t - z_t + \omega_\zeta v_t + \zeta_{t+1} \tag{4.13}$$

[10] We regard κ as a technological parameter. If the base for logarithms is 2, the unit used to measure information flow is a "bit," and for the natural logarithm e the unit is a "nat." 1 nat is equal to $\log_2 e \approx 1.433$ bits.

[11] Convergence requires that $\kappa > \log(R) \approx R - 1$; see Luo and Young (2010) for a discussion.

and the regular Kalman filter equation

$$\widehat{s}_{t+1} = (1-\theta)\left(R\widehat{s}_t - z_t + \omega_\zeta \nu_t\right) + \theta\left(s_{t+1} + \xi_{t+1}\right) \tag{4.14}$$

Notice that $f(z_t)$ is a quadratic function, so the model is in a linear-quadratic form. As to be shown in the next section, we can explicitly solve the optimal choice for control variable z_t and the worst-case shock ν_t. After substituting these two solutions into the transition equations for s_t and $\widehat{s}_t$, it can easily be shown that the model has a state-space representation.

4.3.2.1 Robust Filtering Under RI

It is clear that the Kalman filter under RI, (4.13), is not only affected by the fundamental shock (ζ_{t+1}) but also affected by the endogenous noise (ξ_{t+1}) induced by finite capacity; these noise shocks could be another source of the demand for robustness. We therefore need to consider this demand for robustness in the RB–RI model. By adding the additional concern for robustness developed here, we are able to strengthen the effects of robustness on decisions.[12] Specifically, we assume that the agent thinks that (4.14) is the approximating model. Following Hansen, Peter and Sargent (2007), we surround (4.14) with a set of alternative models to represent a preference for robustness:

$$\widehat{s}_{t+1} = R\widehat{s}_t - z_t + \omega_\eta \nu_t + \eta_{t+1}, \tag{4.15}$$

where

$$\eta_{t+1} = \vartheta R(s_t - \widehat{s}_t) + \vartheta(\zeta_{t+1} + \xi_{t+1}) \tag{4.16}$$

and $E_t\left[\eta_{t+1}\right] = 0$ because the expectation is conditional on the perceived signals and inattentive agents cannot perceive the lagged shocks perfectly.

Under RI the innovation η_{t+1}, (4.16), that the agent distrusts is composed of two MA(∞) processes and includes the entire history of the exogenous income shock and the endogenous noise, $\{\zeta_{t+1}, \zeta_t, \cdots, \zeta_0; \xi_{t+1}, \xi_t, \cdots, \xi_0\}$. The difference between (4.13) and (4.15) is the third term; in (4.13) the coefficient on ν_t is ω_ζ while in (4.15) the coefficient is ω_η; note that with $\theta < 1$ and $R > 1$ it holds that $\omega_\zeta < \omega_\eta$.

The optimizing problem for this RB–RI model can be formulated as follows:

$$\widehat{v}(\widehat{s}_t) = \max_{c_t} \min_{\nu_t} \left\{ f(z_t) + \eta E_t\left[\vartheta \nu_t^2 + \widehat{v}(\widehat{s}_{t+1})\right]\right\} \tag{4.17}$$

subject to (4.15). Equation (4.17) is a standard dynamic programming problem and can be easily solved using the standard procedure.

[12] Luo, Nie, and Young (2011a) use this approach to study the joint dynamics of consumption, income, and the current account.

4.4 Applications

This section provides several applications of the framework developed in Sect. 4.3.[13] In each application, the model can be mapped into the general framework presented in the previous section. Using these examples, we show how this framework can be analytically solved and can be explicitly mapped to a state-space representation (Sect. 4.4.1). We also show that this state-space representation plays an important role in quantifying the model uncertainty and state uncertainty (Sect. 4.4.4). These applications show how model uncertainty (RB) and state uncertainty (RI or imperfect information) alter the results from the standard framework presented in Sect. 4.2.

4.4.1 Explaining Current Account Dynamics

Return in to Example 4.1 in Sect. 4.2. The model is a small-open economy version of the permanent income model. The standard model is represented by (4.5) and (4.6), while the model incorporating model uncertainty and state uncertainty is represented by (4.12)–(4.14). (Notice that $z_t = c_t$ and $f(x_t) = -\frac{1}{2}(\bar{c} - c_t)^2$.)

As shown in Luo et al. (2011a), given ϑ and θ, the consumption function under RB and RI is

$$c_t = \frac{R-1}{1-\Sigma}\widehat{s}_t - \frac{\Sigma \bar{c}}{1-\Sigma}, \tag{4.18}$$

the mean of the worst-case shock is

$$\omega_\eta \nu_t = \frac{(R-1)\Sigma}{1-\Sigma}\widehat{s}_t - \frac{\Sigma}{1-\Sigma}\bar{c}, \tag{4.19}$$

where $\rho_s = \frac{1-R\Sigma}{1-\Sigma} \in (0,1)$, $\Sigma = R\omega_\eta^2/(2\vartheta)$, $\omega_\eta^2 = \text{var}[\eta_{t+1}] = \frac{\theta}{1-(1-\theta)R^2}\omega_\xi^2$.

Substituting (4.19) into (4.13) and combining with (4.14), the observed s_t and unobserved $\widehat{s}_t$ are governed by the following two equations

$$s_t - \widehat{s}_t = \frac{(1-\theta)\zeta_t}{1-(1-\theta)R\cdot L} - \frac{\theta\xi_t}{1-(1-\theta)R\cdot L} \tag{4.20}$$

$$\widehat{s}_{t+1} = \rho_s \widehat{s}_t + \eta_{t+1}. \tag{4.21}$$

where

$$\eta_{t+1} = \theta R(s_t - \widehat{s}_t) + \theta(\zeta_{t+1} + \xi_{t+1}) \tag{4.22}$$

Thus, it is clear to see that (4.20) and (4.21) form a state-space representation the model in which (4.20) is the measurement equation that links the observed variable

[13] These illustrations are based on the research by Luo and Young (2010) and Luo, Nie, and Young (2011a, 2011b, 2011c).

s_t to unobserved variable $\widehat{s_t}$ and (4.21) is the transition equation which describes the dynamics of $\widehat{s_t}$.

Notice that Σ measures the effects of both model uncertainty and state uncertainty, which is bounded by 0 and 1.[14] As argued in Sims (2003), although the randomness in an individual's response to aggregate shocks will be idiosyncratic because it arises from the individual's own information-processing constraint, there is likely a significant common component. The intuition is that people's needs for coding macroeconomic information efficiently are similar, so they rely on common sources of coded information. Therefore, the common term of the idiosyncratic error, $\overline{\xi}_t$, lies between 0 and the part of the idiosyncratic error, ξ_t, caused by the common shock to permanent income, ζ_t. Formally, assume that ξ_t consists of two independent noises: $\xi_t = \overline{\xi}_t + \xi_t^i$, where $\overline{\xi}_t = E^i[\xi_t]$ and ξ_t^i are the common and idiosyncratic components of the error generated by ζ_t, respectively. A single parameter,

$$\lambda = \frac{\text{var}\left[\overline{\xi}_t\right]}{\text{var}[\xi_t]} \in [0,1],$$

can be used to measure the common source of coded information on the aggregate component (or the relative importance of $\overline{\xi}_t$ vs. ξ_t).[15]

Next, we briefly list the facts we focus on (Table 4.1). First, the correlation between the current account and net income is positive but small (and insignificant when detrended with the Hodrick–Prescott (HP) filter). Second, the relative volatility of the current account to net income is smaller in emerging countries than in developed economies, although the difference is not statistically significant when the series are detrended with the HP filter. Third, the persistence of the current account is smaller than that of net income and less persistent in emerging economies. And fourth, the volatility of consumption growth relative to income growth is larger in emerging economies than in developed economies.

Finally, let's compare the model implications, as summarized in Table 4.2. First, we have seen that in this case ($\lambda = 1$ and $\theta = 50\%$) the interaction of RB and RI makes the model fit the data quite well along dimensions (3) and (4), while also quantitatively improving the model's predictions along dimensions (1) and (2). Second, this improvement does not preclude the model from matching the first two dimensions as well (i.e., the contemporaneous correlation between the current account and net income and the volatility of the current account). For example, holding λ equal to 1 and further reducing θ can generate a smaller contemporaneous correlation between the current account and net income which is closer to the data. And holding $\theta = 50\%$ and reducing λ to 0.1 can make the relative volatility of the current account to net income very close to the data.

[14] See Luo, Nie, and Young (2011a) for the proof.

[15] It is worth noting that the special case that $\lambda = 1$ can be viewed as a representative-agent model in which we do not need to discuss the aggregation issue.

Table 4.1 Emerging vs. developed countries (averages)

A: Emerging vs. developed countries (HP filter)		
$\sigma(y)/\mu(y)$	4.09(0.23)	1.98(0.09)
$\sigma(\Delta y)/\mu(y)$	4.28(0.23)	1.89(0.07)
$\rho(y_t, y_{t-1})$	0.53(0.03)	0.66(0.02)
$\rho(\Delta y_t, \Delta y_{t-1})$	0.28(0.05)	0.46(0.03)
$\sigma(c)/\sigma(y)$	0.74(0.02)	0.59(0.02)
$\sigma(\Delta c)/\sigma(\Delta y)$	0.71(0.02)	0.59(0.02)
$\sigma(ca)/\sigma(y)$	0.79(0.03)	0.85(0.04)
$\rho(c, y)$	0.85(0.02)	0.78(0.02)
$\rho(ca_t, ca_{t-1})$	0.30(0.05)	0.41(0.03)
$\rho(ca, y)$	−0.59(0.05)	−0.35(0.04)
$\rho\left(\frac{ca}{y}, y\right)$	−0.54(0.04)	−0.36(0.04)
B: Emerging vs. developed countries (linear filter)		
$\sigma(y)/\mu(y)$	7.97(0.40)	4.79(0.22)
$\sigma(\Delta y)/\mu(y)$	4.28(0.23)	1.89(0.07)
$\rho(y_t, y_{t-1})$	0.79(0.02)	0.89(0.01)
$\rho(\Delta y_t, \Delta y_{t-1})$	0.28(0.05)	0.46(0.03)
$\sigma(c)/\sigma(y)$	0.72(0.02)	0.58(0.02)
$\sigma(\Delta c)/\sigma(\Delta y)$	0.71(0.02)	0.59(0.02)
$\sigma(ca)/\sigma(y)$	0.54(0.03)	0.65(0.04)
$\rho(c, y)$	0.88(0.02)	0.85(0.02)
$\rho(ca_t, ca_{t-1})$	0.53(0.04)	0.71(0.02)
$\rho(ca, y)$	−0.17(0.06)	−0.08(0.05)
$\rho\left(\frac{ca}{y}, y\right)$	−0.32(0.05)	−0.20(0.04)

Table 4.2 Implications of different models (emerging countries)

	Data	RE	RB	RB + RI ($\theta = 0.9$)	RB + RI ($\theta = 0.8$)	RB + RI ($\theta = 0.7$)	RB + RI ($\theta = 0.5$)
				($\lambda = 1$)			
$\rho(ca, y)$	0.13	1.00	0.62	0.57	0.56	0.56	0.58
$\rho(ca_t, ca_{t-1})$	0.53	0.80	0.74	0.57	0.50	0.45	0.36
$\sigma(ca)/\sigma(y)$	0.80	0.71	0.49	0.52	0.55	0.59	0.79
$\sigma(\Delta c)/\sigma(\Delta y)$	1.35	0.28	0.90	0.89	0.89	0.91	1.36
				($\lambda = 0.5$)			
$\rho(ca, y)$	0.13	1.00	0.62	0.59	0.58	0.59	0.64
$\rho(ca_t, ca_{t-1})$	0.53	0.80	0.74	0.63	0.59	0.55	0.46
$\sigma(ca)/\sigma(y)$	0.80	0.71	0.49	0.50	0.52	0.53	0.64
$\sigma(\Delta c)/\sigma(\Delta y)$	1.35	0.28	0.90	0.85	0.81	0.79	0.99
				($\lambda = 0.1$)			
$\rho(ca, y)$	0.13	1.00	0.62	0.61	0.60	0.61	0.67
$\rho(ca_t, ca_{t-1})$	0.53	0.80	0.74	0.67	0.64	0.62	0.56
$\sigma(ca)/\sigma(y)$	0.80	0.71	0.49	0.49	0.50	0.51	0.57
$\sigma(\Delta c)/\sigma(\Delta y)$	1.35	0.28	0.90	0.84	0.79	0.75	0.82

4.4.2 Resolving the International Consumption Puzzle

The same framework can be used to address an old puzzle in the international economics literature. That is, the cross-country consumption correlations are very low in the data (lower than the cross-country correlations of outputs) while standard models imply the opposite.[16]

To show the flexibility of the general framework summarized by (4.5) and (4.6), we slightly deviate from the assumption we used in the previous subsection (Example 4.1) to introduce state uncertainty. We assume that consumers in the model economy cannot observe the true state s_t perfectly and only observe the noisy signal

$$s_t^* = s_t + \xi_t, \tag{4.23}$$

when making decisions, where ξ_t is the iid Gaussian noise due to imperfect observations. The specification in (4.23) is standard in the signal extraction literature and captures the situation where agents happen or choose to have imperfect knowledge of the underlying shocks.[17] Since imperfect observations on the state lead to welfare losses, agents use the processed information to estimate the true state.[18] Specifically, we assume that households use the Kalman filter to update the perceived state $\widehat{s}_t = E_t[s_t]$ after observing new signals in the steady state:

$$\widehat{s}_{t+1} = (1-\theta)(R\widehat{s}_t - c_t) + \theta(s_{t+1} + \xi_{t+1}), \tag{4.24}$$

where θ is the Kalman gain (i.e., the observation weight).[19]

In the signal extraction problem, the Kalman gain can be written as

$$\theta = \Upsilon \Lambda^{-1}, \tag{4.25}$$

where Υ is the steady state value of the conditional variance of s_{t+1}, $\text{var}_{t+1}[s_{t+1}]$, and is the variance of the noise, $\Lambda = \text{var}_t[\xi_{t+1}]$. Υ and Λ are linked by the following equation which updates the conditional variance in the steady state:

$$\Lambda^{-1} = \Upsilon^{-1} - \Psi^{-1}, \tag{4.26}$$

where Ψ is the steady state value of the ex ante conditional variance of s_{t+1}, $\Psi_t = \text{var}_t[s_{t+1}]$.

[16] For example, Backus, Kehoe, and Kydland (1992) solve a two-country real business cycles model and argue that the puzzle that empirical consumption correlations are actually lower than output correlations is the most striking discrepancy between theory and data.

[17] For example, Muth (1960), Lucas (1972), Morris and Shin (2002), and Angeletos and La'O (2009). It is worth noting that this assumption is also consistent with the rational inattention idea that ordinary people only devote finite information-processing capacity to processing financial information and thus cannot observe the states perfectly.

[18] See Luo (2008) for details about the welfare losses due to information imperfections within the partial equilibrium permanent income hypothesis framework.

[19] Note that θ measures how much uncertainty about the state can be removed upon receiving the new signals about the state.

Multiplying ω_ζ^2 on both sides of (4.26) and using the fact that $\Psi = R^2\Upsilon + \omega_\zeta^2$, we have

$$\omega_\zeta^2 \Lambda^{-1} = \omega_\zeta^2 \Upsilon^{-1} - \left[R^2 \left(\omega_\zeta^2 \Upsilon^{-1} \right)^{-1} + 1 \right]^{-1}, \tag{4.27}$$

where $\omega_\zeta^2 \Upsilon^{-1} = \left(\omega_\zeta^2 \Lambda^{-1} \right) \left(\Lambda \Upsilon^{-1} \right)$.

Define SNR as $\pi = \omega_\zeta^2 \Lambda^{-1}$. We obtain the following equality linking SNR (π) and the Kalman gain (θ):

$$\pi = \theta \left(\frac{1}{1-\theta} - R^2 \right). \tag{4.28}$$

Solving for θ from the above equation yields

$$\theta = \frac{-(1+\pi) + \sqrt{(1+\pi)^2 + 4R^2(\pi + R^2)}}{2R^2}, \tag{4.29}$$

where we omit the negative values of θ because both Υ and Λ must be positive. Note that given π, we can pin down Λ using $\pi = \omega_\zeta^2 \Lambda^{-1}$ and Υ using (4.25) and (4.29).

Combining (4.4) with (4.24), we obtain the following equation governing the perceived state $\widehat{s}_t$:

$$\widehat{s}_{t+1} = R\widehat{s}_t - c_t + \eta_{t+1}, \tag{4.30}$$

where

$$\eta_{t+1} = \theta R (s_t - \widehat{s}_t) + \theta (\zeta_{t+1} + \xi_{t+1}) \tag{4.31}$$

is the innovation to the mean of the distribution of perceived permanent income,

$$s_t - \widehat{s}_t = \frac{(1-\theta)\zeta_t}{1-(1-\theta)R \cdot L} - \frac{\theta \xi_t}{1-(1-\theta)R \cdot L} \tag{4.32}$$

is the estimation error where L is the lag operator and $E_t[\eta_{t+1}] = 0$. Note that η_{t+1} can be rewritten as

$$\eta_{t+1} = \theta \left[\left(\frac{\zeta_{t+1}}{1-(1-\theta)R \cdot L} \right) + \left(\xi_{t+1} - \frac{\theta R \xi_t}{1-(1-\theta)R \cdot L} \right) \right], \tag{4.33}$$

where $\omega_\xi^2 = \text{var}[\xi_{t+1}] = \frac{1}{\theta} \frac{1}{1/(1-\theta)-R^2} \omega_\zeta^2$. Expression (4.33) clearly shows that the estimation error reacts to the fundamental shock positively, while it reacts to the noise shock negatively. In addition, the importance of the estimation error is decreasing with θ. More specifically, as θ increases, the first term in (4.33) becomes less important because $(1-\theta)\zeta_t$ in the numerator decreases, and the second term also becomes less important because the importance of ξ_t decreases as θ increases.[20]

[20] Note that when $\theta = 1$, $var[\xi_{t+1}] = 0$.

Although the assumption we use to introduce state uncertainty is different, the general framework is still the same. More importantly, the solution strategy is also the same. Basically, we can explicitly derive the expressions for consumption and the worst-case shock and then substitute them into (4.30). Together with (4.32), it forms a state-space representation of the model.

Table 4.3 reports the implied consumption correlations (between the domestic country and ROW) between the RE, RB, and RB–SU models. There are two interesting observations in the table. First, given the degrees of RB and SU (θ), $\text{corr}(c_t, c_t^*)$ decreases with the aggregation factor (λ). Second, when λ is positive (even if it is very small, e.g., 0.1 in the table), $\text{corr}(c_t, c_t^*)$ is decreasing with the degree of inattention (i.e., increasing with θ). The intuition is that when there are common noises, the effect of the noises could dominate the effect of gradual consumption adjustments on cross-country consumption correlations.

Table 4.3 Theoretical $\text{corr}(c, c^*)$ from different models

	Data	RE	RB	RB + SU ($\theta = 0.9$)	RB + SU ($\theta = 0.6$)	RB + SU ($\theta = 0.3$)
Canada						
($\lambda = 1$)	0.38	0.41	0.33	0.27	0.17	0.12
($\lambda = 0.5$)	0.38	0.41	0.33	0.31	0.26	0.23
($\lambda = 0.1$)	0.38	0.41	0.33	0.32	0.32	0.32
Italy						
($\lambda = 1$)	0.25	0.54	0.50	0.42	0.27	0.19
($\lambda = 0.5$)	0.25	0.54	0.50	0.48	0.41	0.36
($\lambda = 0.1$)	0.25	0.54	0.50	0.50	0.50	0.49
UK						
($\lambda = 1$)	0.21	0.69	0.45	0.38	0.25	0.17
($\lambda = 0.5$)	0.21	0.69	0.45	0.44	0.38	0.32
($\lambda = 0.1$)	0.21	0.69	0.45	0.46	0.46	0.45
France						
($\lambda = 1$)	0.46	0.51	0.49	0.40	0.26	0.18
($\lambda = 0.5$)	0.46	0.51	0.49	0.46	0.40	0.34
($\lambda = 0.1$)	0.46	0.51	0.49	0.49	0.48	0.48
Germany						
($\lambda = 1$)	0.04	0.45	0.40	0.33	0.22	0.15
($\lambda = 0.5$)	0.04	0.45	0.40	0.38	0.33	0.29
($\lambda = 0.1$)	0.04	0.45	0.40	0.40	0.40	0.40

As we can see from Table 4.3, for all the countries we consider here, introducing SU into the RB model can make the model better fit the data on consumption correlations at many combinations of the parameter values. For example, for Italy, when $\theta = 60\%$ (60% of the uncertainty is removed upon receiving a new signal about the innovation to permanent income) and $\lambda = 1$, the RB–SU model predicts that $\text{corr}(c_t, c_t^*) = 0.27$, which is very close to the empirical counterpart,

0.25.[21] For France, when $\theta = 90\%$ and $\lambda = 0.5$, the RB–SU model predicts that $\text{corr}(c_t, c_t^*) = 0.46$, which exactly matches the empirical counterpart. Note that a small value of θ can be rationalized by examining the welfare effects of finite channel capacity.[22]

4.4.3 Other Possible Applications

This linear-quadratic framework which incorporates model uncertainty (due to RB) and state uncertainty (either due to RI or imperfect information) can be applied to study other topics as well. We will briefly discuss several more in this subsection. We will not write down the model equations again as we have shown in Sects. 4.2 and 4.3 that these models can be written in a similar framework.

First, as shown in the previous section, model uncertainty due to RB is particularly promising and interesting for studying emerging and developed small-open economies because it has the potential to generate the *different* joint behaviors of consumption and current accounts observed *across the two groups of economies*. This novel theoretical contribution can also be used to address the observed U.S. Great Moderation in which the volatility of output changed after 1984. Specifically, this feature can be used to address different macroeconomic dynamics (e.g., consumption volatility) given that output volatility changed *before and after* the Great Moderation.

Second, inventories in the standard production smoothing model can be viewed as a stabilizing factor. Cost-minimizing firms facing sales fluctuations smooth production by adjusting their inventories. As a result, production is less volatile than sales. However, in the data, real GDP is more volatile than final sales measured by real GDP minus inventory investment. The existing studies find supportive evidence that real GNP is more volatile than final sales in industry-level data. The key question is that if cost-minimizing firms use inventories to smooth their production, why is production more volatile than sales? In the future research, we can examine whether introducing RB can help improve the prediction of an otherwise standard production smoothing model with inventories on the joint dynamics of inventories, production, and sales.

Third, as shown in Luo, Nie, and Young (2011c), the standard tax-smoothing model proposed by Barro (1979) cannot explain the observed volatility of the tax rates and the joint behavior of the government spending and deficits. As shown in Example 4.2 of Sect. 4.2, the tax-smoothing model used in the literature falls well into the linear-quadratic framework we described. It is easy to show that the same mechanisms presented in Sects. 4.4.1 and 4.4.2 will work in the tax-smoothing

[21] For example, Adam (2007) found $\theta = 40\%$ based on the response of aggregate output to monetary policy shocks. Luo (2008) found that if $\theta = 50\%$, the otherwise standard permanent income model can generate realistic relative volatility of consumption to labor income.

[22] See Luo and Young (2010) for details about the welfare losses due to imperfect observations in the RB model; they are uniformly small.

model which incorporates model uncertainty and state uncertainty. Specifically, Luo, Nie, and Young (2011c) shows that it can help the standard model to better explain the relative volatility of the changes in tax rates to government spending and the comovement between government deficits and spending in the data.

Fourth, this framework can also be extended to study optimal monetary policy under model uncertainty and imperfect state observation. A central bank sets nominal interest rate to minimize prices fluctuations and the output gap (i.e., the deviation of the output from the potential maximum output level). Following the literature, the standard objective function of a central bank can be described by a quadratic function which is a weighted average of the deviation of the inflation from its target and the output gap.[23] Therefore, the framework presented in this chapter can be used to study optimal monetary policy when a central bank has concerns that the model is misspecified and it faces noisy data when making decisions.[24]

4.4.4 Quantifying Model Uncertainty

One remaining question from previous sections is how to quantify the incorporated degree of model uncertainty.[25] In this section, we will show how to use the state-space representation of s_t and $\widehat{s_t}$ to simulate the model and calibrate the key parameters. For convenience and consistence, we continue to use the small-open economy model described in Example 4.1 as the illustration example.

Let model A denote the approximating model and model B be the distorted model. Define p_A as

$$p_A = \text{Prob}\left(\log\left(\frac{L_A}{L_B}\right) < 0 \middle| A\right), \tag{4.34}$$

where $\log\left(\frac{L_A}{L_B}\right)$ is the log-likelihood ratio. When model A generates the data, p_A measures the probability that a likelihood ratio test selects model B. In this case, we call p_A the probability of the model detection error. Similarly, when model B generates the data, we can define p_B as

$$p_B = \text{Prob}\left(\log\left(\frac{L_A}{L_B}\right) > 0 \middle| B\right). \tag{4.35}$$

[23] For example, see Svensson (2000), Gali and Monacelli (2005), Walsh (2004), Leitemo and Soderstrom (2008a; 2008b).

[24] For the examples of the model equations describing the inflation and output dynamics in a closed economy, see Leitemo and Soderstrom (2008a).

[25] This includes the two versions of the model presented in previous sections which incorporates the model uncertainty due to RB: one uses the regular Kalman filter; the other one assumes that the agent does not trust the Kalman filter either (robust filtering).

Following Hansen, Sargent, and Wang (2002) and Hansen and Sargent (2007b), the detection error probability, p, is defined as the average of p_A and p_B:

$$p(\vartheta) = \frac{1}{2}(p_A + p_B), \tag{4.36}$$

where ϑ is the robustness parameter used to generate model B. Given this definition, we can see that $1-p$ measures the probability that econometricians can distinguish the approximating model from the distorted model.

Now we show how to compute the model detection error probability due to model uncertainty and state uncertainty.

In the model with both the RB preference and RI, the approximating model can be written as

$$s_{t+1} = Rs_t - c_t + \zeta_{t+1}, \tag{4.37}$$
$$\widehat{s}_{t+1} = (1-\theta)(R\widehat{s}_t - c_t) + \theta(s_{t+1} + \lambda\xi_{t+1}), \tag{4.38}$$

and the distorted model is

$$s_{t+1} = Rs_t - c_t + \zeta_{t+1} + \omega_\zeta \nu_t, \tag{4.39}$$
$$\widehat{s}_{t+1} = (1-\theta)\left(R\widehat{s}_t - c_t + \omega_\zeta \nu_t\right) + \theta(s_{t+1} + \lambda\xi_{t+1}), \tag{4.40}$$

where we remind the reader that $\lambda = \frac{\text{var}[\overline{\xi}_t]}{\text{var}[\xi_t]} \in [0,1]$ is the parameter measuring the relative importance of $\overline{\xi}_t$ vs. ξ_t.

After substituting the consumption function and the worst-case shock expression into (4.38) and (4.40) we can put the equations in the following matrix form:

$$\begin{bmatrix} s_{t+1} \\ \widehat{s}_{t+1} \end{bmatrix} = \begin{bmatrix} R & -\frac{R-1}{1-\Sigma} \\ \theta R & \frac{1-R+R(1-\theta)(1-\Sigma)}{1-\Sigma} \end{bmatrix} \begin{bmatrix} s_t \\ \widehat{s}_t \end{bmatrix} + \begin{bmatrix} \zeta_{t+1} \\ \theta(\zeta_{t+1} + \lambda\xi_{t+1}) \end{bmatrix} + \begin{bmatrix} \frac{\Sigma}{1-\Sigma}\overline{c} \\ \frac{\Sigma}{1-\Sigma}\overline{c} \end{bmatrix} \tag{4.41}$$

and

$$\begin{bmatrix} s_{t+1} \\ \widehat{s}_{t+1} \end{bmatrix} = \begin{bmatrix} R & -(R-1) \\ \theta R & 1-\theta R \end{bmatrix} \begin{bmatrix} s_t \\ \widehat{s}_t \end{bmatrix} + \begin{bmatrix} \zeta_{t+1} \\ \theta(\zeta_{t+1} + \lambda\xi_{t+1}) \end{bmatrix}. \tag{4.42}$$

Given the RB parameter, ϑ, and RI parameter, θ, we can compute p_A and p_B and thus the detection error probability as follows:

1. Simulate $\{s_t\}_{t=0}^T$ using (4.41) and (4.42) a large number of times. The number of periods used in the simulation, T, is set to be the actual length of the data for each individual country.
2. Count the number of times that $\log\left(\frac{L_A}{L_B}\right) < 0 \Big| A$ and $\log\left(\frac{L_A}{L_B}\right) > 0 \Big| B$ are each satisfied.
3. Determine p_A and p_B as the fractions of realizations for which $\log\left(\frac{L_A}{L_B}\right) < 0 \Big| A$ and $\log\left(\frac{L_A}{L_B}\right) > 0 \Big| B$, respectively.

4.4.5 Discussions: Risk-Sensitivity and Robustness Under Rational Inattention

Risk-sensitivity (RS) was first introduced into the LQG framework by Jacobson (1973) and extended by Whittle (1981). Exploiting the recursive utility framework of Hansen and Sargent (1995) introduce discounting into the RS specification and show that the resulting decision rules are time-invariant. In the RS model agents effectively compute expectations through a distorted lens, increasing their effective risk aversion by overweighting negative outcomes. The resulting decision rules depend explicitly on the variance of the shocks, producing precautionary savings, but the value functions are still quadratic functions of the states.[26] In Hansen et al. (1999) and Hansen, Peter and Sargent (2007), they interpret the RS preference in terms of a concern about model uncertainty (robustness or RB) and argue that RS introduces precautionary savings because RS consumers want to protect themselves against model specification errors.

Following Luo and Young (2010), we formulate an RI version of risk-sensitive control based on recursive preferences with an exponential certainty equivalence function as follows:

$$\widehat{v}(\widehat{s}_t) = \max_{c_t} \left\{ -\frac{1}{2}(c_t - \overline{c})^2 + \eta \mathscr{R}_t \left[\widehat{v}(\widehat{s}_{t+1}) \right] \right\} \tag{4.43}$$

subject to the budget constraint (4.6) and the Kalman filter equation (4.14). The distorted expectation operator is now given by

$$\mathscr{R}_t \left[\widehat{v}(\widehat{s}_{t+1}) \right] = -\frac{1}{\alpha} \log E_t \left[\exp\left(-\alpha \widehat{v}(\widehat{s}_{t+1}) \right) \right],$$

where $s_0 | \mathscr{I}_0 \sim N\left(\widehat{s}_0, \overline{\sigma}^2\right)$, $\widehat{s}_t = E_t[s_t]$ is the perceived state variable, θ is the optimal weight on the new observation of the state, and ξ_{t+1} is the endogenous noise. The optimal choice of the weight θ is given by $\theta(\kappa) = 1 - 1/\exp(2\kappa) \in [0,1]$. The following proposition summarizes the solution to the RI–RS model when $\eta R = 1$:

Proposition 4.1. *Given finite channel capacity κ and the degree of risk-sensitivity α, the consumption function of a risk-sensitive consumer under RI*

$$c_t = \frac{R-1}{1-\Pi}\widehat{s}_t - \frac{\Pi \overline{c}}{1-\Pi}, \tag{4.44}$$

[26] Formally, one can view risk-sensitive agents as ones who have non-state-separable preferences, as in, but with a value for the intertemporal elasticity of substitution equal to one.

where

$$\Pi = R\alpha\omega_\eta^2 \in (0,1), \tag{4.45}$$

$$\omega_\eta^2 = \mathrm{var}\left[\eta_{t+1}\right] = \frac{\theta}{1-(1-\theta)R^2}\omega_\xi^2, \tag{4.46}$$

η_{t+1} *is defined in (4.16), and* $\theta(\kappa) = 1 - 1/\exp(2\kappa)$.

Comparing (4.18) and (4.44), it is straightforward to show that it is impossible to distinguish between RB and RS under RI using only consumption-savings decisions.

Proposition 4.2. *Let the following expression hold:*

$$\alpha = \frac{1}{2\vartheta}. \tag{4.47}$$

Then consumption and savings are identical in the RS–RI and RB–RI models.

Note that (4.47) is *exactly the same as* the observational equivalence condition obtained in the full-information RE model (see Backus, Routledge, and Zin 2004). That is, under the assumption that the agent distrusts the Kalman filter equation, the OE result obtained under full-information RE still holds under RI.[27]

HST (1999) show that as far as the *quantity* observations on consumption and savings are concerned, the robustness version ($\vartheta > 0$ or $\alpha > 0, \widetilde{\eta}$) of the PIH model is observationally equivalent to the standard version ($\vartheta = \infty$ or $\alpha = 0, \eta = 1/R$) of the PIH model for a unique pair of discount factors.[28] The intuition is that introducing a preference for risk-sensitivity (RS) or a concern about robustness (RB) increases savings in the same way as increasing the discount factor, so that the discount factor can be changed to offset the effect of a change in RS or RB on consumption and investment.[29] Alternatively, holding all parameters constant except the pair (α, η), the RI version of the PIH model with RB consumers ($\vartheta > 0$ and $\eta R = 1$) is observationally equivalent to the standard RI version of the model ($\vartheta = \infty$ and $\widetilde{\eta} > 1/R$).

[27] Note that the OE becomes

$$\frac{\alpha\theta}{1-(1-\theta)R^2} = \frac{1}{2\vartheta},$$

if we assume that the agents distrust the income process hitting the budget constraint, but trust the RI-induced noise hitting the Kalman filtering equation.

[28] HST (1999) derive the observational equivalence result by fixing all parameters, including R, except for the pair (α, η).

[29] As shown in HST (1999), the two models have different implications for asset prices because continuation valuations would alter as one alters (α, η) within the observationally equivalent set of parameters.

Proposition 4.3. *Let*

$$\widetilde{\eta} = \frac{1}{R}\frac{1 - R\omega_\eta^2/(2\vartheta)}{1 - R^2\omega_\eta^2/(2\vartheta)} = \frac{1}{R}\frac{1 - R\alpha\omega_\eta^2}{1 - R^2\alpha\omega_\eta^2} > \frac{1}{R}.$$

Then consumption and savings are identical in the RI, RB–RI, and RS–RI models.

4.5 Conclusions

In this chapter we show that a state-space representation can be explicitly derived from a classic macroeconomic framework which has incorporated model uncertainty due to concerns for model misspecification (robustness or RB) and state uncertainty due to limited information constraints (rational inattention or RI). We show the state-space representation can also be used to quantify the key model parameters. Several applications are also provided to show how this general framework can be used to address a range of interesting economic questions.

Acknowledgments Luo thanks the Hong Kong GRF under grant No. 748209 and 749510 and HKU seed funding program for basic research for financial support. All errors are the responsibility of the authors. The views expressed here are the opinions of the authors only and do not necessarily represent those of the Federal Reserve Bank of Kansas City or the Federal Reserve System.

Appendix

A.1 Solving the Current Account Model Explicitly Under Model Uncertainty

To solve the Bellman equation (4.7), we conjecture that

$$v(s_t) = -As_t^2 - Bs_t - C,$$

where A, B, and C are undetermined coefficients. Substituting this guessed value function into the Bellman equation gives

$$-As_t^2 - Bs_t - C = \max_{c_t}\min_{v_t}\left\{-\frac{1}{2}(\bar{c} - c_t)^2 + \eta E_t\left[\vartheta v_t^2 - As_{t+1}^2 - Bs_{t+1} - C\right]\right\}. \tag{4.48}$$

We can do the min and max operations in any order, so we choose to do the minimization first. The first-order condition for v_t is

$$2\vartheta v_t - 2AE_t\left[\omega_\zeta v_t + Rs_t - c_t\right]\omega_\zeta - B\omega_\zeta = 0,$$

which means that

$$v_t = \frac{B + 2A\left(Rs_t - c_t\right)}{2\left(\vartheta - A\omega_\zeta^2\right)}\omega_\zeta. \tag{4.49}$$

Substituting (4.49) back into (4.48) gives

$$-As_t^2 - Bs_t - C = \max_{c_t}\left\{-\frac{1}{2}\left(\bar{c} - c_t\right)^2 + \eta E_t\left[\vartheta\left[\frac{B + 2A\left(Rs_t - c_t\right)}{2\left(\vartheta - A\omega_\zeta^2\right)}\omega_\zeta\right]^2 - As_{t+1}^2 - Bs_{t+1} - C\right]\right\},$$

where

$$s_{t+1} = Rs_t - c_t + \zeta_{t+1} + \omega_\zeta v_t.$$

The first-order condition for c_t is

$$\left(\bar{c} - c_t\right) - 2\eta\vartheta\frac{A\omega_\zeta}{\vartheta - A\omega_\zeta^2}v_t + 2\eta A\left(1 + \frac{A\omega_\zeta^2}{\vartheta - A\omega_\zeta^2}\right)\left(Rs_t - c_t + \omega_\zeta v_t\right) + \eta B\left(1 + \frac{A\omega_\zeta^2}{\vartheta - A\omega_\zeta^2}\right) = 0.$$

Using the solution for v_t the solution for consumption is

$$c_t = \frac{2A\eta R}{1 - A\omega_\zeta^2/\vartheta + 2\eta A}s_t + \frac{\bar{c}\left(1 - A\omega_\zeta^2/\vartheta\right) + \eta B}{1 - A\omega_\zeta^2/\vartheta + 2\eta A}. \tag{4.50}$$

Substituting the above expressions into the Bellman equation gives

$$\begin{aligned}
&-As_t^2 - Bs_t - C \\
&= -\frac{1}{2}\left(\frac{2A\eta R}{1 - A\omega_\zeta^2/\vartheta + 2\eta A}s_t + \frac{-2\eta A\bar{c} + \eta B}{1 - A\omega_\zeta^2/\vartheta + 2\eta A}\right)^2 \\
&+ \frac{\eta\vartheta\omega_\zeta^2}{\left(2\left(\vartheta - A\omega_\zeta^2\right)\right)^2}\left(\frac{2AR\left(1 - A\omega_\zeta^2/\vartheta\right)}{1 - A\omega_\zeta^2/\vartheta + 2\eta A}s_t + B - \frac{2\bar{c}\left(1 - A\omega_\zeta^2/\vartheta\right)A + 2\eta AB}{1 - A\omega_\zeta^2/\vartheta + 2\eta A}\right)^2 \\
&- \eta A\left(\left(\frac{R}{1 - A\omega_\zeta^2/\vartheta + 2\eta A}s_t - \frac{-B\omega_\zeta^2/\vartheta + 2c + 2B\eta}{2\left(1 - A\omega_\zeta^2/\vartheta + 2\eta A\right)}\right)^2 + \omega_\zeta^2\right) \\
&- \eta B\left(\frac{R}{1 - A\omega_\zeta^2/\vartheta + 2\eta A}s_t - \frac{-B\omega_\zeta^2/\vartheta + 2c + 2B\eta}{2\left(1 - A\omega_\zeta^2/\vartheta + 2\eta A\right)}\right) - \eta C.
\end{aligned}$$

Given $\eta R = 1$, collecting and matching terms, the constant coefficients turn out to be

$$A = \frac{R(R-1)}{2 - R\omega_\zeta^2/\vartheta}, \tag{4.51}$$

$$B = -\frac{R\overline{c}}{1 - R\omega_\zeta^2/(2\vartheta)}, \tag{4.52}$$

$$C = \frac{R}{2\left(1 - R\omega_\zeta^2/2\vartheta\right)}\omega_\zeta^2 + \frac{R}{2\left(1 - R\omega_\zeta^2/2\vartheta\right)(R-1)}\overline{c}^2. \tag{4.53}$$

Substituting (4.51) and (4.52) into (4.50) yields the consumption function. Substituting (4.53) into the current account identity and using the expression for s_t yields the expression for the current account.

References

1. Adam, Klaus (2007), "Optimal monetary policy with imperfect common knowledge," *Journal of Monetary Economics*, Elsevier 54(2), 267–301.
2. Angeletos, George-Marios and Jennifer La'O (2009), "Noisy Business Cycles," *NBER Macroeconomics Annual 2009*, 319–378.
3. Backus, David K., Patrick J. Kehoe, and Finn E. Kydland (1992), "International Real Business Cycles," *Journal of Political Economy* 101(4), 745–775.
4. Backus, David K., Bryan R. Routledge, and Stanley E. Zin (2004), "Exotic Preferences for-Macroeconomists," *NBER Macroeconomics Annual 2004*, 319–414.
5. Bohn, Henning (1990), Tax Smoothing with Financial Instruments, American Economic Review 80(5), 1217–1230.
6. Gali, J., Monacelli, T., (2005), "Monetary policy and exchange rate volatility in a small open economy," *Review of Economic Studies* 72(3), 702–734.
7. Barro, Robert J. (1979), "On the Determination of the Public Debt," *Journal of Political Economy* 87, 940–971.
8. Ghosh, Atish R. (1995), "Intertemporal Tax-smoothing and the Government Budget Surplus: Canada and the United States," *Journal of Money, Credit and Banking* 27, 1033–1045.
9. Hall, Robert E. (1978), "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy* 86(6), 971–87.
10. Hansen, Lars Peter and Thomas J. Sargent (1995), "Discounted Linear Exponential Quadratic Gaussian Control," *IEEE Transactions on Automatic Control* 40, pp. 968–971.
11. Hansen, Lars Peter, Thomas J. Sargent, and Thomas D. Tallarini, Jr. (1999), "Robust Permanent Income and Pricing," *Review of Economic Studies* 66(4), 873–907.
12. Hansen, Lars Peter and Thomas J. Sargent (2007), Robustness, Princeton University Press.
13. Hansen, Lars Peter and Thomas J. Sargent (2007a), *Robustness*, Princeton University Press.
14. Hansen, Lars Peter and Thomas J. Sargent (2007b), "Robust Estimation and Control without Commitment," *Journal of Economic Theory* 136(1), 1–27.
15. Hansen, Lars Peter, Thomas J. Sargent, and Neng Wang (2002), "Robust Permanent Income and Pricing with Filtering," *Macroeconomic Dynamics* 6(1), 40–84.
16. Huang, Chao-Hsi, Lin, Kenneth S. (1993), "Deficits, Government Expenditures and Tax Smoothing in the United States: 1929–1988," *Journal of Monetary Economics* 31, 317–339.
17. Jacobson, David H. (1973), "Optimal Stochastic Linear Systems with Exponential Performance Criteria and Their Relation to Deterministic Differential Games," *IEEE Transactions on Automatic Control* 18, 124–131.

18. Leitemo, Kai, Ulf Soderstrom (2008), "Robust Monetary Policy in the New Keynesian Framework," *Macroeconomic Dynamics* 12 (supplement 1), 126–135.
19. Leitemo, Kai, Ulf Soderstrom (2008), "Robust Monetary Policy in a Small Open Economy," *Journal of Economic Dynamics and Control* 32, 3218–3252.
20. Lucas, Robert E., Jr. (1972), "Expectations and the Neutrality of Money," *Journal of Economic Theory* 4, 103–24.
21. Luo, Yulei (2008), "Consumption Dynamics under Information Processing Constraints," *Review of Economic Dynamics* 11(2), 366–385.
22. Luo, Yulei and Eric R. Young (2010), "Risk-sensitive Consumption and Savings under Rational Inattention," *American Economic Journal: Macroeconomics* 2(4), 281–325.
23. Luo, Yulei, Jun Nie, and Eric R. Young (2011a), "Robustness, Information-processing Constraints, and the Current Account in Small Open Economies," forthcoming, *Journal of International Economics* 88(1), 104–120.
24. Luo, Yulei, Jun Nie, and Eric R. Young (2011b), "Robust Control, Informational Frictions, and International Consumption Correlations," Manuscript.
25. Luo, Yulei, Jun Nie, and Eric R. Young (2011c), "Model Uncertainty and Intertemporal Tax Smoothing," Manuscript.
26. Morris, Stephen and Hyun Song Shin (2002), "The Social Value of Public Information," *American Economic Review* 92, 1521–34.
27. Muth, John F. (1960), "Optimal Properties of Exponentially Weighted Forecasts", *Journal of the American Statistical Association* 55: 299–306.
28. Sims, Christopher A. (2003), "Implications of Rational Inattention," *Journal of Monetary Economics* 50(3), 665–690.
29. Sims, Christopher A. (2006), "Rational Inattention: Beyond the Linear-Quadratic Case," *American Economic Review* 96(2), 158–163.
30. Svensson, L.E.O. (2000), "Open-economy inflation targeting," *Journal of International Economics* 50(1), 155–184.
31. Walsh, Carl E. (2004), "Robust optimal instrument rules and robust control: An equivalent result," *Journal of Money, Credit, and Banking* 36(6), 1105–1113.
32. Whittle, Peter (1981), "Risk-Sensitive Linear/Quadratic/Gaussian Control," *Advances in Applied Probability* 13, 764–777.

Chapter 5
Hong Kong Inflation Dynamics: Trend and Cycle Relationships with the USA and China

Pym Manopimoke

5.1 Introduction

Inflation modeling is an important topic in macroeconomics, particularly because being able to understand and predict inflation plays a central role in monetary policy analysis. Recently, the approach of modeling inflation as having two components, trend and cycle, has become an appealing way to study the inflation process. Within this framework, changes to the trend component are driven by permanent shocks and correspond to long-horizon forecasts of inflation. Shocks to the cycle component are transitory and generally arise from short-run fluctuations in aggregate demand. Policymakers closely monitor movements in trend inflation as it indicates the future course of inflation that is rid of short-term noise. At the same time, knowledge about the driving forces behind the cyclical movements can help improve near-term inflation forecasts as well as deliver an improved understanding about the monetary policy transmission mechanism.

One challenge in studying the trend and cycle components of inflation is that they are unobserved, thus an appropriate econometric technique must be used to extract their movements from the observed data. In recent work, a popular method for trend and cycle decomposition is to model the inflation dynamics as an unobserved components (UC) model, which can then be easily estimated with the Kalman filter once cast into state-space form. For example, Stock and Watson (2007) in [28] propose a univariate UC model that decomposes US inflation into a trend component which follows a driftless random walk, and a stationary white noise component. Harvey (2008) in [17] estimates a UC model for US inflation that corresponds to a reduced form Phillips curve, with lags of inflation replaced by a random walk process. Lee and Nelson (2007) in [24] and Kim et al. (2012) in [23] consider estimation of trend-cycle UC models of inflation that are consistent with the forward-looking

P. Manopimoke (✉)
Department of Economics, University of Kansas, Lawrence, KS 66045, USA
e-mail: pymm@ku.edu

Y. Zeng and S. Wu (eds.), *State-Space Models: Applications in Economics and Finance*, Statistics and Econometrics for Finance 1, DOI 10.1007/978-1-4614-7789-1_5,

New Keynesian Phillips curve (NKPC). Within this context, trend inflation corresponds to long-horizon forecasts of inflation and the cycle component is driven by current and future forecasts of the real activity variable. Modeling inflation in accordance with a Phillips curve has the advantage of giving the extracted trend and cycle components more economic content when compared to a univariate decomposition that is based solely on the statistical properties of inflation.

Thus far, the majority of studies that estimate UC models for inflation limit the driving factors for trend and cycle movements to domestic ones. This chapter extends the UC model of [23] to account for external factors as well and applies the model to study Hong Kong inflation dynamics. The case of Hong Kong is particularly interesting for at least three reasons. First, the direction of trend inflation is usually assumed to be driven by domestic monetary policy. For example, in the USA, the usual approach is to attribute movements in trend inflation to changes in the Federal Reserve Bank's implicit inflation target (see [7, 21]). However, Hong Kong relinquished the control of its monetary policy stance since it entered into the Linked Exchange Rate System in October 1983. To establish stability and confidence in the economy, Hong Kong fixed its currency at a rate of 1 HKD to 7.80 US dollars, leaving little room for the conduct of discretionary monetary policy. Accordingly, Hong Kong trend inflation may be heavily reliant on external factors such as US trend inflation movements.

Second, inflation in a country under a currency board arrangement such as Hong Kong is believed to be highly dependent on external forces. Moreover, Hong Kong is a small open economy that engages in substantial amounts of international trade which could lead to volatile price movements. While these swings may originate from many sources, it is often understood that shocks from the USA and Mainland China are most responsible in shaping the macroeconomic landscape of Hong Kong. This is because these two countries are Hong Kong's leading trading partners and investors. Also, as mentioned earlier, the economy of Hong Kong is tied to some degree to the USA via the Linked Exchange Rate System. As for China, the close geographic proximity and the return of Hong Kong to the Chinese sovereignty in 1997 has led to tight economic integration between Hong Kong and the Mainland through activities in trade, FDI, tourism, and financial markets. While it is undisputed that these developments with the USA and China are important towards Hong Kong's price movements, the exact nature and transmission mechanism is less well understood. The UC framework developed in this chapter can help shed some light on this issue.

Last, despite the economic influence that China has exerted on the rest of the world, its macroeconomic variables and linkages with its trading partners is still an understudied topic. In this chapter, the output gaps in the empirical model are treated as latent variables, thus a by-product from estimation is a measure of China's unobserved output gap. A number of studies attempt to estimate China's output gap through UC approaches such as [11], but the majority of work rely on within country relationships and Chinese data alone. By exploiting the macroeconomic linkages between Hong Kong, US, and China, information in Hong Kong and US data may be able to help deliver a more accurate measure of China's output gap.

The rest of this chapter is organized as follows. Section 5.2 briefly describes the characteristics of Hong Kong's inflation dynamics as well as the related literature. The model specification is outlined in Sect. 5.3 and Sect. 5.4 discusses the empirical findings. Section 5.5 concludes.

5.2 Literature Review

Since the establishment of the currency board arrangement in 1983, consumer price inflation in Hong Kong has varied substantially from highly inflationary to deflationary periods. Figure 5.1 plots headline inflation as calculated from the consumer price index (CPI), CPI inflation that excludes rental components, and underlying CPI inflation. The latter measure is headline inflation that strips out the impact of one-off government relief measures. As shown, Hong Kong experienced high inflation for the most part of the 1980s and 1990s. Then, it underwent a 6-year prolonged period of deflation starting in 1998 which may have been spurred by events such as the Asian Financial Crisis and the greater integration with Mainland China. During the recent period, it can be observed that Hong Kong inflation has been on the rise since mid-2004, albeit with a slight dip due to the recent recession. Rising global food prices along with a number of other factors such as rising energy prices, the appreciation of the renminbi, and the weakening US dollar may all be responsible for this increase in headline inflation. However, it can be inferred from Fig. 5.1 that

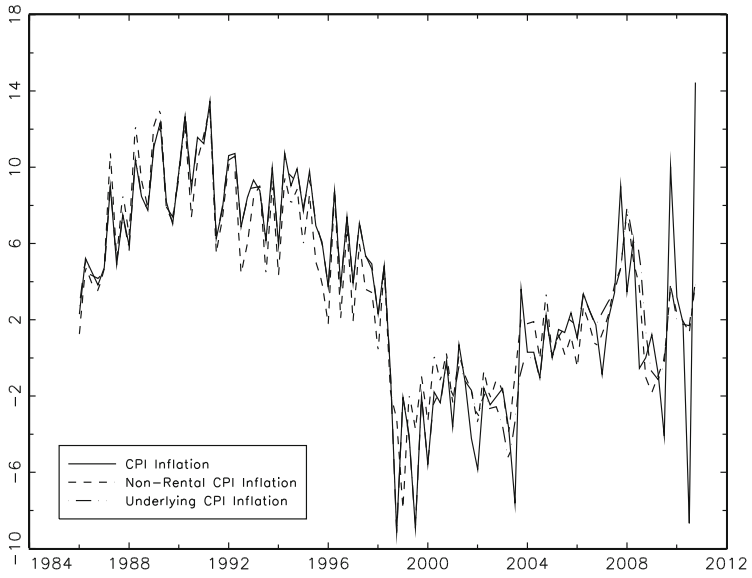

Fig. 5.1 Hong Kong CPI inflation

during the recent period, the rise in the rental component has driven the overall increase in Hong Kong CPI inflation as well.

In the literature, various versions of the NKPC have been used to analyze the inflation process. In the baseline NKPC, monopolistic competitive firms exhibit forward-looking price-setting behavior in an environment of sticky prices. Accordingly, current inflation is postulated to depend on expected future inflation and a measure of real economic activity such as the output gap. However, this so-called forward-looking NKPC has difficulty in explaining high persistence observed in the inflation data. Therefore, the forward-looking NKPC is often augmented to include a backward-looking component or a lagged inflation term to help fit with the data, resulting in an NKPC of hybrid form. The inclusion of a backward-looking term is somewhat ad hoc but is often justified by the existence of price-indexation or rule-of-thumb price-setting behavior (see [4, 9]).

For the case of Hong Kong, an open-economy hybrid NKPC is often the preferred model used to study the inflation process as the economy relies heavily on international trade. The open-economy NKPC is an extension of the baseline model where pricing decisions are also allowed to depend on external sector macrovariables such as fluctuations in the multilateral terms of trade or imported intermediate goods (see [5, 10]). Among the few studies that have estimated open-economy Phillips curves for Hong Kong, the majority report evidence supporting the empirical relevance of the model. For example, using the instrumental variable approach, Genberg and Pauwels (2005) in [11] find that the Phillips curve can provide a good description of movements in Hong Kong inflation and that both backward and forward-looking components are important in the NKPC. In addition, they report that using either a output gap, a unit labor cost gap or a specification of marginal cost as the driving variable for inflation yield similar results. Liu and Tsang (2008) in [25] also employs a Phillips curve model to study Hong Kong domestic inflation but they focus on analyzing the pass-through effect of exchange rate movements to Hong Kong domestic inflation. They find that although the degree of exchange rate pass-through is high in Hong Kong compared to OECD averages, domestic factors are also important in explaining Hong Kong inflation dynamics and can even dominate external factors in the medium-run.

Recent studies have also attempted to gain a better understanding about Hong Kong inflation trend and cycle movements. For example, Leung et al. (2009) in [26] employs various approaches such as the exclusion method and the principal components analysis to extract trend inflation movements from the data. Ha et al. (2002) in [15] estimates a backward-looking Phillips curve augmented by an error-correction term to relate Hong Kong's inflation to those of the USA and China in the long-run, and the output gap, import prices, and property prices in the short-run. They find that US inflation explains 92% of Hong Kong's long-run price movements whereas in the short-run, lags of the output gap, import prices, and property prices are found to be important. Cheung and Yuen (2002) in [3] also find long-run price movements in Hong Kong to be tied to the USA via cointegration tests, and using a vector error correction model, they also show that US inflation has a significant impact on Hong Kong inflation in the short-run with a lag of 2 years. Ha and Fan (2002) in

[14] uses a panel of city-level commodity prices in Hong Kong, Beijing, Shanghai, Guangzhou, and Shenzhen to examine the price convergence between Hong Kong and Mainland China. They find that price convergence with the Mainland is responsible for about less than a quarter of the deflation in Hong Kong, whereas domestic cyclical conditions may play a larger role.

5.3 Model Specification

Consider the following NKPC:

$$\pi_t = E_t(\pi_{t+1}) + kx_t + \eta_t, \tag{5.1}$$

where $E_t(.)$ refers to expectation formed conditional on information up to time t, π_t is current inflation, k is the slope of the Phillips curve, and x_t is the output gap, defined as the difference between actual and potential output. As explained in [23], η_t may be serially correlated if the backward-looking component or additional leads of inflation beyond $t+1$ are important in the NKPC.[1]

Iterating (5.1) forward results in the NKPC of the following form:

$$\pi_t = \lim_{j\to\infty} E_t(\pi_{t+j}) + k\sum_{j=0}^{\infty} E_t(x_{t+j}) + \tilde{z}_t, \tag{5.2}$$

where $\tilde{z}_t = \sum_{j=0}^{\infty} E_t(\eta_{t+j})$. Since a number of studies such as [19] and [12] fail to reject the null of a unit root for Hong Kong inflation, the $\lim_{t\to\infty} E_t(\pi_{t+j})$ term can be interpreted as long-horizon forecasts of inflation or the Beveridge and Nelson stochastic trend. The remaining term, $k\sum_{j=0}^{\infty} E_t(x_{t+j}) + \tilde{z}_t$, is the stationary cycle component of inflation, also known as the inflation gap. Note that in theory, $\tilde{z}_t$ would be driven by backward-looking or additional forward-looking price dynamics. However, empirically, $\tilde{z}_t$ could also be influenced by a variety of other macroeconomic factors. For the case of a small open economy such as Hong Kong, fluctuations in terms of trade, import prices, and property price movements as well as exchange rate variability could all be relevant towards explaining $\tilde{z}_t$. This chapter is particularly interested in investigating how US and China output gaps influence $\tilde{z}_t$ movements.

[1] To illustrate this point, notice that the widely estimated hybrid NKPC:

$$\pi_t = (1-\alpha)E_t(\pi_{t+1}) + \alpha\pi_{t-1} + kx_t + \eta_t, \;\; 0 \le \alpha < 1,$$

can be rewritten as (5.1) with $\eta_t = \alpha(\pi_{t-1} - E_t(\pi_{t+1}))$. It then follows that if $\alpha > 0$, firms are backward-looking and η_t will be serially correlated. The above model is a hybrid NKPC that is derived under the assumption that inflation is stationary. However, as shown in [7], additional leads of inflation beyond $t+1$ may enter the NKPC if inflation is assumed to have a unit root. Thus, in the presence of stochastic trend inflation, serial correlation in η_t may not necessarily stem from backward-looking price-setting dynamics. Rather, serial correlation in η_t may serve as a spurious proxy for additional forward-looking elements.

Following [23], (5.2) can be written as the following UC model for inflation:

$$\pi_t = \tilde{\pi}_t + k \sum_{j=0}^{\infty} E_{t-1}(x_{t+j}) + z_t, \tag{5.3}$$

$$\tilde{\pi}_t = \tilde{\pi}_{t-1} + e_t, \tag{5.4}$$

$$z_t = \varepsilon_t, \tag{5.5}$$

where the inflation gap is rewritten with $z_t = \tilde{z}_t + (\sum_{j=0}^{\infty} E_t(x_{t+j}) - \sum_{j=0}^{\infty} E_{t-1}(x_{t+j}))$ for feasible estimation of the model. Since z_t is a function of economic agents' revision on the present value of future output gaps, shocks to z_t are allowed to be correlated with those of x_t through a nonzero correlation coefficient $\rho_{\varepsilon v}$. Also, note that although z_t may be serially correlated, it is defined as a white noise process in (5.5). This is because with the use of Akaike's Information Criterion (AIC) tests, it is determined that the fit of the data during the time period studied is best when z_t is modeled as a white noise process with $\varepsilon_t \sim i.i.d.N(0, \sigma_\varepsilon^2)$. This result implies that a forward-looking NKPC can explain Hong Kong inflation data well.

In dealing with the unobserved output gap x_t, Kim et al. (2012) in [23] assumes that the US output gap is an observed process and proxies it with the Congressional Budget Office's (CBO) estimate of the output gap. For the case of Hong Kong, there is less of an established measure for the output gap. Hence, the output gap in the UC model for Hong Kong inflation is also treated as an unobserved process and is extracted from the following UC model for output:

$$y_t = \tau_t + x_t, \tag{5.6}$$

$$\tau_t = \mu + \tau_{t-1} + w_t, \tag{5.7}$$

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + v_t. \tag{5.8}$$

The above model follows [16] and [6], in which equilibrium output y_t is decomposed into a stochastic trend component τ_t and a cyclical component x_t which corresponds to the output gap.[2] The output trend and cycle components are assumed to be uncorrelated. Note that the unobserved output gap x_t backed out from the full UC model will be consistent with the NKPC, and its movements are influenced not only by information contained in its own lags but also by information in inflation as well as the trend output growth rate.

[2] As an alternative to a UC model for real output, a UC model for the unemployment rate can be estimated with the inflation equations instead. The unemployment rate can be specified as a sum of the natural rate, which typically is assumed to follow a driftless random walk, and an unemployment gap, which may follow an autoregressive process. In such a case, the unemployment gap will replace the output gap as the driving variable for inflation in the NKPC specification. For examples of UC models that use the unemployment gap as a measure of economic slack, see [1] and [24].

The UC model denoted by (5.3)–(5.8) is henceforth referred to as the one-country model. Its corresponding state-space representation can be written as follows:

Measurement equation

$$\begin{bmatrix} \pi_t \\ y_t \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \tilde{\pi}_t \\ z_t \\ \tau_t \\ x_t \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} k\sum_{j=0}^{\infty} E_{t-1}(x_{t+j}) \\ 0 \end{bmatrix}, \tag{5.9}$$

Transition equation

$$\begin{bmatrix} \tilde{\pi}_t \\ z_t \\ \tau_t \\ x_t \\ x_{t-1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \mu \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \phi_1 & \phi_2 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \tilde{\pi}_{t-1} \\ z_{t-1} \\ \tau_{t-1} \\ x_{t-1} \\ x_{t-2} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} e_t \\ \varepsilon_t \\ w_t \\ v_t \end{bmatrix}, \tag{5.10}$$

$$\begin{bmatrix} e_t \\ \varepsilon_t \\ w_t \\ v_t \end{bmatrix} \sim i.i.d.N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_e^2 & 0 & 0 & 0 \\ 0 & \sigma_\varepsilon^2 & 0 & \sigma_{\varepsilon v} \\ 0 & 0 & \sigma_w^2 & 0 \\ 0 & \sigma_{\varepsilon v} & 0 & \sigma_v^2 \end{bmatrix} \right)$$

where $\sigma_{\varepsilon v} = \rho_{\varepsilon v}\sigma_\varepsilon\sigma_v$. Note that $\sum_{j=0}^{\infty} E_{t-1}(x_{t+j})$ in the measurement equation can be calculated as:

$$\sum_{j=0}^{\infty} E_{t-1}(x_{t+j}) = \tilde{e}_1' F(I_2 - F)^{-1}\tilde{x}_{t-1}, \tag{5.11}$$

where $\tilde{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $F = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix}$ and $\tilde{x}_{t-1} = \begin{bmatrix} x_{t-1} \\ x_{t-2} \end{bmatrix}$.

To investigate how Hong Kong's inflation dynamics may be affected by external factors from the USA and China, the one-country model is extended to the following three-country model:

NKPC for Hong Kong:

$$\pi_{1,t} = \tilde{\pi}_{1,t} + k_1 \sum_{j=0}^{\infty} E_{t-1}(x_{1,t+j}) + z_{1,t}, \tag{5.12}$$

$$\tilde{\pi}_{1,t} = \eta\tilde{\pi}_{2,t-1} + (1-\eta)\tilde{\pi}_{1,t-1} + e_{1,t}, \tag{5.13}$$

$$z_{1,t} = \gamma_1 x_{2,t-1} + \gamma_2 x_{2,t-2} + \gamma_3 x_{3,t-1} + \gamma_4 x_{3,t-2} + \varepsilon_{1,t}, \tag{5.14}$$

$$y_{1,t} = \tau_{1,t} + x_{1,t}, \tag{5.15}$$

$$\tau_{1,t} = \mu_1 + \tau_{1,t-1} + w_{1,t}, \tag{5.16}$$

$$x_{1,t} = \phi_{1,1}x_{1,t-1} + \phi_{1,2}x_{1,t-2} + v_{1,t}, \tag{5.17}$$

NKPC for the USA:

$$\pi_{2,t} = \tilde{\pi}_{2,t} + k_2 \sum_{j=0}^{\infty} E_{t-1}(x_{2,t+j}) + z_{2,t}, \tag{5.18}$$

$$\tilde{\pi}_{2,t} = \tilde{\pi}_{2,t-1} + e_{2,t}, \tag{5.19}$$

$$z_{2,t} = \varepsilon_{2,t}, \tag{5.20}$$

$$y_{2,t} = \tau_{2,t} + x_{2,t}, \tag{5.21}$$

$$\tau_{2,t} = \mu_2 + \tau_{2,t-1} + w_{2,t}, \tag{5.22}$$

$$x_{2,t} = \phi_{2,1} x_{2,t-1} + \phi_{2,2} x_{2,t-2} + v_{2,t}, \tag{5.23}$$

Output equation for China:

$$y_{3,t} = \tau_{3,t} + x_{3,t}, \tag{5.24}$$

$$\tau_{3,t} = \mu_3 + \tau_{3,t-1} + w_{3,t}, \tag{5.25}$$

$$x_t = \phi_{3,1} x_{3,t-1} + \phi_{3,2} x_{3,t-2} + v_{3,t}, \tag{5.26}$$

where variables with subscripts 1, 2, and 3 belong to Hong Kong, the USA, and China, respectively, except for all γ coefficients in (5.14) that belong to the domestic country, Hong Kong. In the NKPC representation for Hong Kong, two departures are made from the one-country model. First, since Hong Kong ties its monetary policy to the USA via the Linked Exchange Rate System, the three-country model allows Hong Kong trend inflation to be influenced by US trend inflation movements. The importance of US trend inflation is captured through the significance of the coefficient η. Note that in theory, Hong Kong's inflation rate should converge to the levels of the USA in the long-run, in which case the two countries will share a common trend with η equal to one. However, as shown in Fig. 5.2, it is unclear whether this is empirically the case since the differences between Hong Kong and US price movements are quite substantial.

Next, the cycle component of Hong Kong inflation that follows (5.14) is allowed to depend on the lagged output gap effects from the USA and China through the coefficients γ_1, γ_2, γ_3, and γ_4. In general equilibrium, the term of trade gap is driven by the difference between the domestic and foreign countries' output gaps, thus the significance of these coefficients may denote the importance of terms of trade fluctuations onto Hong Kong inflation dynamics.

Note that the UC model for the USA in (5.18)–(5.23) is similar to the one-country NKPC model for Hong Kong and is not influenced by any variables that belong to Hong Kong or China.[3] As for China, only an output equation is included as the literature suggests that the fit of Chinese inflation data to standard Phillips curves are

[3] This is similar to [23]'s UC model for US inflation except here, z_t is specified as a white-noise process instead of an AR(1) process. However, the empirical findings in [23] suggest that z_t follows a white noise process for the post mid-1980s which corresponds to the sample period studied in this chapter.

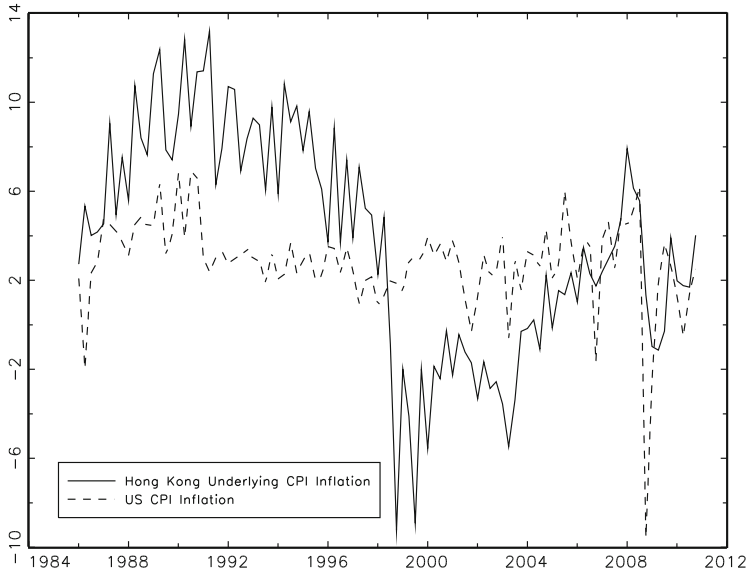

Fig. 5.2 Hong Kong and US inflation

problematic (see [11]). This is not surprising since there have been large swings and structural changes in Chinese inflation that may stem from events such as changes in the exchange rate regime, trade liberalization, and the impact of price deregulation. Finally, in the three-country model, the shocks to all three output gaps are allowed to be correlated through correlation coefficients $\rho_{12,v}$, $\rho_{13,v}$, and $\rho_{23,v}$. Likewise, all shocks to trend output are allowed to be correlated through correlation coefficients $\rho_{12,w}$, $\rho_{13,w}$, and $\rho_{23,w}$.

The corresponding state-space representation for the three-country model can be written as:

Measurement equation

$$\begin{bmatrix} \pi_{1,t} \\ \pi_{2,t} \\ y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \tilde{\pi}_{1,t} \\ z_{1,t} \\ \tilde{\pi}_{2,t} \\ z_{2,t} \\ \tau_{1,t} \\ x_{1,t} \\ x_{1,t-1} \\ \tau_{2,t} \\ x_{2,t} \\ x_{2,t-1} \\ \tau_{3,t} \\ x_{3,t} \\ x_{3,t-1} \end{bmatrix} + \begin{bmatrix} k_1 \sum_{j=0}^{\infty} E_{t-1}(x_{1,t+j}) \\ k_2 \sum_{j=0}^{\infty} E_{t-1}(x_{2,t+j}) \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{5.27}$$

Transition equation

$$
\begin{bmatrix} \tilde{\pi}_{1,t} \\ z_{1,t} \\ \tilde{\pi}_{2,t} \\ z_{2,t} \\ \tau_{1,t} \\ x_{1,t} \\ x_{1,t-1} \\ \tau_{2,t} \\ x_{2,t} \\ x_{2,t-1} \\ \tau_{3,t} \\ x_{3,t} \\ x_{3,t-1} \end{bmatrix} = \begin{bmatrix} (1-\eta) & 0 & \eta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \gamma_1 & \gamma_2 & 0 & \gamma_3 & \gamma_4 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \phi_{1,1} & \phi_{1,2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi_{2,1} & \phi_{2,2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi_{3,1} & \phi_{3,2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \tilde{\pi}_{1,t-1} \\ z_{1,t-1} \\ \tilde{\pi}_{2,t-1} \\ z_{2,t-1} \\ \tau_{1,t-1} \\ x_{1,t-1} \\ x_{1,t-2} \\ \tau_{2,t-1} \\ x_{2,t-1} \\ x_{2,t-2} \\ \tau_{3,t-1} \\ x_{3,t-1} \\ x_{3,t-2} \end{bmatrix}
$$

$$
+ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \mu_1 \\ 0 \\ 0 \\ \mu_2 \\ 0 \\ 0 \\ \mu_3 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} e_{1,t} \\ \varepsilon_{1,t} \\ e_{2,t} \\ \varepsilon_{2,t} \\ w_{1,t} \\ v_{1,t} \\ w_{2,t} \\ v_{2,t} \\ w_{3,t} \\ v_{3,t} \end{bmatrix}, \tag{5.28}
$$

$$
\begin{bmatrix} e_{1,t} \\ \varepsilon_{1,t} \\ e_{2,t} \\ \varepsilon_{2,t} \\ w_{1,t} \\ v_{1,t} \\ w_{2,t} \\ v_{2,t} \\ w_{3,t} \\ v_{3,t} \end{bmatrix} \sim i.i.d.N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{1,e}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{1,\varepsilon}^2 & 0 & 0 & 0 & \sigma_{1,\varepsilon v} & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{2,e}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{2,\varepsilon}^2 & 0 & 0 & 0 & \sigma_{2,\varepsilon v} & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{1,w}^2 & 0 & \sigma_{12,w} & 0 & \sigma_{13,w} & 0 \\ 0 & \sigma_{1,\varepsilon v} & 0 & 0 & 0 & \sigma_{1,v}^2 & 0 & \sigma_{12,v} & 0 & \sigma_{13,v} \\ 0 & 0 & 0 & 0 & \sigma_{12,w} & 0 & \sigma_{2,w}^2 & 0 & \sigma_{23,w} & 0 \\ 0 & 0 & 0 & \sigma_{2,\varepsilon v} & 0 & \sigma_{12,v} & 0 & \sigma_{2,v}^2 & 0 & \sigma_{23,v} \\ 0 & 0 & 0 & 0 & \sigma_{13,w} & 0 & \sigma_{23,w} & 0 & \sigma_{3,w}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{13,v} & 0 & \sigma_{23,v} & 0 & \sigma_{3,v}^2 \end{bmatrix} \right)
$$

where $\sigma_{i,\varepsilon v} = \rho_{i,\varepsilon v}\sigma_{i,\varepsilon}\sigma_{i,v}$ for $i = 1,2$ and $\sigma_{jk,v} = \rho_{jk,v}\sigma_{j,v}\sigma_{k,v}$, $\sigma_{jk,w} = \rho_{jk,w}\sigma_{j,w}\sigma_{k,w}$ for $j,k = 1,2,3; j < k$.

Similar to (5.11), the infinite sum term $\sum_{j=0}^{\infty} E_{t-1}(x_{1,t+j})$ term in the measurement equation can be calculated as:

$$\sum_{j=0}^{\infty} E_{t-1}(x_{1,t+j}) = \tilde{e}_1' F(I_2 - F)^{-1} \tilde{x}_{1,t-1}, \tag{5.29}$$

where $\tilde{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $F = \begin{bmatrix} \phi_{1,1} & \phi_{1,2} \\ 1 & 0 \end{bmatrix}$ and $\tilde{x}_{1,t-1} = \begin{bmatrix} x_{1,t-1} \\ x_{1,t-2} \end{bmatrix}$. Likewise,

$$\sum_{j=0}^{\infty} E_{t-1}(x_{2,t+j}) = \tilde{e}_1' F(I_2 - F)^{-1} \tilde{x}_{2,t-1}, \tag{5.30}$$

where $\tilde{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $F = \begin{bmatrix} \phi_{2,1} & \phi_{2,2} \\ 1 & 0 \end{bmatrix}$ and $\tilde{x}_{2,t-1} = \begin{bmatrix} x_{2,t-1} \\ x_{2,t-2} \end{bmatrix}$.

5.4 Empirical Results

The empirical analysis employs quarterly data that spans 1986Q1–2010Q4. The Hong Kong inflation series are calculated as the one-quarter log change of the CPI, the non-rental component of the CPI, and Hong Kong's underlying CPI. The underlying CPI inflation is the preferred measure for Hong Kong inflation in this study as it strips out the impact of one-off government relief measures. This relief is designed to reduce the final cost of various goods and services to people burdened by inflation. Headline inflation does not adjust for this, which may have caused distortions and increased volatility in the data. Inflation data for Hong Kong is obtained from the CEIC database and from the database at the Hong Kong Monetary Authority.

For the USA, the inflation rate is calculated as the one-quarter log change of the CPI obtained from the Federal Reserve Economic Database. As for data on output, the Purchasing Power Parity (PPP)-adjusted Gross Domestic Product (GDP) is used, with 2005 PPP data obtained from the Penn World Table. GDP data for Hong Kong, USA, and China are obtained from the Hong Kong Census and Statistics Department, the US Bureau of Economic Analysis and the China National Bureau of Statistics, respectively. Note that the series in which quarterly data is not available, monthly data is converted to quarterly data by averaging monthly data within the quarter.

First, the one-country model is fitted to Hong Kong inflation to see how well the model performs without explicitly accounting for the external influences from the USA and China. Table 5.1 reports the estimation results for the three Hong Kong inflation series. All parameter estimates have the right sign and are of reasonable magnitudes, and they are statistically significant at the 5% level except for the correlation coefficient $\rho_{\varepsilon v}$. The estimation results produced from all three Hong Kong inflation series are reasonably similar. Some minor differences are as expected, which include the variability of shocks to the component z_t being slightly smaller for underlying CPI inflation and shocks to trend inflation being less volatile for the

Table 5.1 Estimation results for the one-country model [Hong Kong: 1986Q1–2010Q4]

Parameters	CPI	Non-rental CPI	Underlying CPI
Phillips curve slope, output trend drift, and AR coefficients of the unobserved gap			
k	0.345 (0.168)	0.364 (0.153)	0.379 (0.185)
μ	0.965 (0.132)	0.966 (0.130)	0.976 (0.137)
ϕ_1	1.588 (0.078)	1.574 (0.077)	1.582 (0.081)
ϕ_2	−0.714 (0.070)	−0.709 (0.065)	−0.687 (0.068)
Standard deviations and correlations			
σ_e	1.088 (0.197)	0.841 (0.158)	1.145 (0.181)
σ_w	1.229 (0.118)	1.218 (0.112)	1.280 (0.108)
σ_v	0.463 (0.110)	0.478 (0.102)	0.397 (0.110)
σ_ε	2.064 (0.210)	1.940 (0.182)	1.692 (0.192)
$\rho_{\varepsilon v}$	−0.548 (0.378)	−0.286 (0.271)	−0.743 (0.419)
Log-likelihood:	−289.902	−280.604	−279.526

Note: Standard errors are in parentheses.

non-rental component of CPI inflation. The slope of the output gap is estimated to be approximately 0.35 for all inflation measures, suggesting that the domestic output gap plays an important role in explaining Hong Kong's short-run inflation movements. Finally, for the underlying CPI inflation measure, the model implies a trend output growth rate of 3.9%, and the unobserved output gap implied by the NKPC is fairly persistent with a sum of AR coefficients equal to 0.895.

Plots of the unobserved components produced from the three inflation measures are similar, so only the estimates from the underlying CPI inflation series are reported due to space considerations. First, smoothed trend inflation estimates inferred from the one-country model are plotted in Fig. 5.3. As shown, Hong Kong trend inflation tracks the overall movements in actual inflation well. Trend inflation was high in the mid-1980s to early 1990s, experienced a sharp drop in the mid-1990s, was low throughout the late 1990s and early 2000s, and has been picking up since the mid-2000s. In comparison with the literature, the estimates of trend inflation shown here are less volatile than the measures that Leung et al. (2009) in [26] report by using the exclusion method or the principal components analysis.

Figure 5.4 plots the unobserved output gap as implied by the one-country model for Hong Kong alongside the HP-filtered output gap.[4] It is to be emphasized that

[4] The HP gap is a commonly used measure of the output gap in Phillip curve models for Hong Kong. This is despite the well-known shortcomings of the HP-filter which includes difficulty in identifying the appropriate smoothing parameter as well as high end-sample biases (see [18]). For comparisons with the literature, the one-country model is also estimated with the HP-filtered output gap. In doing so, the one-country model is reduced to (5.3)–(5.5) and (5.8), with the HP gap acting as an observed measure of the output gap. For underlying CPI inflation, the reduced model gives an estimated slope of 0.082 which is smaller in magnitude than the reported slope estimate of 0.379 in Table 5.1. This result is as expected since the HP gap undergoes larger swings relative to the unobserved output gap that is extracted from the one-country model.

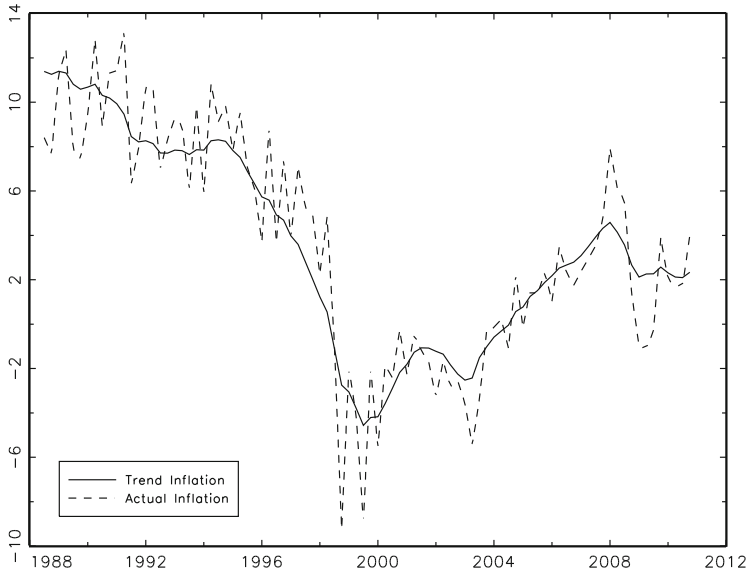

Fig. 5.3 Hong Kong actual inflation and smoothed estimates of trend inflation from the one-country model

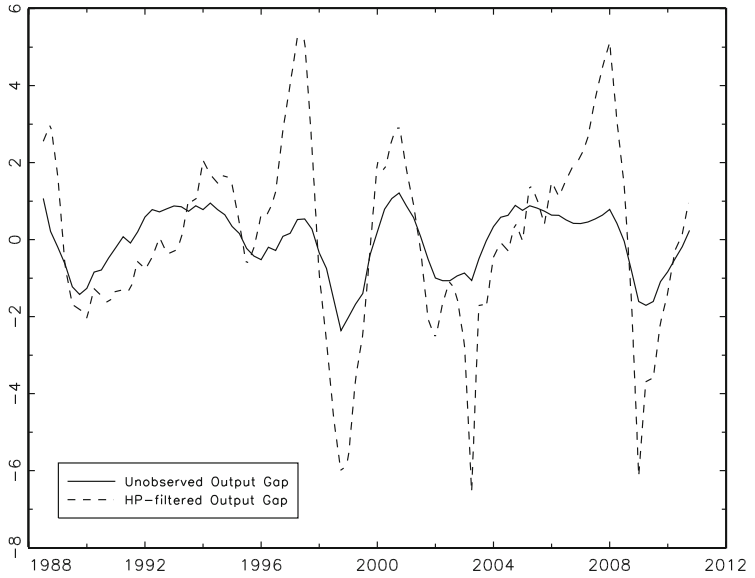

Fig. 5.4 Hong Kong unobserved output gap from the one-country model and HP-filtered output gap

the two output gap measures are very different since the HP-gap is a purely statistical measure, whereas the unobserved output gap is one that is consistent with the NKPC. Nevertheless, it can be observed that the two series share the same general movements, with dates of peaks and troughs in the business cycle that roughly coincide. However, the magnitude of swings in the HP-filtered output gap is much more pronounced. This may be due to the fact that the HP-filter imposes a smooth trend, causing the variability to show up in the cyclical component, whereas the UC model makes no such assumption. Gerlach and Yiu (2004) in [13] and Cheng et al. (2011) in [2] also show that a univariate UC model for output that is similar to (5.6)–(5.8) produces a less volatile output gap that has smaller peaks and troughs when compared to the HP-filtered output gap. However, their estimated magnitudes are still larger than those implied by the one-country model. Thus, it may be the case that information in inflation helps identify a less volatile output gap.

Table 5.2 reports estimation results for the three-country model for Hong Kong, the USA, and China. The inflation measure for Hong Kong is the underlying CPI,

Table 5.2 Estimation results for the three-country model [Hong Kong, USA, China: 1986Q1–2010Q4]

Inflation equation parameters			
	Hong Kong	USA	
k	0.366 (0.183)	0.010 (0.005)	
η	0.052 (0.022)	–	
γ_1	−2.296 (1.255)	–	
γ_2	2.279 (1.240)	–	
γ_3	4.401 (2.287)	–	
γ_4	−4.004 (2.353)	–	
Output equation parameters			
	Hong Kong	USA	China
μ	0.937 (0.135)	0.797 (0.085)	2.395 (0.093)
ϕ_1	1.678 (0.053)	1.714 (0.081)	1.894 (0.036)
ϕ_2	−0.785 (0.048)	−0.719 (0.079)	−0.919 (0.033)
Standard deviations and correlations			
	Hong Kong	USA	China
σ_e	1.079 (0.208)	0.180 (0.112)	–
σ_ε	1.661 (0.209)	1.834 (0.142)	–
σ_w	1.266 (0.100)	0.307 (0.079)	0.845 (0.063)
σ_v	0.357 (0.085)	0.435 (0.077)	0.084 (0.037)
$\rho_{1,\varepsilon v}$	−0.718 (0.449)		
$\rho_{2,\varepsilon v}$	0.123 (0.147)		
$\rho_{13,w}$	0.254 (0.099)		
$\rho_{12,v}$	0.570 (0.142)		
$\rho_{13,v}$	0.998 (0.002)		
$\rho_{23,v}$	0.998 (0.001)		
Log-likelihood: −397.276			

Note: Standard error are in parentheses.

but the estimation results are robust across the three different inflation series for Hong Kong. As shown, the parameter estimates that describe Hong Kong inflation and output dynamics are similar to those reported in Table 5.1. For the USA, they are similar to those that Kim et al. (2012) in [23] report when fitting a one-country model similar to (5.18)–(5.20) to US inflation and CBO output gap data. Comparing the NKPC parameter estimates across the two countries, the major difference is in the Phillips curve slope estimates k. Hong Kong has a steeper slope of 0.37 versus the USA that has a flat slope of 0.01. The estimates for Hong Kong reported here are slightly higher than those in the literature, but these comparisons are against the few studies that estimate Phillips curve models with quite different specifications and with data that span a shorter time period. For example, using the instrumental variable approach, Genberg and Pauwels (2005) in [11] estimate a version of the open-economy hybrid NKPC for the sample period 1984–2002 and report a slope estimate of 0.19. By a similar approach, Dua and Gaur (2010) in [8] obtain a slope estimate of 0.15 for the 1990–2005 sample. As for the USA, NKPCs typically deliver small slope estimates of magnitude 0.02 which is comparable to the one reported in this chapter (see [23, 24]). Finally, in explaining the sharp differences in the two countries' slope estimates, a number of theoretical models such as Romer (1993) in [27] predict that Phillip curve slopes increase with the degree of trade openness. Using 2011 data from the CIA World Factbook database, Hong Kong's degree of openness as measured by the sum of imports and exports as a percentage of GDP is 256.05, whereas for the USA it is 24.41. This difference is quite pronounced and hence the findings in this chapter provide empirical support to the argument that more open economies have steeper Phillips curves.

Examining the variability of inflation trend components through the estimated σ_e parameters, Hong Kong trend inflation is more volatile when compared to the USA. Estimates of Hong Kong trend inflation from the three-country model are similar to those shown in Fig. 5.3, except that they track actual inflation more closely during the decline in inflation that started in 1997. Figure 5.5 plots the estimates of US trend inflation which is substantially less volatile when compared to Hong Kong trend inflation, supporting the view that US inflation has been sufficiently well anchored at around 2% since the mid-1980s. From these observed differences in the long-run components of the two inflation series, it is not surprising that the estimate of η, which denotes the degree in which Hong Kong inflation is dependent on US inflation in the long-run is estimated at 0.05 which is low.[5] This result stands in contrast with [15]'s finding that Hong Kong prices will converge to US levels in the long-run.

Comparing estimates of the output equation parameters, the trend output growth rate was highest for China and lowest for the USA. China had an annual trend out-

[5] Long-run price movements in Hong Kong should effectively be tied to the USA through the Linked Exchange Rate System. Nevertheless, there may be many causes for persistent deviations of Hong Kong price dynamics from that of the USA. For example, the relatively high inflation in Hong Kong during the 1990s may be due to favorable export price shocks which hike up the prices of tradables that ultimately impact the prices of non-tradables. The Balassa–Samuelson effect in which the high productivity growth gap between the tradable sector and the non-tradable sector leads to a real exchange rate appreciation that increases prices of non-tradables could also be responsible for Hong Kong's higher long-term inflation in the 1990s (see [20] and references therein).

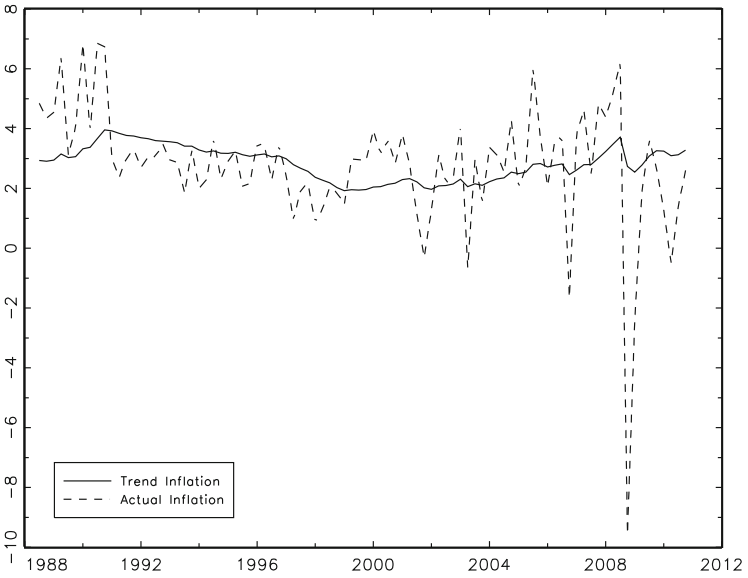

Fig. 5.5 US actual inflation and smoothed estimates of trend inflation from the three-country model

put growth rate of 9.58%, whereas for the USA, the growth rate is estimated at 3.19%. All three output gap measures are highly persistent as evidenced by the sum of their AR coefficients ϕ_1 and ϕ_2. Shocks to the three output gap series are highly correlated, especially the ones between Hong Kong and China and USA and China, yielding evidence of business cycle synchronization. Shocks to Hong Kong trend output is most volatile, with US output exhibiting the smoothest trend. The variability of shocks to China's trend output is also high, in contrast to the shocks to its cyclical component which is lower than Hong Kong and the USA by approximately a factor of 5. From estimation of the three-country model, it turns out that only shocks to Hong Kong and China trend outputs are correlated, thus the results reported here are based on the restricted model in which the correlations between the shocks to the other countries' output trends are set to zero.

In Fig. 5.6, the unobserved output gap estimates from the three-country model are plotted. As shown, the three output gaps began to co-move more closely since the early 2000s. Prior to this period, the output gaps of Hong Kong and the USA were unsynchronized. Given that Hong Kong's monetary policy is tied to the USA but their real economies differ, US monetary policy that aims at domestic output gap stabilization may have delivered adverse effects onto Hong Kong's economy, contributing to the high volatility observed in Hong Kong's inflation dynamics. Another observation from the graph is that the current recession affected the USA the most in terms of loss in output, whereas Hong Kong shows the fastest recovery. The Hong Kong output gap looks similar to the series obtained from the one-country model in Fig. 5.4, except at the end points. For example, examining the lowest point

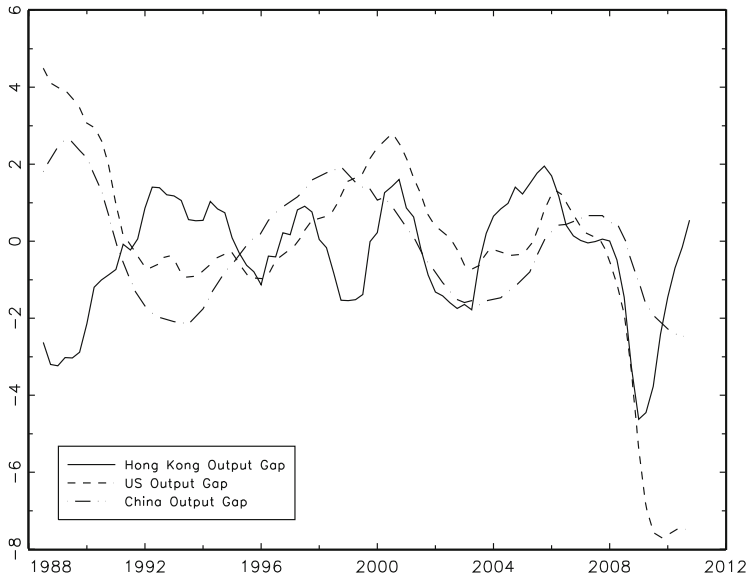

Fig. 5.6 Output gap estimates from the three-country model

of the most recent recession, the three-country model gap was lower than that of the one-country model gap by approximately 3%. In addition, it is sufficiently lower relative to the trough of the 1997 Asian financial crisis. As shown in the plot of Hong Kong trend output and actual output in Fig. 5.7, this suggests that for Hong Kong, the 1997 recession can be characterized by a large permanent loss in output whereas the loss during the most recent recession was largely temporary.

As for China, China's output gap appears smooth in comparison with the US and Hong Kong output gaps. The general movement of China's output gap reported here resembles those of [11], in which the authors estimate a univariate UC model for output similar to (5.24)–(5.26). The authors find an output gap that also peaks around the mid-1980s and mid-1990s but their output gap measure is slightly more volatile. By estimating a UC model for output using both US and China data, Jia and Sinclair (2009) in [22] also report a more volatile gap. However, note that these output gap estimates for China should be viewed with caution. Output data from the Mainland are known to be subject to considerable measurement errors causing output gap estimates to be imprecise. Moreover, Chinese GDP data are found to be very smooth in comparison with those of the USA and Hong Kong, and this limitation may have contributed to the overall smoothness of Chinese output gap estimates.

Finally, as discussed in Sect. 5.2, studies in the literature have found Hong Kong price dynamics to be related to macroeconomic factors in the USA at both short and long time-horizons. However, it has been more difficult to establish a link between Hong Kong inflation dynamics and macroeconomic factors from China. In this chapter, an encouraging finding is that both the coefficients that link the US and

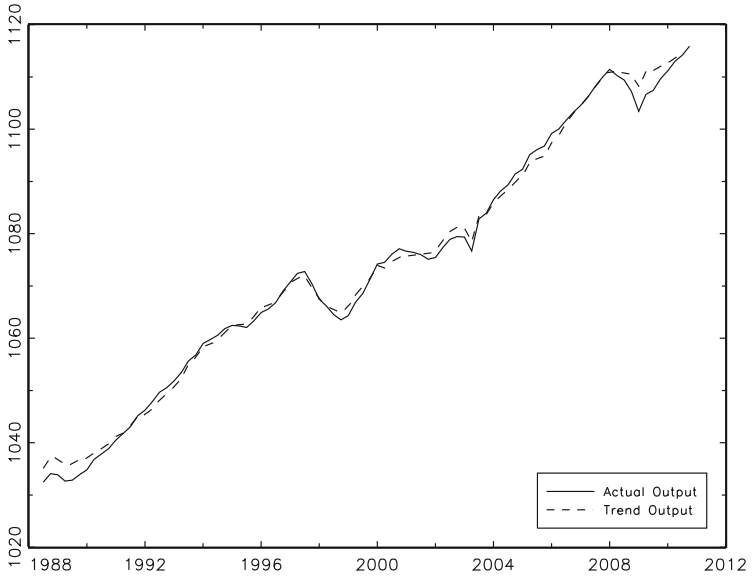

Fig. 5.7 Hong Kong actual output and estimates of trend output from the three-country model

China output gaps to Hong Kong inflation in the short-run are sufficiently large and statistically significant at the 10% level. Examining the estimates of γ_3 and γ_4, the output gap influence from China onto Hong Kong's cyclical component is approximately twice as large when compared to the impact from the US output gap, as reflected through the coefficients γ_1 and γ_2. In addition, from the signs on these gamma coefficients, it seems to be the case that the US and China output gaps have opposing effects on Hong Kong's inflation cycle at the first and second time lags. The finding that China's output gap effects matter more for Hong Kong inflation when compared to the USA is not surprising. This is because according to the statistics compiled by the Trade and Industry Department, the total trade between Hong Kong and Mainland China in 2011 is 48.5%, while its total trade with the USA is smaller at 7.6%.

5.5 Conclusion

This chapter investigates the extent in which domestic and external factors matter for Hong Kong inflation trend and cycle movements within the framework of a NKPC. The empirical model is an unobserved component model for inflation and output where US trend inflation and output gaps from the USA and China are allowed to influence Hong Kong price dynamics at the long and short time-horizons. In contrast to theory, the empirical findings suggest that since the mid-1980s, the degree in which Hong Kong and US inflation rates are related in the long-run is minor.

Over the short-run, the domestic output gap turns out to be a very important driving variable in explaining Hong Kong inflation dynamics. Moreover, foreign output gap effects from the US and China matter as well, with the magnitude on the China output gap estimated at twice as large as that of the USA. Comparing the unobserved output gap series that are backed out from the empirical model, there is evidence that the output gaps of the three countries have become more synchronized since the early 2000s.

The findings in this chapter are encouraging as it is able to identify meaningful relationships between Hong Kong inflation and external factors from the USA and China. Admittedly, there are other factors that could influence Hong Kong inflation that are not explicitly included in the empirical model. For example, permanent price shocks from China may be important in explaining Hong Kong trend inflation movements. Swings in property prices or global food and energy prices may also matter for Hong Kong's short-run price dynamics. Given the flexibility of the empirical model and the fact that both the trend and cycle components are reduced form expressions, incorporating such features into the model would be straightforward. Then, if the relationships between these factors and the model are found to be important and stable, an interesting avenue for future research would be to evaluate the forecasting performance of the empirical model. As a small open economy influenced heavily by international trade, Hong Kong inflation is known to be difficult to forecast. So far, vector autoregression (VAR) models are widely used to forecast Hong Kong inflation. However, giving the forecasting model more structure through the NKPC framework may be able to yield fruitful results.

Acknowledgments The author is grateful for the generous hospitality and financial support provided by the Hong Kong Institute for Monetary Research during the completion of this chapter. Special thanks go to the seminar participants at the Hong Kong Institute for Monetary Research, Michael Cheng at the Hong Kong Monetary Authority for helpful discussions and for providing Hong Kong inflation data, and an anonymous referee for valuable comments and suggestions.

References

1. Basistha A, Nelson C (2007) New measures of the output gap based on the forward-looking New Keynesian Phillips curve. Journal of Monetary Economics 54(2), 498–511
2. Cheng M, Chung L, Yu I-W (2011) On the estimation of the output gap of Hong Kong. HKIMR Working Paper No. 03/2011
3. Cheung Y-W, Yuen J (2002) Effects of US inflation on Hong Kong and Singapore. CESifo Working Paper Series No. 700. HKIMR Working Paper No. 03/2001
4. Christiano L, Eichenbaum M, Evans C (2005) Nominal rigidities and the dynamic effects of a shock to monetary policy. Journal of the Political Economy 113: 1–45
5. Clarida R, Gali J, Gertler M (2002) A simple framework for international monetary policy analysis. Journal of Monetary Economics 49: 879–904
6. Clark P (1987) The cyclical component of U.S. economic activity. The Quarterly Journal of Economics 102(4): 797–814
7. Cogley T, Sbordone A (2008) Trend inflation, indexation, and inflation persistence in the New Keynesian Phillips curve. American Economic Review 98(5): 2101–2126

8. Dua P, Gaur U (2010) Determination of inflation in an open economy Phillips curve framework: the case of developed and developing Asian countries. Macroeconomics and Finance in Emerging Market Economies 3(1): 33–51
9. Gali J, Gertler M (1999). Inflation dynamics: a structural econometric analysis. Journal of Monetary Economics 44(2): 195–222
10. Gali J, Monacelli T (2005) Monetary policy and exchange rate volatility in a small open economy. Review of Economic Studies 72, 707–734
11. Genberg H, Pauwels L (2005) An open economy New Keynesian Phillips curve: evidence from Hong Kong. Pacific Economic Review 10(2): 261–277
12. Gerlach S, Peng W (2006) Output gaps and inflation in Mainland China. China Economic Review 17(2): 210–225
13. Gerlach S, Yiu S (2004) Estimating output gaps in Asia: a cross-country study. Journal of the Japanese International Economies 18(1): 115–136
14. Ha J, Fan K (2002) Price convergence between Hong Kong and the Mainland. Mimeo. Hong Kong Monetary Authority
15. Ha J, Leung C, Shu C (2002) A small macroeconomic model of Hong Kong. Mimeo. Hong Kong Monetary Authority
16. Harvey A (1985) Trends and cycles in macroeconomic time series. Journal of Business & Economic Statistics 3(3): 216–227
17. Harvey A (2008) Modeling the Phillips curve with unobserved components. Cambridge Working Papers in Economics No. 0805. Faculty of Economics, University of Cambridge
18. Harvey A, Jaeger, A (1993) Detrending, stylised facts and the business cycle. Journal of Applied Econometrics 8: 231–247
19. Henry O, Summers P (2003) Long swings and the implication for unit root tests: evidence from the Hong Kong deflation. Working Paper. University of Melbourne
20. Imai H (2010) Hong Kong's inflation and deflation under the US dollar peg: the Balassa-Samuelson effect or export price shocks? The Developing Economies 48(3): 319–44
21. Ireland, P (2007) Changes in the Federal Reserve's inflation target: causes and consequences. Journal of Money, Credit and Banking 39(8): 1851–2110
22. Jia Y, Sinclair T (2009) Permanent and transitory macroeconomic relationships between the US and China. Working Paper. George Washington University
23. Kim C-J, Manopimoke P, Nelson C (2013) Trend inflation and the nature of structural breaks in the New Keynesian Phillips Curve. Working Paper, forthcoming, *Journal of Money, Credit and Banking*
24. Lee J, Nelson C (2007) Expectation horizon and the Phillips curve: the solution to an empirical puzzle. Journal of Applied Econometrics 22: 161–178
25. Liu L, Tsang A (2008) Exchange rate pass-through to domestic inflation in Hong Kong. Working Paper No.2, Hong Kong Monetary Authority
26. Leung F, Chow K, Chan S (2009) Measures of trend inflation in Hong Kong. BIS Working Paper 49
27. Romer D (1993) Openness and inflation: theory and evidence. Quarterly Journal of Economics 100: 1169–1189
28. Stock J, Watson M (2007) Why has U.S. inflation become harder to forecast? Journal of Money, Credit and Banking 39(1): 3–33

Chapter 6
The State Space Representation and Estimation of a Time-Varying Parameter VAR with Stochastic Volatility

Taeyoung Doh and Michael Connolly

6.1 Introduction

Vector autoregressions (VARs) are widely used in macroeconomics to detect comovements among multiple economic time series. In a nutshell, VARs regress each time series onto various lags of multiple time series included in the model. When coefficients are assumed to be stable, each equation in a VAR becomes an example of a multiple linear regression. In the simplest form, error terms in the VAR are assumed to have constant variances.

While convenient, assuming time-invariant coefficients and variances turns out to be quite restrictive in capturing the evolution of economic time series. For example, US business cycle dynamics and monetary policy have changed substantially over the post-war period. To describe these changes in the VAR framework requires one to allow shifts in coefficients or volatility (e.g. Canova and Gambetti (2009); Clark (2009); Cogley and Sargent (2005); Cogley, Primiceri, and Sargent (2010); Primiceri (2005); Sims and Zha (2006)).

When time variation is introduced to either coefficients or volatility in a VAR, the state space representation of the VAR is typically used in empirical analysis to estimate unobserved time-varying coefficients or volatility. Since allowing time variation in coefficients or volatility introduces too many parameters unless restricted, the literature evolved in a way of introducing random processes to time-varying coefficients or volatility to avoid the "overparameterization" problem (Koop and Korobilis (2010)). The randomness in these parameters fits quite well with Bayesian methods because there is no strict distinction between fixed "true" parameters and random samples in the Bayesian tradition.

This chapter will discuss applying Bayesian methods for estimating a time-varying parameter VAR with stochastic volatility using the state space representa-

T. Doh (✉) • M. Connolly
Federal Reserve Bank of Kansas City, 1 Memorial Drive, Kansas City, MO 64198, USA
e-mail: Taeyoung.Doh@kc.frb.org; Michael.Connolly@kc.frb.org

Y. Zeng and S. Wu (eds.), *State-Space Models: Applications in Economics and Finance*, Statistics and Econometrics for Finance 1, DOI 10.1007/978-1-4614-7789-1_6,

tion of the VAR. Section 6.2 describes the state space representation and estimation methods for VARs. In particular, each step in the Bayesian estimation procedure of a time-varying parameter VAR with stochastic volatility is explained. Section 6.3 provides empirical analysis of a time-varying parameter VAR with stochastic volatility using three US macroeconomic variables. We will focus on implications of estimates for the time-varying trend and volatility of each variable during the recent period since the start of the recession of 2007–2009. Section 6.4 concludes.

6.2 State Space Representation and Estimation of VARs

6.2.1 State Space Representation

Let y_t be an $n \times 1$ vector of observed variables and q the length of lags. A canonical representation of a VAR(q)) with time-invariant parameters and volatility takes the following form;

$$y_t = c_0 + c_1 y_{t-1} + \cdots + c_q y_{t-q} + e_t \, , \, e_t \sim (0, \Sigma_e). \tag{6.1}$$

Since all the state variables are observed, there is no need to distinguish a state transition equation from a measurement equation in this case. However, if we allow for time variation in coefficients $(c_0, c_1, \cdots, c_q)$ or volatility (Σ_e), the model includes some unobserved components as state variables. To estimate these unobserved components based on the observed data, it is useful to distinguish a state transition equation from a measurement equation as in the canonical representation of a state space model. Here are examples of the state space representation of VARs with time-varying coefficients and volatility.

Example 6.1 (Time-Varying Parameter VAR with Time-Invariant Volatility).

$$y_t = X_t' \theta_t + \varepsilon_t \, , \, \varepsilon_t \sim \mathcal{N}(0, \Sigma_\varepsilon) \, , \, \text{(Measurement Equation)}, \tag{6.2}$$

$$\theta_t = \theta_{t-1} + v_t \, , \, v_t \sim \mathcal{N}(0, \Sigma_v) \, , \, \text{(State Transition Equation)}. \tag{6.3}$$

Here X_t includes a constant plus lags of y_t, and θ_t is a vector of VAR parameters. ε_t and v_s are assumed to be independent of one another for all t and s. Given the linear and Gaussian state space representation of the above VAR, we can apply the Kalman filter to estimate θ_t conditional on the time series of observed variables y_t. If we further allow a possible correlation between ε_t and v_t, the model studied in Cogley and Sargent (2001) belongs to this example.

Example 6.2 (VAR with Stochastic Volatility).

$$y_t = X_t' \theta + \Sigma_t \varepsilon_t \, , \, \varepsilon_t \sim \mathcal{N}(0, I_n) \, , \, \Sigma_t = \text{diag}(\sqrt{H_{i,t}}) \, , \, \text{(Measurement Equation)}, \tag{6.4}$$

$$\ln H_{i,t} = \ln H_{i,t-1} + u_{i,t} \, , \, u_{i,t} \sim \mathcal{N}(0, Q) \, , \, \text{(State Transition Equation)}. \tag{6.5}$$

Here Σ_t is a diagonal matrix whose diagonal elements are $\sqrt{H_{i,t}} (i = 1, \cdots, n)$. The measurement equation is a nonlinear function of the unobserved log stochastic volatility $(\ln H_{i,t})$. Hence, the Kalman filter is not applicable in this case. Simulation-based filtering methods are typically used to back out stochastic volatility implied by the observed data. The above model is close to the one studied in Clark (2011), who shows that allowing stochastic volatility improves the real time accuracy of density forecasts out of the VAR model.

Example 6.3 (Time-Varying Parameter VAR with Stochastic Volatility).

As emphasized by Sims (2001), ignoring time-varying volatility may overstate the role of time-varying coefficients in explaining structural changes in the dynamics of macroeconomic variables. Adding stochastic volatility to a time-varying parameter VAR will alleviate this concern. The time-varying parameter VAR with stochastic volatility can be described as follows:

$$y_t = X_t' \theta_t + \varepsilon_t \ , \ \varepsilon_t \sim \mathcal{N}(0, B^{-1} H_t B^{-1'}) \ , \text{(Measurement Equation)}, \quad (6.6)$$

$$\theta_t = \theta_{t-1} + v_t \ , \ v_t \sim \mathcal{N}(0, \Sigma_v) \quad (6.7)$$

$$\ln H_{i,t} = \ln H_{i,t-1} + u_{i,t} \ , \ u_{i,t} \sim \mathcal{N}(0, Q_i) \ , \text{(State Transition Equation)}. \quad (6.8)$$

Here H_t is a diagonal matrix whose diagonal element is $H_{i,t}$. B^{-1} is a matrix used to identify structural shocks from VAR residuals. If we allow for a correlation between ε_t and v_t, this model is the one studied in Cogley and Sargent (2005).[1]

6.2.2 Estimation of VARs

Without time-varying coefficients or volatility, the VAR can be estimated by equation-by-equation ordinary least squares (OLS) which minimizes the sum of residuals in each equation of the VAR. However, estimating VARs with time-varying coefficients or volatility requires one to use filtering methods to extract information about unobserved states from observed time series. For example, in the time-varying parameter VAR model with time-invariant volatility, we can use the Kalman filter to obtain the estimates of time-varying coefficients conditional on parameters determining the covariance matrix and initial values of coefficients.

Under the frequentist approach, we estimate the covariance matrix and initial values of coefficients first and obtain estimates of time-varying coefficients conditional on the estimated covariance matrix and initial values of coefficients. While conceptually natural, implementing this procedure faces several computational issues especially for a high-dimensional model. The likelihood is typically highly

[1] If we allow for time variation in the B matrix, the model becomes a time-varying structural VAR in Primiceri (2005). And we can also incorporate the time-varying volatility of v_t too to capture fluctuations in variances of innovations in trend components as in Cogley, Primiceri, and Sargent (2010).

nonlinear with respect to parameters to be estimated and maximizing it over a high-dimensional space is computationally challenging.

As emphasized by Primiceri (2005), Bayesian methods can deal efficiently with these types of models using the numerical evaluation of posterior distributions of parameters and unobserved states. The goal of Bayesian inference is to obtain joint posterior distributions of parameters and unobserved states. In many cases such as VARs with time-varying parameters or volatility, these joint distributions are difficult or impossible to characterize analytically. However, distributions of parameters and unobserved states conditional on each other are easier to characterize or simulate. Gibbs sampling, which iteratively draws parameters and unobserved states conditional on each other, provides draws from joint distributions under certain regularity conditions.[2]

As an illustration of Bayesian estimation methods in this context, consider Example 6.1. Denote z^T be a vector or matrix of variable z_t from $t = 0$ to $t = T$. In this model, unobserved states are time-varying coefficients θ_t and parameters are covariance matrices of VAR residuals and innovations in coefficients $(\Sigma_\varepsilon, \Sigma_v)$. Prior distributions for θ_0 and Σ_ε can be represented by $p(\theta_0)$ and $p(\Sigma_\varepsilon)$. The Bayesian estimation procedure for this model is described as follows:

(***Bayesian Estimation Algorithm for a Homoskedastic Time-Varying Parameter VAR***)

Step 1: Initialization

Draw Σ_ε from the prior distribution $p(\Sigma_\varepsilon)$.

Step 2: Draw VAR coefficients θ^T

The model is a linear and Gaussian state space model. Assuming that $p(\theta_0)$ is Gaussian, the conditional posterior distribution of $p(\theta_t|y^t, \Sigma_\varepsilon, \Sigma_v)$ is also Gaussian. A forward recursion using the Kalman filter provides expressions for posterior means and the covariance matrix.

$$\begin{aligned} p(\theta_t|y^t, \Sigma_\varepsilon, \Sigma_v) &= N(\theta_{t|t}, P_{t|t}), \\ P_{t|t-1} &= P_{t-1|t-1} + \Sigma_v, \\ K_t &= P_{t|t-1}X_t(X_t'P_{t|t-1}X_t + \Sigma_\varepsilon)^{-1}, \\ \theta_{t|t} &= \theta_{t-1|t-1} + K_t(y_t - X_t'\theta_{t-1|t-1}), \\ P_{t|t} &= P_{t|t-1} - K_tX_t'P_{t|t-1}. \end{aligned} \tag{6.9}$$

Starting from $\theta_{T|T}$ and $P_{T|T}$, we can run the Kalman filter backward to characterize posterior distributions of $p(\theta^T|y^T, \Sigma_\varepsilon, \Sigma_v)$.

$$\begin{aligned} p(\theta_t|\theta_{t-1}, y^T, \Sigma_\varepsilon, \Sigma_v) &= N(\theta_{t|t+1}, P_{t|t+1}), \\ \theta_{t|t+1} &= \theta_{t|t} + P_{t|t}P_{t+1|t}^{-1}(\theta_{t+1} - \theta_{t|t}), \\ P_{t|t+1} &= P_{t|t} - P_{t|t}P_{t+1|t}^{-1}P_{t|t}. \end{aligned} \tag{6.10}$$

[2] See Lancaster (2004, Chap. 4) for necessary conditions.

We can generate a random trajectory for θ^T using backward recursion starting with a draw of θ^T from $\mathcal{N}(\theta_{T|T}, P_{T|T})$ as suggested by Carter and Kohn (1994).

Step 3: Draw covariance matrix parameters for VAR coefficients Σ_v
Conditional on a realization for θ^T, innovations in VAR coefficients v_t are observable. Assuming the inverse-Wishart prior for Σ_v with scale parameter $\overline{\Sigma_v}$ and degree of freedom T_{v0}, the posterior is also inverse-Wishart[3]:

$$\begin{aligned} p(\Sigma_v|y^T,\theta^T) &= IW(\Sigma_{v,1}^{-1}, T_{v,1}), \\ \Sigma_{v,1} &= \overline{\Sigma_v} + \sum_{t=1}^{T} v_t v_t' \ , \ T_{v,1} = T_{v,0} + T. \end{aligned} \tag{6.11}$$

Step 4: Draw covariance matrix parameters for VAR residuals Σ_ε
Conditional on a realization for θ^T, VAR residuals ε_t are observable. Assuming the inverse-Wishart prior for Σ_ε with scale parameter $\overline{\Sigma_\varepsilon}$ and degree of freedom $T_{\varepsilon,0}$, the posterior is also inverse-Wishart:

$$\begin{aligned} p(\Sigma_\varepsilon|y^T,\theta^T) &= IW(\Sigma_{\varepsilon,1}^{-1}, T_{\varepsilon,1}), \\ \Sigma_{\varepsilon,1} &= \overline{\Sigma_\varepsilon} + \sum_{t=1}^{T} \varepsilon_t \varepsilon_t' \ , \ T_{\varepsilon,1} = T_{\varepsilon,0} + T. \end{aligned} \tag{6.12}$$

Step 5: Posterior inference
Go back to step 1 and generate new draws of θ^T, Σ_v, and Σ_ε. Repeat this $M_0 + M_1$ times and discard the initial M_0 draws. Use the remaining M_1 draws for posterior inference. Since each draw is generated conditional on the previous draw, posterior draws are generally autocorrelated. To reduce the autocorrelation, we can thin out posterior draws by selecting every 20th draw from M_1 draws, for example.

Although posterior draws are obtained from conditional distributions, their empirical distributions approximate the following joint posterior distribution $p(\theta^T, \Sigma_v, \Sigma_\varepsilon|y^T)$. Hence, integrating out uncertainties about other components of the model is trivial. For instance, if we are interested in the median estimate of θ^T, which integrates out uncertainties of Σ_v and Σ_ε, we can simply use the median value of M_1 draws of θ^T.

In the above example, conditional distributions of parameters and unobserved states are known. However, in models with stochastic volatility such as Examples 6.2 and 6.3, the conditional posterior distribution of volatility is known up to a constant. Without knowing the constant, we cannot directly sample from the conditional posterior distribution for stochastic volatility. Instead, we can use Metropolis–Hastings algorithm following Jacquier, Polson, and Rossi (1994) to generate posterior draws for stochastic volatility.

As an illustration, consider Example 6.3. Posterior simulation for time-varying coefficients θ^T and Σ_v are essentially the same as before. Below, I will describe steps to generate posterior draws for $H_{i,t}$ and B conditional on θ^T, y^T, and Σ_v.

[3] Notations here closely follow those in the appendix of Cogley and Sargent (2005).

(***Drawing Stochastic Volatility*** H_t ***and Covariance Parameters*** B)
Step 1 Given y^T, θ^T, we can generate a new draw for B. Notice the following relationship between VAR residuals ε_t and structural shocks u_t:

$$B\varepsilon_t = u_t. \tag{6.13}$$

Conditional on y^T and θ^T, ε_t is observable. Since B governs only covariance structures among different shocks, $\frac{n(n+1)}{2}$ elements of the matrix are restricted. For example, if B is the following 2×2 matrix,

$$B = \begin{pmatrix} 1 & 0 \\ B_{21} & 1 \end{pmatrix},$$

with $B_{21} \sim \mathcal{N}(\overline{B}_{21}, V_{21})$, the relation between ε_t and u_t implies the following transformed regressions.

$$\begin{aligned} \varepsilon_{1t} &= u_{1t} \\ (H_{2t}^{-.5}\varepsilon_{2t}) &= B_{21}(-H_{2t}^{-.5}\varepsilon_{1t}) + (H_{2t}^{-.5}u_{2t}). \end{aligned} \tag{6.14}$$

As explained by Cogley and Sargent (2005), the above regressions imply the normal posterior for B_{21}.

$$B_{21}|y^T, H^T, \theta^T \sim \mathcal{N}(\hat{B}_{21}, \hat{V}_{21}) \ , \ \hat{V}_{21} = (V_{21}^{-1} + \sum(\frac{\varepsilon_{1t}^2}{H_{2t}}))^{-1} \ , \ \hat{B}_{21} = \hat{V}_{21}(V_{21}^{-1}\overline{B}_{21} + \sum(\frac{\varepsilon_{1t}\varepsilon_{2t}}{H_{2t}})). \tag{6.15}$$

Step 2 Conditional on ε_t, we can write down the following state representation for H_t:

$$\sum_{j=1}^{n} B_{ij}\varepsilon_{jt} = \sqrt{H_{it}}w_{it} \ , \ w_{it} \sim i.i.d.\mathcal{N}(0,1), \tag{6.16}$$

$$\ln H_{i,t} = \ln H_{i,t-1} + u_{i,t} \ , \ u_{i,t} \sim \mathcal{N}(0, Q_i). \tag{6.17}$$

The above system is not linear and Gaussian with respect to H_{it}. The conditional posterior density of H_{it} is difficult to characterize analytically but known up to a constant.

$$\begin{aligned} p(H_{it}|H_{i,t-1}, H_{i,t+1}, y^T, \theta^T, B, Q_i) &\propto p(u_{it}|H_{it}, B)p(H_{it}|H_{i,t-1})p(H_{i,t+1}|H_{it}), \\ &\propto H_{it}^{-1.5} exp(-0.5\frac{u_{it}^2}{H_{it}}) exp(-0.5\frac{(\ln H_{it} - \mu_{it})^2}{0.5Q_i}), \\ \mu_{it} &= 0.5(\ln H_{i,t-1} + \ln H_{i,t+1}). \end{aligned} \tag{6.18}$$

We can use the Metropolis–Hastings algorithm which draws H_{it} from a certain proposal density $q(H_{it})$.[4] Each mth draw is accepted with probability α_m,

[4] One example of such a proposal density is $\mathcal{N}(\mu_{it}, 0.5Q_i)$.

$$\alpha_m = \frac{p(H_{it}^m | H_{i,t-1}^m, H_{i,t+1}^{m-1}, y^T, \theta^T, B, Q_i) q(H_{i,t}^{m-1})}{p(H_{it}^{m-1} | H_{i,t-1}^m, H_{i,t+1}^{m-1}, y^T, \theta^T, B, Q_i) q(H_{i,t}^m)}. \tag{6.19}$$

As shown by Jacquier, Polson, and Rossi (1994), this sampling scheme generates posterior draws for H_{it}.

6.3 Application: A Time-Varying Parameter VAR with Stochastic Volatility for Three US Macroeconomic Variables

As an application of state space modelling, we estimate a time-varying parameter VAR with stochastic volatility for US macroeconomic time series consisting of inflation, the unemployment rate, and the long-term interest rate.[5] The model is close to Cogley and Sargent (2005) but there are two main differences. First, we shut down the correlation VAR residuals and innovations in time-varying parameter transition equations. Second, we use the long-term interest rate rather than the short-term interest rate to cover the overall monetary policy stance at the recent zero lower bound period. The estimated model can be casted into the following state space representation like Example 6.3 in the previous section:

$$\begin{aligned} y_t &= \theta_{0,t} + \theta_{1,t} y_{t-1} + \theta_{2,t} y_{t-2} + \varepsilon_t \ , \ \varepsilon_t \sim \mathcal{N}(0, B^{-1} H_t B^{-1\prime}), \\ \theta_t &= \theta_{t-1} + v_t \ , \ \theta_t = [\theta_{0,t}', vec(\theta_{1,t})', vec(\theta_{2,t})']' \ , \ v_t \sim \mathcal{N}(0, \Sigma_v), \\ \ln H_{i,t} &= \ln H_{i,t-1} + u_{i,t} \ , \ u_{i,t} \sim \mathcal{N}(0, Q_i) \ , \ (i = 1,2,3). \end{aligned} \tag{6.20}$$

y_t contains the dynamics of three variables in the order of the 10-year yield, core PCE (Personal Consumption Expenditure price index) inflation, and the civilian unemployment rate. The sample period is from 1960:Q1 to 2011:Q4. We assume that B is a lower triangular matrix whose diagonal elements are all equal to 1.

6.3.1 Priors

Priors are set in the same way as Cogley and Sargent (2005), using pre-sample data information from 1953:Q2 to 1959:Q4.[6]

First, we estimate seemingly unrelated regressions for the pre-sample data and use the point estimate of coefficients as the prior mean for θ_0 and its asymptotic variance $\overline{P}$ as the prior variance. Second, we use an inverse-Wishart distribution as the prior for Σ_v with degree of freedom $T_0 = 22$ and scale matrix $\overline{\Sigma_v} = T_0 \times 0.001 \times$

[5] Doh (2011) estimates the same model with shorter sample data and focuses on the time-varying relationship between inflation and unemployment.

[6] For pre-sample data, we use total PCE inflation because core PCE inflation is not available for this period.

$\overline{P}$. Third, the prior distribution of the log of the initial volatility is set to the normal distribution whose mean is equal to the variance of regression residuals using the pre-sample data. The prior variance is set to 10. Fourth, the prior distributions of elements in B are normal with the mean equal to 0 and the covariance matrix equal to $10{,}000 \times I_3$. Finally, the prior for the variance of the innovation to volatility process is inverse gamma with the scale parameter equal to 0.01^2 and the degree of freedom parameter equal to 1.

6.3.2 Posterior Simulation

We generate 100,000 posterior draws and discard the first 50,000 draws. Among the remaining 50,000 draws, we use every 20th draw to compute posterior moments. Following Cogley and Sargent (2005), we throw away draws implying the non-stationarity of the VAR. Hence, if θ^T contains coefficients which indicate the non-stationarity of the VAR at any point of time, we redraw θ^T until the stationarity is ensured all the time.

Consider a companion VAR(1) for $[y_t', y_{t-1}']'$. Technically speaking, stationarity is guaranteed if all the eigenvalues of $A_t = \begin{pmatrix} \theta_{1,t} & \theta_{2,t} \\ I_3 & 0 \end{pmatrix}$ are inside the unit circle. The truncation is particulary useful when we back out time-varying trend components from estimated coefficients. When we use the companion form for long-horizon forecasts, the stochastic trend in $[y_t', y_{t-1}']'$ can be approximated as $(I - A_t)^{-1}[\theta_{0,t}', 0]'$. Below, we will use this approximation to obtain posterior estimates of time-varying trends in y_t.

6.3.3 Posterior Estimates of Time-Varying Trends and Volatility

Over the last 50 years, the US economy has shown substantial changes. In particular, there is considerable evidence that trend inflation and volatility of inflation rose during the mid-1970s and the early 1980s but then declined after the Volcker disinflation.[7] Also, the decline in the volatility of inflation is one primary factor for explaining the decline in the term premium of long-term government bonds since the late 1980s (Wright (2011)). On the other hand, economic slack seems to be less important in predicting inflation since 1984.[8] In addition, the volatility of real activity declined since the mid-1980s.[9]

Most papers on these issues rely on data before the most recent recession that started in late 2007. The severity of the recession and the unprecedented policy

[7] For example, see Cogley, Primiceri, and Sargent (2010) and papers cited there.

[8] See Doh (2011) and papers discussed there.

[9] See Canova and Gambetti (2009) and papers cited there.

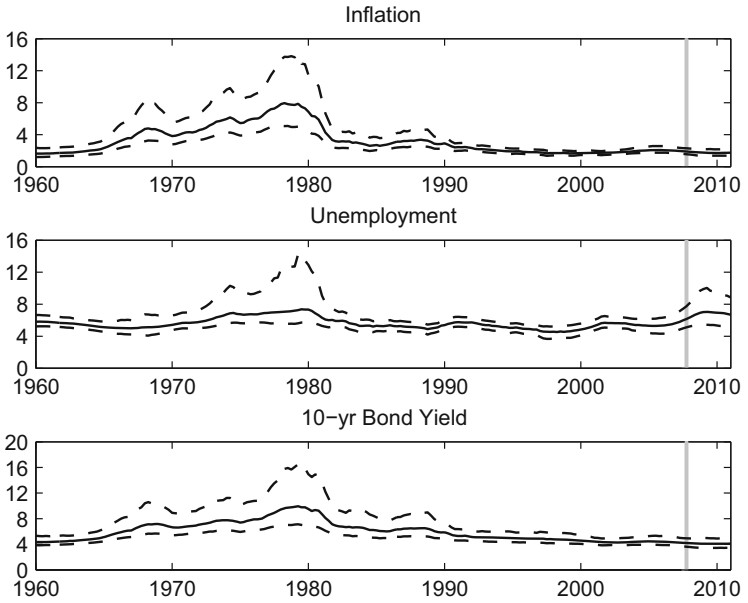

Fig. 6.1 Time-varying trend. The *solid line* stands for posterior median estimates and *dashed lines* for estimates of the 70% highest posterior density regions. The *vertical bar* indicates the fourth quarter of 2007 when the recession started.

actions including keeping the short-term interest rate at the effective zero lower bound and implementing large-scale asset purchases raised a question about the robustness of the above-mentioned changes.

Our time-varying parameter VAR model with stochastic volatility can shed light on this question. First of all, we can investigate if the recession and the subsequent policy responses affected mainly trend components or cyclical components of the three macroeconomic variables. Impacts on cyclical components are expected to be temporary while those on trend components are supposed to be more persistent. Second, we can do a similar exercise for the volatility of three macroeconomic variables. For instance, we can compute the short-run and the long-run volatility of the three variables in the VAR and see if there might be shifts during the recent period.

Our posterior estimates of time-varying trends in Fig. 6.1 suggest that trends in nominal variables such as inflation and the long-term interest rate were little affected by the recent episode while the trend unemployment rate was affected more substantially. This result is interesting because the level of all the variables moved significantly during the same period as shown in Fig. 6.2. For the inflation rate, movements in the trend component explain about 9% of the overall movements in the level of the inflation rate.[10] The relative contribution of the trend component further declines to about 6% for the nominal 10-year bond yield. In contrast, the

[10] This calculation is based on comparing the standard deviation of each variable during the relevant period.

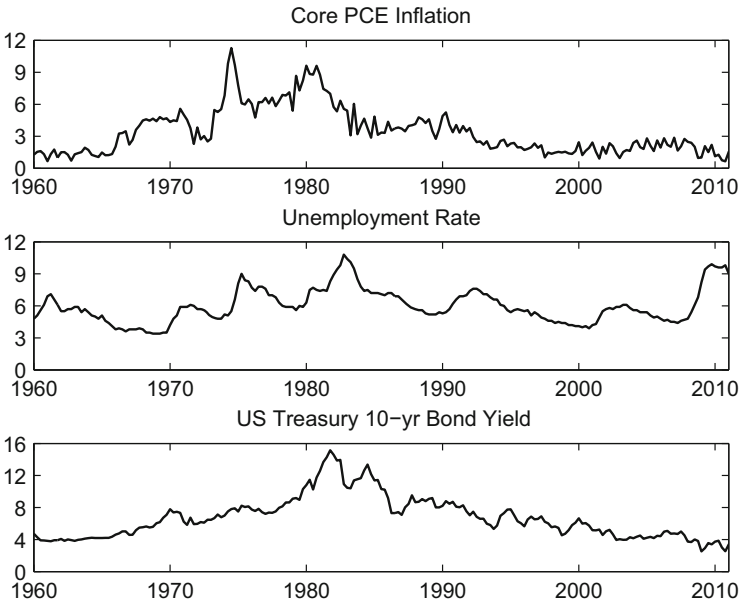

Fig. 6.2 Data.

movement in the trend unemployment rate explains more than 15% of the overall movement in the unemployment rate for the same period.

These differences in the relative contribution of time-varying trends across variables suggest that it will take a longer time for the unemployment rate to return to its pre-recession level than other variables. However, it is possible that the relatively small role of trend component volatility was driven by our assumption of constant volatility of innovations in time-varying coefficients. To check the robustness of our finding, we allowed for the time-varying volatility for innovations in θ_t in an alternative specification. Even in this version of the model, we got essentially the same relative contribution of trend components during the recent period.[11]

[11] The drawback of this generalization of time-varying volatility is that so many volatility estimates become explosive during the mid-1970s, casting doubts on the convergence property of the model estimates. For the model without stochastic volatility for innovations in θ_t, we do not observe such a convergence issue.

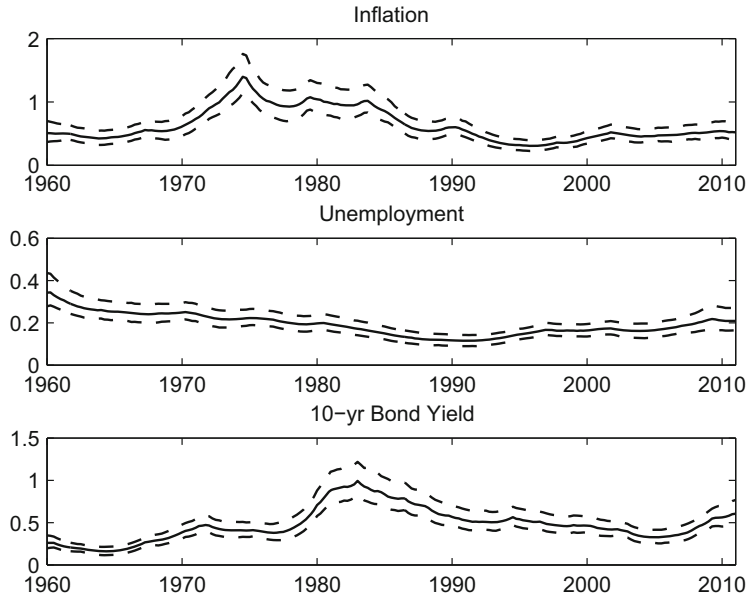

Fig. 6.3 Time-varying volatility of residuals. The *solid line* stands for posterior median estimates and *dashed lines* for estimates of the 70% highest posterior density regions.

We can apply the similar trend-cycle decomposition for volatility estimates, too. Following Cogley, Primiceri, and Sargent (2010), we approximate the unconditional variance of $[y_t', y_{t-1}']'$ by

$$\sum_{h=0}^{\infty} (A_t)^h B^{-1} H_t (B^{-1})' ((A_t)^h)'. \tag{6.21}$$

This unconditional variance is dominated by slowly moving trend components of volatility estimates while H_t is mainly based on the short-run movements of volatility estimates. The volatility of residuals went up for all the variables as shown in Fig. 6.3. In addition, the unconditional volatility of the unemployment rate moved up more noticeably during the recent period to the historical peak level as shown in Fig. 6.4. This finding suggests that the increase in volatility since the recession may not be driven by a common factor affecting the entire economy. This interpretation is in line with the observation in Clark (2009) that the recent increase in volatility is concentrated on certain sectors of the economy (goods production and investment but not services components, total inflation but not core).

Overall, our posterior analysis indicates that the trend and volatility of the unemployment rate have experienced substantial changes during the recent episode while core inflation and the nominal long-term interest rate have been relatively immune from these changes. Analyzing causes of different responses across variables may require a more structural model of the economy built on decisions of agents. Our analysis can be a starting point for such a project.

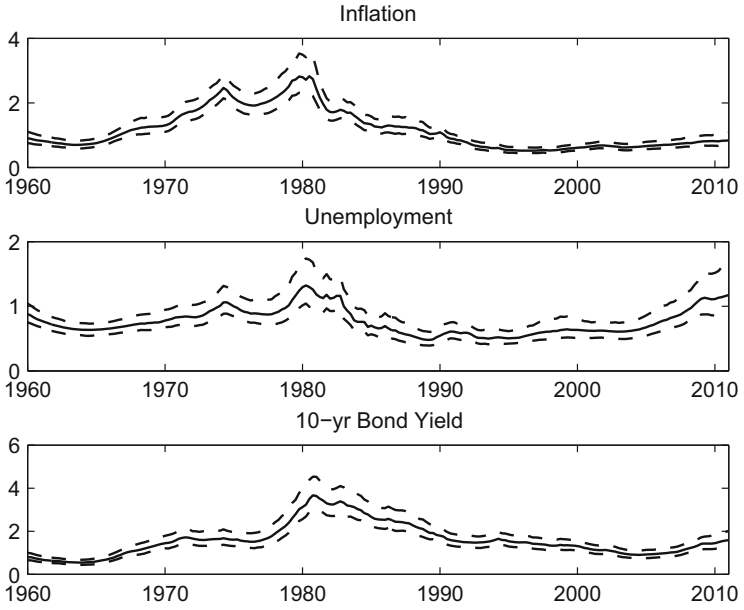

Fig. 6.4 Time-varying unconditional volatility. The *solid line* stands for posterior median estimates and *dashed lines* for estimates of the 70% highest posterior density regions.

6.4 Conclusion

VARs are widely used in macroeconomics and finance to describe the historical dynamics of multiple time series. When the VAR is extended to incorporate time-varying coefficients or volatility to capture structural shifts in the economy over time, the state space representation is necessary for estimation. Applying the Kalman filter in the state space representation of a time-varying parameter VAR provides estimates of unobserved time-varying coefficients that we are interested in. Also, we can obtain estimates of time-varying volatility by applying the simulation-based filtering method to the state space representation of the volatility process.

We illustrate the value of applying the state space representation to the time-varying parameter VAR with stochastic volatility by estimating such a model with three US macro variables. Our empirical analysis suggests that the recession of 2007–2009 was driven by a particulary bad shock to the unemployment rate which increased the trend and volatility of the unemployment rate substantially. In contrast, nominal variables such as the core PCE inflation rate and the 10-year Treasury bond yield have exhibited relatively less noticeable movements in terms of their trend and volatility. Further identifying underlying causes of unemployment dynamics may require us to go beyond the small scale time-varying parameter VAR model that we are considering in this chapter.

Acknowledgments The views expressed here are the opinions of the authors only and do not necessarily represent those of the Federal Reserve Bank of Kansas City or the Federal Reserve System.

References

1. Canova, F. and L. Gambetti (2009): "Structural changes in the US economy: Is there a role for monetary policy?,"*Journal of Economic Dynamics and Control*, **33**, 477-490.
2. Carter, C. and R. Kohn (1994): "On Gibbs sampling for state space models,"*Biometrika*, **81**, 541-553.
3. Clark, T. (2009): "Is the Great moderation over? An empirical analysis,"*Economic Review*, 2009:Q4, 5-42, Federal Reserve Bank of Kansas City.
4. Clark, T. (2011): "Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility, "*Journal of Business and Economic Statistics*, **29**(3), 327-341.
5. Cogley, T. and T. Sargent (2001): "Evolving post-world war II U.S. inflation dynamics,"*NBER Macroeconomics Annual*, **16**, Edited by B.S. Bernanke and K. Rogoff, Cambridge,MA:MIT Press, 331-373.
6. Cogley, T. and T. Sargent (2005): "Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S.,"*Review of Economic Dynamics*, **8**(2), 262-302.
7. Cogley, T., G. Primiceri, and T. Sargent (2010): "Inflation-gap persistence in the US,"*American Economic Journal: Macroeconomics*, **2**(1), 43-69.
8. Doh, T. (2011): "Is unemployment helpful for understanding inflation?,"*Economic Review*, 2011:Q4, 5-26, Federal Reserve Bank of Kansas City.
9. Jacquier, E., N. Polson, and P. Rossi (1994): "Bayesian analysis of stochastic volatility,"*Journal of Business and Economic Statistics*, **12**, 371-417.
10. Koop, G. and D. Korobilis (2010): "Bayesian multivariate time series methods for empirical macroeconomics,"*Manuscript*, University of Strathclyde.
11. Lancaster, T. (2004): "An introduction to modern Bayesian econometrics," Malden, MA: Blackwell Publishing.
12. Primiceri, G. (2005): "Time-varying structural vector autoregressions and monetary policy," *Review of Economic Studies*, **72**, 821-852.
13. Sims, C. (2001): "Comment on Cogley and Sargent (2001),"*NBER Macroeconomics Annual*, **16**, Edited by B.S. Bernanke and K. Rogoff, Cambridge,MA:MIT Press, 373-379.
14. Sims, C. and T. Zha (2006): "Were there regime switches in macroeconomic policy?," *American Economic Review*, **96**(1), 54-81.
15. Wright, J. (2011): "Term Premiums and Inflation Uncertainty: Empirical Evidence from an International Panel Dataset, "*American Economic Review*, **101**, 1514-1534.

Chapter 7
A Statistical Investigation of Stock Return Decomposition Based on the State-Space Framework

Jun Ma and Mark E. Wohar

7.1 Introduction

John Cochrane (2011) writes an incredibly insightful 2011 Presidential Address for the American Finance Association. The first part of the address surveys the empirical findings, theories, and applications related to fluctuations in discount rates (expected returns) that are reported in the extant literature with respect to time series analysis.[1] Finance theory tells us that the price of an asset is equal to the discounted expected future cash flows that the asset generates. It follows that there are two primary factors that can influence prices: expectations regarding discount rates (expected returns) and expectations regarding future cash flows.

We can think of two branches of literature. We begin with the first branch of literature that suggests that aggregate expected cash flows do not generate significant aggregate stock price variability. Most of the work in this area has concluded that expectations of future excess returns rather than future real dividend growth or real interest rates are responsible for most of the fluctuations in stock prices. Campbell (1991), for aggregate US stock returns, employs the Campbell and Shiller (1988a) log-linearization of the dividend-price ratio in order to decompose the variance of equity returns into three components: (1) expected future excess returns (i.e., risk

J. Ma (✉)
Department of Economics, Finance, and Legal Studies, University of Alabama, Tuscaloosa, AL 35487, USA
e-mail: jma@cba.ua.edu

M.E. Wohar
Department of Economics, University of Nebraska at Omaha, Omaha, NE 68182, USA
e-mail: mwohar@unomaha.edu

[1] In the second half of the address he discusses how fluctuations in discount rates influence portfolio theory, capital structure, the cross section of returns, and macroeconomics. The focus in this chapter relates to issues discussed in the first half of the address.

Y. Zeng and S. Wu (eds.), *State-Space Models: Applications in Economics and Finance*, Statistics and Econometrics for Finance 1, DOI 10.1007/978-1-4614-7789-1_7,

premium), (2) future real risk free interest rate (the other component of discount rates) and (3) future dividend cash flows. Campbell (1991) finds that cash flow news explains only 15% in the variation of market returns over the period 1952–1988, while Campbell and Vuolteenaho (2004) find that cash flow news explains only 20% of the variation in returns.[2]

Campbell and Shiller (1988a, 1988b, 1998), Campbell (1991), Shiller and Beltratti (1992), Cochrane (1992), and Campbell and Ammer (1993) decompose the variance of stock returns and bond returns into contributions of real dividend growth and other factors. In particular, Cochrane (1992) and Campbell and Ammer (1993) break stock return movements into contributions of real dividend growth, real interest rate, and excess stock returns. They find that US return innovations are primarily driven by news about future returns (for stocks) and inflation news (for bonds). They find that most of the variability in stock returns is due to innovations in excess returns and not real dividend or real interest rates.

Cochrane (2001, p. 398) notes that

> It is nonetheless an uncomfortable fact that almost all variation in price/dividend ratios is due to variation in expected excess returns. How nice it would be if high prices reflected expectations of higher future cash flows.

Cochrane (2008a, p. 1573) is still consistent with his view after 7 years as he points out the implications of the finding that excess returns explain the majority of the movements in the price-dividend ratio.

> If all market price-dividend ratio variations come from varying expected returns and none from varying expected growth in dividends or earnings, much of the rest of finance still needs to be rewritten.

The second branch of the literature studies the volatility of stock return (specifically firm level returns) as opposed to the price-dividend ratio (see, for example, Vuolteenaho, 2002). These studies find that cash flows do generate enough variability in stock returns to be consistent with the variability of cash flows. For example, Vuolteenaho (2002) extends the Campbell (1991) analysis of aggregate stock returns to a similar decomposition of firm level stock returns and finds that they are driven primarily by cash flow news.[3] Campbell and Ammer (1993) and Ammer and Mei (1996) extend the above discussed approach to multiple assets, so that the covariance between national stock returns can be characterized in terms of elements similar to time-varying discount rates and the value of future cash flows. Campbell

[2] Consistent with the previously mentioned finding, the extant literature finds that the price-dividend ratio predicts aggregate stock returns but not dividends. There is a vast literature investigating stock return predictability which we do not survey here.

[3] The literature has attempted to reconcile the conflicting results between aggregate returns and firm level returns in the following way. A number of authors have applied the analysis to the cross-section of firm stock returns (e.g., Vuolteenaho, 2002; Cohen, Polk, and Vuolteennaho 2003; Callen and Segal 2004). Expected return news is likely to be highly correlated across firms, while cash flow news is very much firm specific and can almost be completely diversified away in aggregate portfolios. The literature has inferred that firm level news about profitability, unlike returns, must be primarily idiosyncratic in nature and thus almost completely diversifiable. Cochrane (2001, p. 399) concludes,

and Hamao (1992) argue that if asset returns in different countries are generated by an international multivariate linear factor model, the conditional means of the excess returns must move together, as linear combinations of some common risk premiums.

Using an alternative approach to the variance decomposition approach discussed above, Binsbergen and Koijen (2010) employ a state-space framework using US annual data within a present value model to estimate the expected return and expected dividend growth rates of the aggregate US stock market. Their approach aggregates information contained in the price-dividend ratio and dividend growth rates to predict expected returns and dividend growth rates. They treat conditional expected returns and expected dividend growth as latent variables that follow an exogenously specified time series process. They combine this model with the Campbell and Shiller (1988a, 1988b) present value model to derive the implied dynamics of the price-dividend ratio. They find that both expected dividend growth rates and expected returns are time varying and persistent but that expected returns are more persistent than expected dividend growth rates. They also find that the filtered returns for expected returns and expected dividend growth are good predictors of realized returns and realized dividend growth rates, with R^2 ranging from 8.2% to 8.9% for returns and 13.9% to 31.6% for dividend growth rates. Finally, they estimate a process for dividend growth and back out expected returns. They find that expected returns contribute more to fluctuations in the price-dividend ratio than do dividend growth rates.

Balke and Wohar (2002) also estimate unobserved expectations of market fundamentals employing a more general state-space model than Binsbergen and Koijen (2010). An attractive feature of their state-space framework is that it allows for a parsimonious specification of low frequency movements in market fundamentals. A vector auto-regression (VAR) in levels may have difficulty capturing low frequency movements in small samples. Balke and Wohar (2002) model the dynamics of the log price-dividend ratio along with short-term and long-term interest rates, real dividend growth and inflation. One advantage of the state-space approach that they employ is that they can parsimoniously model the low frequency movements present in the data. They find that if one allows persistent changes, albeit small, in real dividend growth, interest rates and inflation (but not excess returns), then expectations of real dividend growth and real interest rates become significant contributors to stock price fluctuations. They also show that stock price decompositions

---

Much of the expected cash-flow variation is idiosyncratic while the expected return variation is common, which is why variation in the index book/market ratio, like variation in the index dividend/price ratio is almost all due to varying expected returns.

The results suggest that fluctuation in expected profitability can explain a large portion of the variation in firm-level returns, book-to-market ratios, and earnings-to-price ratios. The studies attribute the difference between the firm level and aggregate results to the relative strength of the idiosyncratic components of cash flow variation versus the systematic components of expected returns.

are very sensitive to assumptions about which market fundamental has a permanent component.[4]

Surprisingly, studies that investigate the variance decomposition of stock return in an effort to find the contribution that particular factors have on stock prices have focuses on point estimates and have ignored issues of inference.[5] The current chapter employs Cochrane's (2011) annual data covering the period 1947–2009 to illustrate the point that focusing on point estimates without regard to conducting analysis of inference can lead to inaccurate conclusions that have weak statistical foundation.[6] When issues of inference are considered we find that there is little evidence to support the notion that either expected returns or expected dividend growth contributes to movements in the price-dividend ratio. In an effort to explore this finding in detail, we employ a state-space modeling framework. We find that within this framework, it is the existence of weak identification combined with a low signal-to-noise ratio that leads to the conclusion that there is too much uncertainty to make any claims about the relative contribution of expected returns and expected dividend growth to movements in the price-dividend ratio. We propose a procedure that could potentially correct for the inference problem and offer more reliable results. The corrected inference indicates that the large contribution of the expected returns to fluctuations in the price-dividend ratio found in previous studies has no statistical significance. We also evaluate the statistical significance in the variance decomposition as well as in the state-space modeling framework.

The remainder of the chapter is organized as follows. Section 7.2 describes the stock return variance decomposition based on VAR framework and proposes a bootstrap procedure to document the uncertainty of such decomposition in this approach. Section 7.3 presents the state-space decomposition of stock returns. Section 7.4 discusses the existence of spurious inference in the presence of weak identification. Section 7.5 concludes the chapter.

[4] When they allow excess stock returns to have a permanent component (but not real dividend growth) then it is excess stock returns that become the important contributor to fluctuations to stock price movements, while real dividend growth is not. They conclude that the data is not informative about which one of the above models is the appropriate model. The important finding of Balke and Wohar (2002) that the factor with the largest degree of persistence is the factor that contributes most to movements in the price-dividend ratio is also consistent with the findings of this chapter as well as those of Binsbergen and Koijen (2010) who found that the persistent expected excess return factor explains most of the movements in stock prices.

[5] Although some papers such as Campbell (1991) and Larrain and Yogo (2008) have conducted inference exercises for such decompositions, the rest of the large literature have largely ignored issues of inference.

[6] We thank John Cochrane for generously providing all the data used in this chapter which he also used in Cochrane's (2011).

7.2 VAR Variance Decomposition of the Stock Prices

Consider the log-linearization of the stock return due to Campbell and Shiller (1988a):

$$r_{t+1} = \kappa + \Delta d_{t+1} + \rho \cdot pd_{t+1} - pd_t, \tag{7.1}$$

where r is the log return, Δd is the log dividend growth, pd is the log price-dividend ratio, $\rho = 1/(1+e^{-\bar{pd}})$, and $\bar{pd}$ is the steady-state price/dividend ratio, κ is a constant.

Iterating the above equation forward and taking expectations, we obtain:

$$pd_t = \frac{\kappa}{1-\rho} + E_t \sum_{j=0}^{\infty} \rho^j (\Delta d_{t+1+j} - r_{t+1+j}). \tag{7.2}$$

Equation (7.2) says that the price-dividend is the sum of discounted future expected dividend growth. Therefore, if the price-dividend varies, its variation must come from either the expected future dividend growth (cash flow or CF) or the expected future returns (discount rate or DR). CF is more related to firm fundamentals because of its link to production, while DR news can reflect time-varying risk aversion or investor sentiment. Since we do not directly observe these expectations, we have to make certain assumptions about the information set (such as past price-dividend ratios, dividends growth, and stock returns) and explore possible dynamics patterns in these variables to estimate the agent's expectations.

Following seminal articles by Campbell (1991) and Campbell and Ammer (1993) it has become common to employ VAR-based return decomposition to determine the contribution that cash flow (CF) news and discount rate (DR) news have on fluctuations in stock prices. Although some papers such as Campbell (1991) and Larrain and Yogo (2008) have conducted inference for such variance decomposition, we attempt to provide more extensive inference exercises in order to evaluate the statistical significance of the stock prices decompositions.

Next we aim to employ a nonparametric bootstrap procedure to explicitly take into account the parameter uncertainty of such variance decompositions in order to illustrate some of the statistical issues in this type of exercise. We show that the empirical distributions of the bootstrapped contributions of returns and dividends both have heavy tails. Following this literature we directly model the log return along with the log price-dividend ratio in a bivariate VAR. The contribution of the expected future returns can be computed directly using the model parameter estimates while the contribution of the expected future dividend growth is backed out using (7.2).

Using Cochrane's (2011) annual data from 1947 to 2009 we estimate the VAR with four lags and implement the variance decomposition. The variance decomposition result is presented in Table 7.1. Similar to what have been documented in this literature, the variance decomposition seems to attribute most of the stock price variation to the expected returns but little to the dividends. In order to investigate

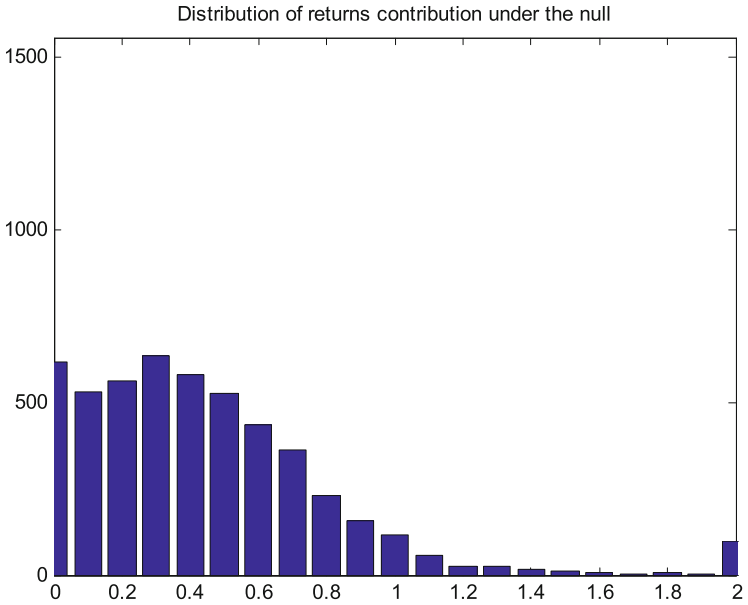

Fig. 7.1 Empirical distribution of returns contribution under the null.*Note*: The graph is based on 5,000 replications; the scale of *y*-axis is made the same as that in Fig. 7.2 to facilitate comparison; the last bar on the right counts the percentage of the contributions greater than 200%; because of negative covariance the contributions are often greater than 100%.

the statistical significance of the decomposition results, we resort to a nonparametric bootstrap procedure. Since the variance contribution estimate is highly nonlinear in the parameter estimates, the distribution of the variance contribution must not be normal. Therefore we prefer an inference based on the bootstrapping procedure.

In our bootstrapping exercise, we use the parameter estimates and the estimated residuals to generate data from the VAR model. In this way, we explicitly take into account the actual fat-tailed residuals that are commonly observed in the equity market. To generate the distribution under the null we set all coefficients in the returns equation to zero. In doing this we assume no return forecastability as the null hypothesis and also the benchmark case. Specifically, we follow the procedure of nonparametric bootstrapping with replacement in Davidson and MacKinnon (2004). For each set of bootstrapped data, we estimate the VAR model, compute and record the variance decomposition results. The number of draws is 5,000. We plot the empirical distributions of the contributions of expected returns and expected dividend growth to price-dividend variations under the null hypothesis in Figs. 7.1 and 7.2, respectively. The empirical distributions of the contributions allow us to evaluate the statistical significance of the point estimates of the variance decomposition. Notice that in Fig. 7.1 even the data are generated from the null of no return predictability, the return contributions are often estimated close to or even great than 1. This has to do with the fact that the variance contribution estimate is highly nonlinear in the

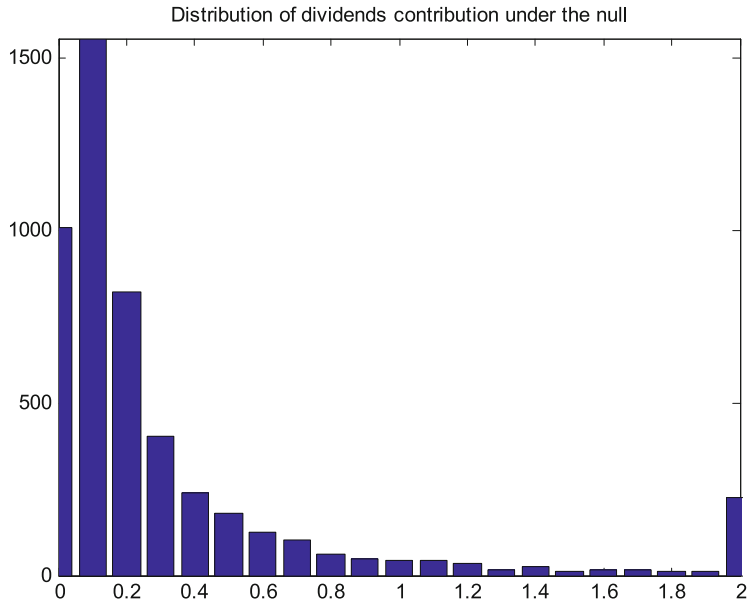

Fig. 7.2 Empirical distribution of dividends contribution under the null. *Note*: The graph is based on 5,000 replications; the scale of *y*-axis is made the same as that of Fig. 7.1 to facilitate comparison; the last bar on the right counts the percentage of those contributions greater than 200%; because of negative covariance the contributions are often greater than 100%.

Table 7.1 VAR variance decomposition (four lags)

Variance decomposition of price-dividend ratio (%)	
Variance due to expected return	64.32%
Variance due to expected dividend growth	7.81%
Covariance part	27.87%

Note: Annual data from 1947 to 2009 is used; the covariance part is constructed as 100% minus the first two contributions.

parameter estimates and that makes it very sensitive to even small changes in the parameter estimates.

Both distributions turn out to have heavy tails at the right end. In particular, the 95th percentile for returns contribution is 110.75%, which is far larger than the point estimate 64.32% in Table 7.1. Therefore, the seemingly large role of the expected returns is in fact insignificant at 5% level. Similarly, the dividend contribution is also insignificant. The data does not seem informative enough to tell us whether it is expected returns or expected dividends that contribute most to the price-dividend variation.

We conduct a second bootstrap exercise for a robustness check. In the estimated VAR, the price-dividend ratio appears very persistent as the sum of its four AR coefficients is 0.9695, very close to 1. Therefore, to study the effect of the very persistent price-dividend ratio on the empirical distributions of the variance decom-

position, we bootstrap 5,000 sets of data by further imposing the sum of the four AR coefficients for price-dividend to be unity. The resulting empirical distributions of the variance decomposition turn out to be quite similar to those obtained in the first exercise but more heavily tailed on the right. For example, the 95th percentile for returns contribution is 133.20%, even greater than that from the first bootstrap exercise.

7.3 The State-Space Model for Decomposing Stock Prices

The above section employs VAR to illustrate that the stock price decomposition involves a great deal of uncertainty. To better understand where this uncertainty comes from we build a state-space framework for decomposing the stock prices. Furthermore, the state-space model can naturally model the unobservable expectations in the stock price decomposition as latent factors and extract them through filtering procedures. Balke and Wohar (2002) and Binsbergen and Koijen (2010) have provided such examples. Furthermore, this latent factor approach within the state-space framework can capture the long-run serial correlations that a VAR with finite number of lags have difficulty in doing, since the state-space model usually leads to a reduced-form of Vector Auto-Regressive Moving Average (VARMA) model.

Let us denote the conditional expectation of dividends growth by $\bar{g}_t = E_t[\Delta d_{t+1} - a_g]$, the conditional expectation of returns by $\bar{\mu}_t = E_t[r_{t+1}] - a_\mu$, and model them both as stationary $AR(p)$ processes:

$$\phi_g(L) \cdot \bar{g}_t = \varepsilon_t^g, \tag{7.3}$$

$$\phi_\mu(L) \cdot \bar{\mu}_t = \varepsilon_t^\mu, \tag{7.4}$$

where $\phi_i(L) = 1 - \sum_{j=1}^{p} \phi_{i,j} L^j$, $i = g, \mu$ and $\phi_{i,0} = 1$. And $\varepsilon_t^i, i = g, \mu$ are shocks to the expectation processes. Each shock is *i. i. d*, respectively, with variances $Var(\varepsilon_t^g) = \sigma_g^2, Var(\varepsilon_t^\mu) = \sigma_\mu^2$. The shocks may be contemporaneously correlated. At each period, the realized dividends growth and returns are then their expected values plus any realized (news) shocks:

$$\Delta d_{t+1} = a_g + \bar{g}_t + \varepsilon_{t+1}^d, \tag{7.5}$$

$$r_{t+1} = a_\mu + \bar{\mu}_t + \varepsilon_{t+1}^r, \tag{7.6}$$

where a_g, a_μ are average dividend growth and return, respectively. ε_{t+1}^d and ε_{t+1}^r are the realized (news) shocks to the realized dividends growth and returns. Again each shock is *i. i. d*, respectively, with variances $Var(\varepsilon_t^d) = \sigma_d^2, Var(\varepsilon_t^r) = \sigma_r^2$. The shocks may be contemporaneously correlated.

Such state-space model can be most efficiently estimated using the Kalman filter as long as the model is linear and the shocks are normally distributed. Once the hyper-parameters are estimated the latent factors (the expectations) can be estimated as the filtered estimate from the realized values.

To facilitate the computation, write out the companion form of (7.3):

$$X_t^g = F_g \cdot X_{t-1}^g + V_t^g, \tag{7.7}$$

where $X_t^g = (\bar{g}_t, \cdots, \bar{g}_{t-p+1})', V_t^g = (\varepsilon_t^g, \cdots, 0)'$, and F_g is the proper companion matrix. It is then straightforward to derive the contribution of dividends growth (dropping constants):

$$E_t \sum_{j=0}^{\infty} \rho^j (\Delta d_{t+1+j}) = e_1' \cdot (I - \rho \cdot F_g)^{-1} \cdot X_t^g, \tag{7.8}$$

where e_1 is the selection vector that has 1 on its first element and zero elsewhere. Likewise, write out the companion form of (7.4):

$$X_t^\mu = F_\mu \cdot X_{t-1}^\mu + V_t^\mu, \tag{7.9}$$

where $X_t^\mu = (\bar{\mu}_t, \cdots, \bar{\mu}_{t-p+1})', V_t^\mu = (\varepsilon_t^\mu, \cdots, 0)'$, and F_μ is the corresponding companion matrix. We then derive the contribution of returns (dropping constants):

$$E_t \sum_{j=0}^{\infty} \rho^j (r_{t+1+j}) = e_1' \cdot (I - \rho \cdot F_\mu)^{-1} \cdot X_t^\mu. \tag{7.10}$$

Combining (7.7) and (7.9), the stock price decomposition (7.2) becomes:

$$pd_t = e_1' \cdot (I - \rho \cdot F_g)^{-1} \cdot X_t^g - e_1' \cdot (I - \rho \cdot F_\mu)^{-1} \cdot X_t^\mu. \tag{7.11}$$

Equation (7.11) integrates the information of the expectations dynamics into the Campbell–Shiller decomposition formulae and shows that the price-dividend variation comes from two major sources $Var[e_1' \cdot (I - \rho \cdot F_g)^{-1} \cdot X_t^g]$ and $Var[e_1' \cdot (I - \rho \cdot F_\mu)^{-1} \cdot X_t^\mu]$, if ignoring the correlation term for illustrative purpose.

Before estimating the hyper-parameters in the above state-space model and extracting the expectations we notice that the Campbell and Shiller identity (7.1) implicitly imposes a restriction on the four shocks $\varepsilon_{t+1}^g, \varepsilon_{t+1}^\mu, \varepsilon_{t+1}^d, \varepsilon_{t+1}^r$. This restriction is:

$$\varepsilon_{t+1}^r = \varepsilon_{t+1}^d + e_1' \cdot (I - \rho \cdot F_g)^{-1} \cdot \rho V_{t+1}^g - e_1' \cdot (I - \rho \cdot F_\mu)^{-1} \cdot \rho V_{t+1}^\mu. \tag{7.12}$$

See Ma and Wohar (2012) for a detailed derivation and also Cochrane (2008b) for a discussion.

Therefore, it is sufficient to estimate the variance matrix of any arbitrary set of three shocks and then the restriction (7.12) may be invoked to derive the variance matrix of all four shocks. Likewise, due to the identity (7.1) one only needs to choose two out of three observed variables $\Delta d_{t+1}, pd_{t+1}, r_{t+1}$ to set up the state-space model for estimation, and then the last variable may be backed out from the identity.[7]

[7] Ma and Wohar (2012) extensively study the robustness of the decomposition results by employing different variables and model specifications.

In Binsbergen and Koijen's (2010) estimation, they select $[\Delta d_{t+1}, pd_{t+1}]'$ and set lag $p = 1$ for both expectation dynamics. The state-space representation of the model is presented below:

Measurement equation:

$$\begin{bmatrix} \Delta d_{t+1} \\ pd_{t+1} \end{bmatrix} = \begin{bmatrix} a_g \\ H \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 & 1 \\ G_2 & 0 & -G_1 & 0 \end{bmatrix} \begin{bmatrix} \bar{g}_{t+1} \\ \bar{g}_t \\ \bar{\mu}_{t+1} \\ \varepsilon^d_{t+1} \end{bmatrix}. \tag{7.13}$$

Transition equation:

$$\begin{bmatrix} \bar{g}_{t+1} \\ \bar{g}_t \\ \bar{\mu}_{t+1} \\ \varepsilon^d_{t+1} \end{bmatrix} = \begin{bmatrix} \phi_g & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & \phi_\mu & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \bar{g}_t \\ \bar{g}_{t-1} \\ \bar{\mu}_t \\ \varepsilon^d_t \end{bmatrix} + \begin{bmatrix} \varepsilon^g_{t+1} \\ 0 \\ \varepsilon^\mu_{t+1} \\ \varepsilon^d_{t+1} \end{bmatrix}, \tag{7.14}$$

where $H = \frac{\kappa}{1-\rho} + \frac{a_g - a_\mu}{1-\rho}, G_1 = \frac{1}{1-\rho\phi_\mu}, G_2 = \frac{1}{1-\rho\phi_g}$, and $Var\left(\begin{bmatrix} \varepsilon^d_{t+1} \\ \varepsilon^\mu_{t+1} \\ \varepsilon^g_{t+1} \end{bmatrix}\right) = \begin{bmatrix} \sigma_d^2 & \times & \times \\ \sigma_{d\mu} & \sigma_\mu^2 & \times \\ \sigma_{dg} & \sigma_{\mu g} & \sigma_g^2 \end{bmatrix}$.

It has been pointed out in the literature that not all correlations are necessarily identifiable in this type of state-space models. Morley, Nelson and Zivot (2003) work on identification conditions for a particular type of univariate models for decomposing the output into trend and cycle and find that rich dynamics (long lags p) can help identify the correlations. Cochrane (2008b) and Rytchkov (2008) study the identification issues in this state-space model of decomposing the stock price and find that exactly one correlation is not identifiable for the case $p = 1$. As a result we need to impose one restrict in order to estimate the model. To facilitate comparisons we follow Binsbergen and Koijen (2010) and restrict $\sigma_{dg} = 0$.

We use the annual log dividends growth, log price-dividend ratio, and the log returns from 1947 to 2009 in Cochrane's (2011) chapter to estimate the above state-space model. The log-likelihood function of the model is written out via Kalman filter (see Kim and Nelson (1999)) and is maximized over the admissible parameter space using various starting values. The estimation results are laid out in Table 7.2. Our estimation results turn out to be similar to what Binsbergen and Koijen (2010) obtained. In particular, the persistent estimate for the expected returns is very high. Besides its standard error is very small, leading to an extremely tight 95% confidence interval for it: [0.8360, 1.0079]. At the same time, the expected dividend growth is far less persistent and its point estimate is even negative.

It is important to realize that the size of the persistence estimate in this decomposition framework has profound implication to the decomposition result. To illustrate this, write out (7.11) for the case $p = 1$ (dropping constants):

$$pd_t = \frac{g_t}{1-\rho\phi_g} - \frac{\mu_t}{1-\rho\phi_\mu}. \tag{7.15}$$

Table 7.2 State-space estimation results

Parameters	Estimates	Standard errors
a_g	0.0542	0.0127
ϕ_g	−0.4049	0.1660
a_μ	0.0806	0.0162
ϕ_μ	0.9219	0.0438
σ_d	0.0783	0.0152
σ_μ	0.0156	0.0071
σ_g	0.0868	0.0136
$\rho_{d\mu}$	−0.4285	0.0962
$\rho_{\mu g}$	0.9035	0.0456
Log-likelihood value	88.2314	
Implied parameters estimates		
σ_r	0.1540	0.0129
ρ_{dr}	0.9001	0.0451
$\rho_{\mu r}$	−0.7786	0.0468
ρ_{gr}	−0.4346	0.0934
Model constants		
κ	0.1401	–
ρ	0.9685	–
ZILC indication		
σ_μ/σ_r	0.1012	–
σ_g/σ_d	1.1092	–

Note: Data is annual from 1947 to 2009. The model is estimated by imposing the restriction $\rho_{dg} = 0$.

If we take the variance of both sides and ignore the correlation, we obtain (ignoring the correlation):

$$Var(pd_t) = \frac{1}{(1-\rho\phi_g)^2} \cdot Var(g_t) + \frac{1}{(1-\rho\phi_\mu)^2} \cdot Var(\mu_t). \tag{7.16}$$

Therefore, price-dividend variation consists of two contributions: $\frac{1}{(1-\rho\phi_g)^2} \cdot Var(g_t)$ and $\frac{1}{(1-\rho\phi_\mu)^2} \cdot Var(\mu_t)$. Each contribution percentage depends on both the variance of two factors $Var(g_t)$ and $Var(\mu_t)$, and the loading parameters $\frac{1}{(1-\rho\phi_g)^2}$ and $\frac{1}{(1-\rho\phi_\mu)^2}$. But both the variance of latent factors and the loadings are increasing and convex functions in the persistence parameters ϕ_g and ϕ_μ. Consequently, since the expected returns are estimated to be much more persistent than the expected dividend growth the estimation would tend to attribute most of the price-dividend variation to the expected return. Table 7.3 presents the variance decomposition of the price-dividend variation in this case. The variation of expected return indeed dominates that of the expected dividend growth in contributing to the price-dividend variation.

Figure 7.3 compares the expected dividend growth with the realized one. There is essentially very little difference between these two, which indicates that a large part of the realized dividend growth is well expected and there is little surprise. This

Table 7.3 State-space variance decomposition

Variance decomposition of price-dividend ratio (%)	
Contribution of expected return μ_t	76.24%
Contribution of expected dividend growth g_t	2.51%
Covariance contribution	21.25%

Note: Data is annual from 1947 to 2009; the covariance contribution is constructed as 100% minus the first two contributions.

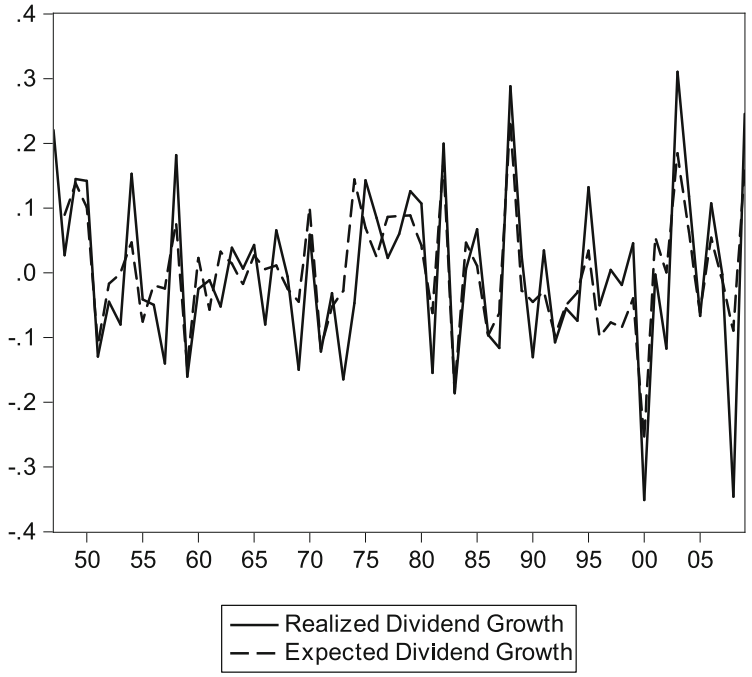

Fig. 7.3 Realized dividend growth and expected dividend growth. *Note*: Data is annual from 1947 to 2009.

result is primarily driven by the estimated signal-to-noise ratio $\sigma_g/\sigma_d = 1.1092$ reported at the bottom of Table 7.2.[8]

Figure 7.4 compares the expected return with the realized one and it portrays a very different picture from dividend growth. The expected return is very persistent and appears much smoother than the realized one. Most realized returns variation does not seem to be explained by the expected one. This finding is driven primarily by the small signal-to-noise ratio $\sigma_\mu/\sigma_r = 0.1012$ as reported at the bottom of Table 7.2.

To further illustrate this point we regress the realized returns on the filtered expected return in the following regression:

[8] Please refer to (7.3) and (7.5) for the definition of the expectation shock and news shock.

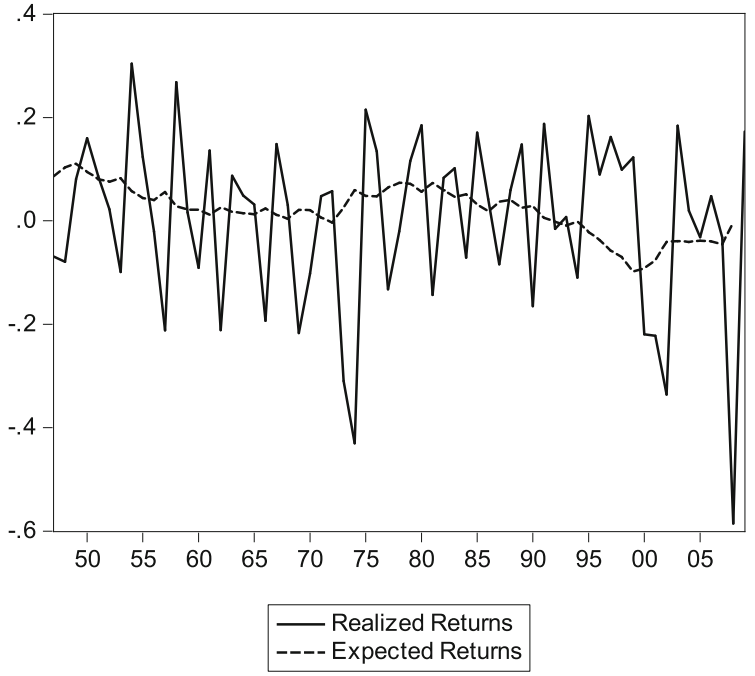

Fig. 7.4 Realized return and expected return. *Note*: Data is annual from 1947 to 2009.

$$r_{t+1} = a_1 + b_1 \cdot \hat{\mu}_t + v_{t+1}. \tag{7.17}$$

The expected return is extracted using all the past information of dividends growth and price-dividend in the state-space framework. Therefore this regression is able to tell us how much variation of returns are predictable using both past returns and price-dividend ratios. In this regression the adjusted R^2 is only about 10%, which is about the same size as one would get by regressing the realized returns on the one-period lagged price-dividend ratio:

$$r_{t+1} = a_2 + b_2 \cdot pd_t + \eta_{t+1}. \tag{7.18}$$

The adjusted R^2 for this regression is about 8.9%. Such poor predictability in the returns equation is, however, entirely consistent with what we highlight in Sect. 7.2 with respect to the VAR decomposition. To summarize, independent of the models chosen for variance decomposition, the evidence of return predictability appears small, and this turns out to have a profound impact on the statistical significance of the variance decomposition.

The state-space approach essentially filters out the expected return (signal) from the realized one (signal plus noise). The signal-to-noise ratio turns out to be very small for the expected return ($\sigma_\mu / \sigma_r = 0.1012$), which indicates the signal is far too small compared with the noise. Intuitively, when this happens we can only know

very little about the dynamics of the latent factors and there should be a great deal of uncertainty of the parameter estimates, in particular, the persistence parameter ϕ_μ. However, its standard error appears very small and indicates a counter-intuitive conclusion that the estimate is very accurate. We show next that the small signal-to-noise ratio implies a weak identified model and leads to spurious inference that needs to be corrected.

7.4 The Weak Identification and the Corrected Inference

Ma and Nelson (2010) find that state-space models in general are subject to the Zero-Information-Limit-Condition (ZILC) of Nelson and Startz (2007) (NS hereafter) when the signal is small relative to noise. Specifically, when the signal-to-noise ratio is small, the state-space model becomes weakly identified and the standard error of the persistence parameter would appear much smaller than it actually is, resulting in too many rejections of the null when using the standard t-test. To prove that ZILC holds, we relate the state-space model to its ARMA representation. Consider the return process:

$$r_{t+1} = a_\mu + \bar{\mu}_t + \varepsilon^r_{t+1}, \tag{7.19}$$

$$\bar{\mu}_t = \phi_\mu \cdot \bar{\mu}_{t-1} + \varepsilon^\mu_t. \tag{7.20}$$

Plugging (7.20) into (7.19), we can obtain:

$$\bar{r}_{t+1} = \phi_\mu \cdot \bar{r}_t + \varepsilon^r_{t+1} - \phi_\mu \cdot \varepsilon^r_t + \varepsilon^\mu_t, \tag{7.21}$$

where $\bar{r}$ is the demeaned return. By Granger and Newbold's Theorem (1986), (7.21) implies an ARMA(1,1) representation for the realized returns:

$$(1 - \phi_\mu L) \cdot \bar{r}_{t+1} = (1 - \theta L) \cdot u_{t+1}. \tag{7.22}$$

The mappings between the parameters of (7.21) and (7.22) can be solved explicitly by matching the second moments of the RHS of the two equations (assuming orthogonal shocks for illustration purpose):

$$\gamma_0 = (1 + \phi_\gamma^2)\sigma_\mu^2 = (1 + \theta^2)\sigma_u^2, \tag{7.23}$$

$$\gamma_1 = \phi_\mu \sigma_\gamma^2 = \theta \sigma_u^2, \tag{7.24}$$

where γ_0, γ_1 are the variance and first-order covariance of the RHS of (7.21) and (7.22). Solve the above two equations for the moving average parameter θ of ARMA(1,1) by assuming invertibility:

$$\theta = \begin{cases} \frac{[1+\phi_\mu^2+S^2]-\sqrt{((1-\phi_\mu^2)^2+S^4+2(1+\phi_\mu^2)S^2)}}{2\phi_\mu} & \text{for } \phi_\mu \neq 0 \\ 0 & \text{for } \phi_\mu = 0 \end{cases}, \tag{7.25}$$

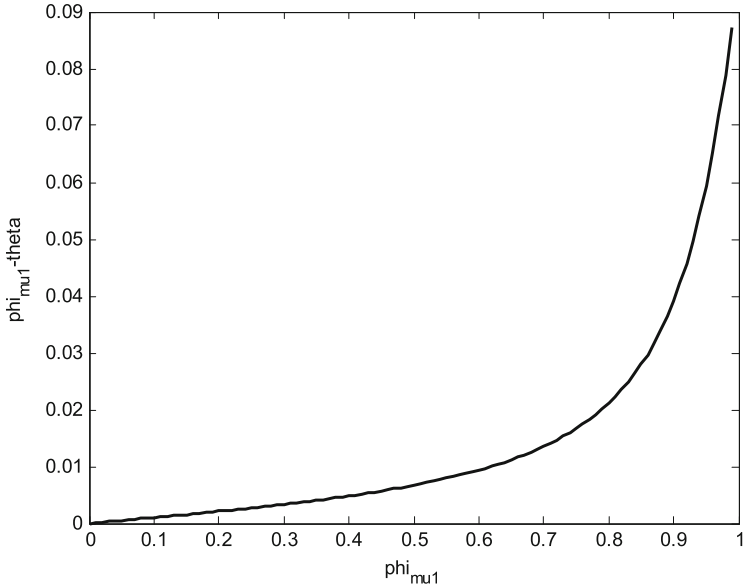

Fig. 7.5 Near root cancellation of the implied ARMA. *Note*: The horizontal axis gives a sequence of values of ϕ_μ between 0 and 1; the vertical axis gives the resulting value of $\phi_\mu - \theta$ from (7.26).

where $S = \sigma_\mu / \sigma_r$ is the signal-to-noise ratio. It is straightforward to prove that as this ratio approaches zero, ϕ_μ converges to θ, the well-known root cancellation issue of the ARMA:

$$\lim_{S \to 0} (\phi_\mu - \theta) = 0. \tag{7.26}$$

Figure 7.5 illustrates numerically the near root cancellation for various possible values of ϕ_μ, assuming S is equal to the estimated value 0.1012. Note that especially for small values of ϕ_μ, that is, low persistence, the difference is extremely small, indicating that the model is more weakly identified when the persistence is low.

NS shows that the ARMA of near root cancellation belongs to the class of models in which ZILC holds. Specifically, when $(\phi_\mu - \theta)$ is small (relative to sample variation), the information of either $\hat{\phi}_\mu$ or $\hat{\theta}$ will be severely underestimated based on conventional inference method, leading to a large size distortion of the standard t-test. Furthermore, under this circumstance, the persistence parameter such as ϕ_μ tends to have an upward bias, that is, the well-known pile-up issue in the ARMA. In the context of state-space model, a small signal-to-noise ratio would imply a weakly identified model that tends to give upward biased persistent estimates along with severely underestimated standard errors. As a result, the standard test would tend to reject the null of low or zero persistence too often.

To deal with this issue, Ma and Nelson (2010) propose a reduced-form test based on a linear approximation to the exact test of Fieller (1954) for a ratio of regression coefficients. The test is also an LM test in the spirit of Breusch and Pagan (1980)

since the test is conducted based on the null. They show that this test has a nearly correct size in finite samples when the standard t-test is severely oversized in the presence of weak identification.

To compute the reduced-form test for ϕ_μ, write out the reduced-form VARMA for the dividends growth and price-dividend ratio as implied by (7.13) and (7.14):

$$\begin{bmatrix} (1-\phi_g L) & 0 \\ 0 & (1-\phi_g L)(1-\phi_\mu L) \end{bmatrix} \begin{bmatrix} \Delta d_{t+1} \\ pd_{t+1} \end{bmatrix} = \begin{bmatrix} (1-\theta_1 L) & 0 \\ 0 & (1-\theta_2 L) \end{bmatrix} \begin{bmatrix} u_{1t+1} \\ u_{2t+1} \end{bmatrix}, \tag{7.27}$$

where the shocks $[u_{1t+1}, u_{2t+1}]'$ may be correlated.[9]

To construct the reduced-form test for ϕ_μ, we first impose the null $\phi_\mu = \phi_{\mu,0}$ and estimate the restricted VARMA. In the second step, we focus on the second equation of (7.27) since ϕ_μ only shows up there and compute the t-statistic for the null $\lambda = 0$ from the following classical linear regression:

$$(1-\tilde{\phi}_{g,1} L) pd_{t+1} = \tau \cdot \sum_{i=1}^{\infty} \phi_{\mu,10}^{i-1} \tilde{u}_{2t+1-i} + \lambda \cdot \sum_{i=2}^{\infty} (i-1) \cdot \phi_{\mu,10}^{i-2} \tilde{u}_{2t+1-i} + \tilde{u}_{2t+1} \tag{7.28}$$

where (7.28) is the first-order Taylor approximation of the second equation of (7.27) around the null $\phi_\mu = \phi_{\mu,0}, \tau = \phi_\mu - \theta_2$, and $\lambda = \tau(\phi_\mu - \phi_{\mu,0})$; $\tilde{\phi}_g$ and $\tilde{u}_{2t+1}$ are the restricted estimates from the first step. The reduced-form test for $\phi_\mu = \phi_{\mu,0}$ is the t-statistic for the null $\lambda = 0$, since intuitively the second term on the RHS of (7.28) should not be significant if the null is true and the first term is enough to capture all serial correlations.

In order to present the whole confidence interval, we calculate the reduced-form test statistics corresponding to each possible null of ϕ_μ, and the 95% confidence interval consists of these null values that produce test statistics not exceeding the 5% critical value. Figure 7.6 plots the reduced-form t-statistic for the data used in this chapter. The test result indicates that the confidence interval for the expected return persistence covers essentially all admissible parameter regions, which surely is much wider than that of the standard inference.

Interestingly, the reduced-form test cannot even reject the zero persistence, that is, $H_0 : \phi_\mu = 0$. Recall that the contribution of the expected returns to price-dividend variation hinges upon the persistence parameter of the expected returns process (see (7.16)). If the null hypothesis $H_0 : \phi_\mu = 0$ cannot be rejected, then we also cannot reject the null hypothesis of zero contribution of the expected returns to price-dividend variation. This finding is, however, remarkably consistent with the primary conclusion based on bootstrap procedures in Sect. 7.2, and thus raises a doubt about the statistical significance of the contribution of the expected returns to the stock price variation.

Finally, the issue of weak identification comes down to the degree of persistence in the particular factor (i.e., expected dividend growth or expected returns). The persistent component is what contains the predominant amount of information

[9] The off-diagonal elements in the matrix of moving average parameters are restricted to be zeros. This restriction does not affect our main results but facilitates the estimation.

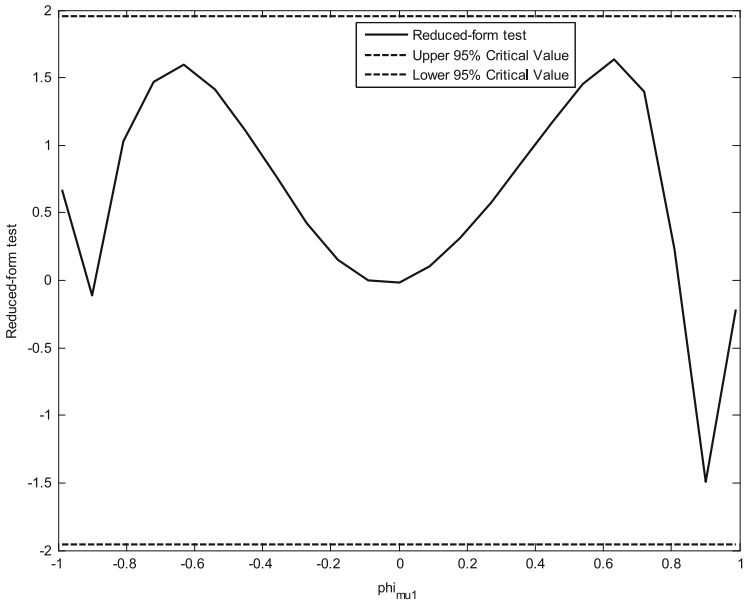

Fig. 7.6 The confidence interval for expected return persistence based on the reduced-form test. *Note*: Data is annual from 1947 to 2009; numbers on the x-axis are the null hypotheses for ϕ_μ. Numbers on the y-axis are reduced-form test statistics.

about movements in the price-dividend ratio. If expected returns are persistent, then the signal-to-noise ratio will be small and most of the variation of the price-dividend ratio will be explained by movements in excess returns. Similarly for expected dividend growth. If the persistence in a factor is low, however, then the signal-to-noise ratio will be large and the contribution of that factor to movements in the price-dividend ratio will be small. The issue of weak identification and the ZILC condition makes it very difficult to determine which factor dominates and explains movements in the price-dividend ratio.

7.5 Conclusion

Finance theory tells us that the price of an asset is equal to the discounted expected future cash flows that the asset generates. It follows that there are primarily two factors that can influence prices; expectations regarding discount rates (expected returns) and expectations regarding future cash flows. Employing the variance decomposition return approach, virtually all of the empirical work in this area has concluded that it is expectations of future returns rather than future dividend growth that are responsible for most of the fluctuations in stock prices (or the price-dividend ratio).

Employing an alternative approach to the variance decomposition method, Binsbergen and Koijen (2010) employ an unobserved component (state-space)

model using US annual data within a present value model to estimate expected return and expected dividend growth rates of the aggregate US stock market. Their approach aggregates information contained in the price-dividend ratio and dividend growth rates to predict expected returns and dividend growth rates. They treat conditional expected returns and expected dividend growth as latent variables that follow an exogenously specified time series process. They find that both expected dividend growth rates and expected returns are time varying and persistent but that expected returns are more persistent than expected dividend growth rates. They find that expected returns contribute more to fluctuations in the price-dividend ratio than do dividend growth rates. Balke and Wohar (2002) also estimate unobserved expectations market fundamentals employing a more general state-space model than Binsbergen and Koijen (2010). They show that stock price decompositions are very sensitive to assumptions about which market fundamental has a permanent component.

Our chapter shows that the existing literature has focused on point estimates with little detailed attention given to issues of inference. When issues of inference are considered we find that there is little evidence to support the notion that either expected returns or expected dividend growth contributes to movements in the price-dividend ratio. In an effort to explore this finding in detail, we employ a state-space modeling framework. We find that within this framework, it is the existence of weak identification combined with a low signal-to-noise ratio that leads to the conclusion that there is too much uncertainty to make any claims about the relative contributions of expected returns and expected dividend growth to movements in the price-dividend ratio. We propose a procedure that could potentially correct for the inference problem and offer more reliable results. The corrected inference indicates that the large contribution of the expected returns to fluctuations in the price-dividend ratio found in previous studies has no statistical significance. Our finding that the weak identification plays an important role in the stock price decomposition naturally calls for further investigations using more information. The additional information can come in the form of either model restrictions or more data, such as disaggregated equity returns or corporate earnings.

References

1. Ammer, J. and J. Mei, 1996, “Measuring International Economic Linkages With Stock Market Data”, *Journal of Finance*, 51, 1743–1763.
2. Balke, N. S. and M. E. Wohar, 2002, “Low Frequency Movements in Stock Prices: A State-Space Decomposition”, *Review of Economics and Statistics*, 84, 649–667.
3. Binsbergen, J. H. van, and R. S.J. Koijen, 2010, “Predictive regressions: A present-value approach”, *Journal of Finance*, 65, 1439–1471.
4. Breusch T. S. and A. R. Pagan, 1980, “The Lagrange Multiplier Test and its Application to Model Specification in Econometrics”, *Review of Economic Studies*, XLVII, 239–253.
5. Callen, J.L. and D. Segal, 2004, “Do Accruals Drive Stock Returns? A Variance Decomposition Analysis”, *Journal of Accounting Research*, 42, 527–560.
6. Campbell J., 1991, “A variance decomposition for stock returns”, *The Economic Journal*, 101, 157–179.

7. Campbell J. and J. Ammer, 1993, "What moves the stock and bond markets? A variance decomposition for long-term asset returns", *Journal of Finance*, 48, 3–37.
8. Campbell J. and Y. Hamao, 1992, "Predictable Returns in the United States and Japan: A Study of Long-term Capital Market Integration", *Journal of Finance*, 47, 43–70.
9. Campbell J. and R. Shiller, 1988a, "The dividend-price ratio and expectations of future dividends and discount factors", *Review of Financial Studies*, 1, 195–228.
10. Campbell J. and R. Shiller, 1988b, "Stock prices, earnings, and expected dividends", *Journal of Finance*, 43, 661–676.
11. Campbell J. and R. Shiller, 1998, "Valuation ratios and the long-run stock market outlook: an update", Unpublished working paper. Harvard University.
12. Campbell J. and T. Vuolteenaho, 2004, "Bad beta, good beta", *American Economic Review*, 94, 1249–75.
13. Cochrane J., 1992, "Explaining the variance of price-dividend ratios", *Review of Financial Studies*, 5, 243–280.
14. Cochrane, J., 2001, Asset Pricing, Princeton, NJ; Princeton University Press.
15. Cochrane J., 2008a, "The dog that did not bark: a defense of return predictability", *Review of Financial Studies*, 21, 1533–1575.
16. Cochrane J., 2008b, "State-space vs. VAR models for stock returns", unpublished working paper, University of Chicago.
17. Cochrane John, 2011, "Discount Rates", *Journal of Finance*,66,1047–1108.
18. Cohen, R.B., C. Polk and T. Vuolteennaho, 2003, "The Value Spread", *Journal of Finance*, 58, 609–641.
19. Davidson, R. and J.G. MacKinnon, 2004, Econometric Theory and Methods, Oxford University Press, 2004.
20. Fieller, E. C., 1954, "Some Problems in Interval Estimation", *Journal of the Royal Statistical Society. Series B (Methodological)*, 16, 175–85.
21. Granger, C. W.J. and P. Newbold, 1986, Forecasting Economic Time Series, Second Edition, Orlando: Academic Press.
22. Kim, C.-J. and C.R. Nelson, 1999, State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications, The MIT Press.
23. Larrain, B., and M. Yogo, 2008, "Does Firm Value Move Too Much to Be Justified by Subsequent Changes in Cash Flow?" *Journal of Financial Economics*, 87, 200–226.
24. Ma, J. and C. R. Nelson, 2010, "Valid Inference for a Class of Models Where Standard Inference Performs Poorly; Including Nonlinear Regression, ARMA, GARCH, and Unobserved Components", working paper, Economic Series 256, Institute for Advanced Studies, Vienna.
25. Ma, J. and M. E. Wohar, 2012, "Expected Returns and Expected Dividend Growth: Time to Rethink an Established Empirical Literature", working paper, University of Alabama.
26. Morley, J., C. R. Nelson and E. Zivot, 2003, "Why Are Unobserved Component and Beveridge-Nelson Trend-Cycle Decompositions of GDP Are So Different?" *Review of Economics and Statistics*, 85, 235–243.
27. Nelson, C. R. and R. Startz, 2007, "The Zero-Information-Limit Condition and Spurious Inference in Weakly Identified Models", *Journal of Econometrics*, 138, 47–62.
28. Rytchkov, O., 2008, "Filtering Out Expected Dividends and Expected Returns", working paper, Fox School of Business and Management.
29. Shiller, R. and A. Beltratti, 1992, "Stock Prices and Bond Yields", *Journal of Monetary Economics*, 30, 25–46.
30. Vuolteenaho T., 2002, "What Drives Firm-Level Stock Returns?" *Journal of Finance*, 57, 233–275.

Part III
Hidden Markov Models, Regime-Switching, and Mathematical Finance

Chapter 8
A HMM Intensity-Based Credit Risk Model and Filtering

Robert J. Elliott and Tak Kuen Siu

8.1 Introduction

Modeling default risk is an important topic in financial risk management. There are two major approaches to modeling default risk, namely, the structural firm value approach initiated by Black and Scholes (1973) [2] and Merton (1974) [29], and the reduced-form intensity-based approach introduced by Jarrow and Turnbull (1995) [24] and Madan and Unal (1998) [28]. The key idea of the structural firm value approach is to model explicitly the relationship between a firm's asset value and the default of the firm. More specifically, the asset value of the firm is described by a geometric Brownian motion and default of the firm is triggered by the event that the asset value falls below a default barrier level. This means that defaults are endogenous events in the structural firm value model. The reduced-form intensity-based approach is based on the premise that defaults are exogenous events and models arrivals of defaults by Poisson point processes. In the structural approach, the default time is a stopping time with respect to the asset's filtration; in the reduced-form approach, the default time is a stopping time with respect to a larger filtration (see Cooper and Martin (1996) [7]). The difference between these two types of default times has important implications for the evaluation of default probabilities of firms and the valuation of defaultable securities. For details, interested readers may refer to Elliott, Jeanblanc and Yor (2000) [18].

Modeling dependent defaults of constituents in a credit portfolio is a central theme in default risk modeling. One approach to modeling dependent defaults is

R.J. Elliott (✉)
Haskayne School of Business, University of Calgary, Calgary, AB, Canada
School of Mathematical Sciences, University of Adelaide, Adelaide, SA 5005, Australia
Centre for Applied Financial Studies, University of South Australia, Adelaide, SA, Australia
e-mail: relliott@ucalgary.ca

T.K. Siu
Cass Business School, City University London, 106 Bunhill Row, London EC1Y 8TZ, UK
e-mail: Ken.Siu.1@city.ac.uk

Y. Zeng and S. Wu (eds.), *State-Space Models: Applications in Economics and Finance*, Statistics and Econometrics for Finance 1, DOI 10.1007/978-1-4614-7789-1_8,

based on the reduced-form intensity-based model. There are two major types of reduced-form intensity-based models for dependent defaults, namely, a bottom-up model and a top-down model. In a bottom-up model, one focuses on modeling default intensities of constituents, and the default intensity of the credit portfolio is given by an aggregation of the default intensities of the constituents. Some works on the bottom-up approach to portfolio credit risk include Kusuoka (1999) [26], Duffie and Garleanu (2001) [12], Jarrow and Yu (2001) [25], Schönbucher and Schubert (2001) [31], Giesecke and Goldberg (2004) [21], Schönbucher (2004) [32], Duffie, Saita and Wang (2006) [13], Giesecke and Weber (2006) [23], Mortensen (2006) [30], Das, Duffe, Kapadia and Saita (2007) [9], Eckner (2007) [15], Feldhütter (2007) [20], Yu (2007) [33], Duffie, Eckner, Horel and Saita (2009) [14], and others. These works differ in the way the default intensities of constituents were specified. In particular, Giesecke and Goldberg (2004) [21], Schönbucher (2004) [32], Duffie, Eckner, Horel and Saita (2009) [14] considered the situation where the default intensities of constituents depend on an unobservable factor, namely, a frailty factor. In a top-down model, the default intensity is specified at the portfolio level without reference to the identities of the constituents. Then a random thinning procedure is adopted to recover default intensities of constituents. Some works on top-down models include Davis and Lo (2001) [10], Giesecke and Goldberg (2005) [22], Brigo, Pallavicini and Torresetti (2006) [3], Ding, Giesecke and Tomecek (2006) [11], Arnsdorf and Halperin (2007) [1], Longstaff and Rajan (2007) [27], Cont and Minca (2008) [6], amongst others.

In this article, we discuss an intensity-based model for portfolio credit risk using a set of hidden Markov-modulated single jump processes, which is a hidden Markov model (HMM). The probability laws of these single jump processes are specified by compensators, or dual predictable projections, modulated by a continuous-time, finite-state, hidden Markov chain. The states of the chain are interpreted as different levels of a common hidden dynamic frailty factor. Firms are exposed to this common source of hidden risk factor. The model considered here is a bottom-up model, where defaults are modelled at the level of individual reference entities. An important feature of the model is that the information structure is fine enough to distinguish the identity of each defaulter, so each of the single jump processes for defaults is observed. Based on observations about single jump processes for defaults, we obtain filtering equations for the dynamic frailty factor. We also derive a (robust) filter-based EM algorithm for the online recursive estimates of unknown model parameters. Finally, a joint default probability of individual reference entities in a credit portfolio is given. This article is for an expository purpose. It is a shorter version of a paper by Elliott and Siu (2011) [19]. Full proofs of the results in this article will be published in the longer paper [19]. Numerical results and applications to Credit Value at Risk will also be published in [19].

The rest of the chapter is organized as follows. The next section discusses the modeling framework. In Sect. 8.3 we first discuss briefly a reference probability approach for filtering. Then we give the filtering equations for some hidden quantities. A filter-based EM algorithm for the online recursive estimates of the model parameters is presented in Sect. 8.4. We present the variance dynamics in Sect. 8.5.

In Sect. 8.6 we give the joint default probability of the credit portfolio. The final section concludes this article.

8.2 A HMM Frailty-Based Default Model

Consider a complete probability space $(\Omega, \mathscr{F}, P)$, where P is a real-world probability measure. The probability space is supposed to be rich enough to incorporate all sources of uncertainty in our modeling framework. Let $\mathscr{T}$ be the extended time interval $[0, \infty)$ on which economic activities take place.

Let $\mathbf{X} := \{\mathbf{X}(t) | t \in \mathscr{T}\}$ be a continuous-time hidden Markov chain with finite-states on $(\Omega, \mathscr{F}, P)$ with state space $\mathscr{E} := \{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$, where $\mathbf{e}_i$ is a standard unit vector in $\Re^N$ and the jth component of $\mathbf{e}_i$ is the Kronecker delta δ_{ij}, for each $i, j = 1, 2, \cdots, n$. Indeed, $\mathscr{E}$ is the standard basis of the Euclidean space $\Re^n$ and is called the canonical state space of the chain $\mathbf{X}$. The states of the chain $\mathbf{X}$ represent different levels of a dynamic frailty factor.

Suppose the generator, or Q-matrix, of the chain $\mathbf{X}$ is the $(n \times n)$-matrix $\mathbf{A}$. The probability laws of the chain $\mathbf{X}$ are specified by the generator $\mathbf{A}$. Since the generator $\mathbf{A}$ does not depend on time, the chain $\mathbf{X}$ is time-homogeneous.

Let $\mathbb{F}^{\mathbf{X}} := \{\mathscr{F}^{\mathbf{X}}(t) | t \in \mathscr{T}\}$ be the P-completion of the natural filtration generated by the chain $\mathbf{X}$. Note that $\mathbb{F}^{\mathbf{X}}$ is right-continuous. It was shown in Elliott et al. (1995) [17] that the chain $\mathbf{X}$ has the following semimartingale dynamics:

$$\mathbf{X}(t) = \mathbf{X}(0) + \int_0^t \mathbf{A}\mathbf{X}(u-)du + \mathbf{M}(t) \ .$$

Here $\{\mathbf{M}(t) | t \in \mathscr{T}\}$ is an $\Re^N$-valued, $(\mathbb{F}^{\mathbf{X}}, P)$-martingale.

We now describe the single jump processes and their associated martingales. Suppose that there are K constituents in a credit portfolio and that after a constituent defaults, it will stay in the default state forever. For each constituent k ($k = 1, 2, \cdots, K$), the default time τ_k of the kth constituent is a totally inaccessible (or an unpredictable), stopping time with respect to some observable filtration to be defined later in this section. We suppose that for each $k = 1, 2, \cdots, K$, τ_k is a random variable on $(\Omega, \mathscr{F}, P)$ taking value in $\mathscr{T}$. For each $k = 1, 2, \cdots, K$, let $N^k := \{N^k(t) | t \in \mathscr{T}\}$ be a right-continuous, non-decreasing process given by:

$$N^k(t) := I_{\{\tau_k \le t\}} \ , \quad t \in \mathscr{T} \ ,$$

where I_A is the indicator function of the event A.

For each $k = 1, 2, \cdots, K$, N^k is the default process of the kth constituent. This process takes the value "zero" before default and "one" afterwards. Here we suppose that P-almost surely there are no common jumps between the chain $\mathbf{X}$ and the default processes N^k, $k = 1, 2, \cdots, K$.

One may consider the situation that structural changes in the state of the economy and defaults of the constituents may happen at the same time. However, we believe that there should be a lead–lag effect between structural changes in the state of the

economy and defaults of the constituents. In other words, defaults of constituents can only happen after structural changes in the state of the economy, and vice versa.

Write, for each $k = 1,2,\cdots,K$, $\mathbb{F}^k := \{\mathscr{F}^k(t)|t \in \mathscr{T}\}$ for the P-complete, right-continuous, natural filtration generated by the default process N^k, so

$$\mathscr{F}^k(t) := \sigma\{N^k(t)|t \in \mathscr{T}\} \vee \mathscr{N} ,$$

where $\mathscr{N}$ is the collection of all P-negligible sets in the σ-field $\mathscr{F}$.

Indeed, the σ-field $\mathscr{F}^k(t)$ is generated by the sets $\{\tau_k \leq u\}$ for each $u \in [0,t]$, and the atom $\{\tau_k > t\}$. Consequently, $\mathbb{F}^k$ is the smallest filtration satisfying the usual hypotheses such that τ_k is an $\mathbb{F}^k$-stopping time.

Define, for each $t \in \mathscr{T}$ and $k = 1,2,\cdots,K$,

$$\mathscr{G}^k(t) := \mathscr{F}^k(t) \vee \mathscr{F}^{\mathbf{X}}(t) .$$

That is, the minimal σ-field generated by both $\mathscr{F}^k(t)$ and $\mathscr{F}^{\mathbf{X}}(t)$. Write $\mathbb{G}^k$ for the right-continuous, complete filtration generated by $\{\mathscr{G}^k(t)|t \in \mathscr{T}\}$.

For each $k = 1,2,\cdots,K$, let $A^k := \{A^k(t)|t \in \mathscr{T}\}$ be the dual predictable projection of the default process N^k with respect to the filtration $\mathbb{G}^k$ under the measure P. Note that A^k is also called the $(\mathbb{G}^k,P)$-compensator of N^k so that the process $\bar{N}^k := \{\bar{N}^k(t)|t \in \mathscr{T}\}$ defined by:

$$\bar{N}^k(t) := N^k(t) - A^k(\tau^k \wedge t) , \quad t \in \mathscr{T} ,$$

is a $(\mathbb{G}^k,P)$-local martingale, where $\tau^k \wedge t := \min\{\tau^k,t\}$. Consequently, N^k is a $(\mathbb{G}^k,P)$-submartingale and has the following decomposition:

$$N^k(t) = \bar{N}^k(t) + A^k(\tau_k \wedge t) , \quad t \in \mathscr{T} .$$

We suppose that the compensator A^k is absolutely continuous with respect to the Lebesgue measure on $\mathscr{T}$. Then there is a nonnegative, $\mathbb{G}^k$-progressively measurable, intensity process $\lambda^k := \{\lambda^k(t)|t \in \mathscr{T}\}$ such that

$$A^k(t) = \int_0^t \lambda^k(u)du , \quad P\text{-a.s.} ,$$

so

$$N^k(t) - \int_0^{\tau^k \wedge t} \lambda^k(u)du , \quad t \in \mathscr{T} ,$$

is a $(\mathbb{G}^k,P)$-local martingale. It was noted in Elliott et al. (2000) [18] that the intensity process λ^k is not uniquely defined after τ^k and that there is no meaning for the "intensity" after τ^k though it is sometimes mentioned.

For each $t \in \mathscr{T}$ and $k = 1,2,\cdots,K$, the intensity $\lambda^k(t)$ is the conditional mean rate of default of the kth constituent given $\mathscr{G}^k(t)$ under the measure P. We consider here a particular form for an intensity process λ^k as follows:

$$\lambda^k(t) := \left\langle \lambda^k, \mathbf{X}(t) \right\rangle , \quad t \in \mathscr{T} .$$

Here $\lambda^k := (\lambda_1^k, \lambda_2^k, \cdots, \lambda_n^k)' \in \Re^n$ with $\lambda_i^k > 0$ for each $i = 1, 2, \cdots, n$. In other words, the intensity process λ^k is modulated by the chain $\mathbf{X}$. Since the chain $\mathbf{X}$ is not unobservable, neither is the intensity process λ^k. The scalar product $\langle \cdot, \cdot \rangle$ in $\Re^N$ selects the component in the vector λ^k of default intensities of the kth constituent in force depending on the current state of the economy described by the value of the chain $\mathbf{X}(t)$.

Under P the HMM intensity-based credit risk model has the following state-space form:

$$N^k(t) = \int_0^{\tau^k \wedge t} \left\langle \lambda^k, \mathbf{X}(u) \right\rangle du + \bar{N}^k(t) \,, \quad k = 1, 2, \cdots, K \,,$$
$$\mathbf{X}(t) = \mathbf{X}(0) + \int_0^t \mathbf{A}\mathbf{X}(u-) du + \mathbf{M}(t) \,.$$

Here the single jump processes N^k, $k = 1, 2, \cdots, K$, are the observation processes and the chain $\mathbf{X}$ is the state process.

8.3 Filtering Equations for the Hidden Dynamic Frailty Factor

In this section, we first discuss an approach based on a reference probability and a version of the Bayes' rule to derive filtering equations for the dynamic frailty factor. Then we present a Zakai stochastic differential equation for the unnormalized filter of the frailty factor. Using a gauge transformation technique, we further simplify the filtering equation and give a linear ordinary differential equation for a (robust) filter of the frailty factor. Here robustness is in the sense of Clark (1978) [5] and refers to the Lipschitz continuity with respect to the observation processes in the Skorohod topology.

Suppose there is a reference probability measure $P^\dagger$ under which

1. the default processes N^k, $k = 1, 2, \cdots, K$, are independent and have an unit intensity. This means that the compensator of N^k is $\{t \wedge \tau^k | t \in \mathscr{T}\}$, for each $k = 1, 2, \cdots, K$;
2. the hidden Markov chain $\mathbf{X}$ has the generator $\mathbf{A}$;
3. the chain $\mathbf{X}$ and the processes N^k, $k = 1, 2, \cdots, K$, are independent.

Write

$$M^k(t) := N^k(t) - \tau^k \wedge t \,, \quad t \in \mathscr{T} \,,$$

so M^k is an $(\mathbb{F}^k, P^\dagger)$-martingale, and hence, a $(\mathbb{G}^k, P^\dagger)$-martingale.

Define, for each $k = 1, 2, \cdots, K$, the family of stochastic exponentials $Z^k := \{Z^k(t) | t \in \mathscr{T}\}$ by:

$$Z^k(t) := \exp\left(- \int_0^{\tau^k \wedge t} (\lambda^k(u) - 1) du + \int_0^t \ln \lambda^k(u) dN^k(u) \right) .$$

Write, for each $t \in \mathscr{T}$,

$$\mathscr{G}(t) := \bigvee_{k=1}^{K} \mathscr{G}^k(t) \ ,$$

and $\mathbb{G} := \{\mathscr{G}(t) | t \in \mathscr{T}\}$.

Consider now the following $\mathbb{G}$-adapted process $Z := \{Z(t) | t \in \mathscr{T}\}$ defined by:

$$Z(t) := Z^1(t)Z^2(t)\cdots Z^K(t) \ , \quad t \in \mathscr{T} \ .$$

Then

$$Z(t) = \exp\left(-\sum_{k=1}^{K} \int_0^{\tau^k \wedge t} (\lambda^k(u) - 1)du + \sum_{k=1}^{K} \int_0^t \ln \lambda^k(u) dN^k(u) \right) .$$

Applying Itô's differentiation rule to $Z(t)$ gives:

$$dZ(t) = \sum_{k=1}^{K} \int_0^t Z(u-)(\lambda^k(u) - 1)dM^k(u) \ ,$$

so Z is a $(\mathbb{G}, P^\dagger)$-(local)-martingale. We suppose here that Z is a positive, uniformly integrable, $(\mathbb{G}, P^\dagger)$-martingale.

By Remark 13.18 in Elliott (1982) [16],

$$Z(\infty) := \lim_{t \to \infty} Z(t) \ ,$$

exists $P^\dagger$-a.s., and for each $t \in \mathscr{T}$,

$$Z(t) = \mathrm{E}^\dagger[Z(\infty) | \mathscr{G}(t)] \ , \quad P\text{-a.s.} \ ,$$

where $\mathrm{E}^\dagger$ is expectation under $P^\dagger$.

Consequently,

$$\mathrm{E}^\dagger[Z(\infty)] = Z(0) = 1 \ .$$

Note that $Z(\infty) > 0$, $P^\dagger$-a.s. We can then re-construct the real-world probability measure P equivalent to $P^\dagger$ on $\mathscr{G}(\infty)$ by putting:

$$\left. \frac{dP}{dP^\dagger} \right|_{\mathscr{G}(\infty)} := Z(\infty) \ .$$

Recall that $\bar{N}^k := \{\bar{N}^k(t) | t \in \mathscr{T}\}$ is defined by:

$$\bar{N}^k(t) := N^k(t) - \int_0^{t \wedge \tau^k} \lambda^k(u)du \ , \quad t \in \mathscr{T} \ .$$

By a version of Theorem 13.19 in Elliott (1982) [16], it can be shown that under P, $\bar{N}^k$ is a local martingale with respect to the filtration $\mathbb{G}^k$, for each $k = 1, 2, \cdots, K$.

In other words, $\bar{N}^k$ has the intensity process λ^k under P prior to the default time τ^k.

Furthermore, since $\mathbf{X}$ and $N^k, k = 1, 2, \cdots, K$, are independent under $P^\dagger$, the probability law of the chain $\mathbf{X}$ remains unchanged after the measure change from $P^\dagger$ to P by the density process Z. Consequently, under P, the chain $\mathbf{X}$ has the generator $\mathbf{A}$.

Write, for each $t \in \mathscr{T}$,

$$\mathscr{F}(t) := \bigvee_{k=1}^{K} \mathscr{F}^k(t) ,$$

and $\mathbb{F} := \{\mathscr{F}(t) | t \in \mathscr{T}\}$, so $\mathbb{F}$ is the observed filtration generated by the default processes of the constituents. We wish to "fuse" together information about defaults of constituents to estimate the hidden dynamic frailty factor.

By a version of the Bayes' rule,

$$\begin{aligned} \mathrm{E}[\mathbf{X}(t)|\mathscr{F}(t)] &= \frac{\mathrm{E}^\dagger[Z(t)\mathbf{X}(t)|\mathscr{F}(t)]}{\mathrm{E}^\dagger[Z(t)|\mathscr{F}(t)]} \\ &= \frac{\sigma(\mathbf{X}(t))}{\sigma(1)} , \quad \text{say} . \end{aligned}$$

Here $\sigma(\mathbf{X}(t)) := \mathrm{E}^\dagger[Z(t)\mathbf{X}(t)|\mathscr{F}(t)]$, so $\sigma(1) = \mathrm{E}^\dagger[Z(t)|\mathscr{F}(t)]$. $\sigma(\mathbf{X}(t))$ is called the unnormalized filter of $\mathbf{X}$ given $\mathscr{F}(t)$.

Note that

$$\langle \mathbf{X}(t), \mathbf{1} \rangle = 1 ,$$

where $\mathbf{1} := (1, 1, \cdots, 1)' \in \Re^n$, so $\sigma(1) = \langle \sigma(\mathbf{X}(t)), \mathbf{1} \rangle$, and hence,

$$\mathrm{E}[\mathbf{X}(t)|\mathscr{F}(t)] = \frac{\sigma(\mathbf{X}(t))}{\langle \sigma(\mathbf{X}(t)), \mathbf{1} \rangle} .$$

Consequently, to determine the normalized filter $\mathrm{E}[\mathbf{X}(t)|\mathscr{F}(t)]$, it suffices to determine the unnormalized one $\sigma(\mathbf{X}(t))$.

Then the following theorem gives the filtering equation for the unnormalized filter $\sigma(\mathbf{X}(t))$.

Theorem 8.1. *Write*

$$\mathbf{diag}(\lambda^k - \mathbf{1}) := \mathbf{diag}(\lambda_1^k - 1, \lambda_2^k - 1, \cdots, \lambda_n^k - 1) ,$$

a diagonal matrix with diagonal elements being $(\lambda_1^k - 1, \lambda_2^k - 1, \cdots, \lambda_n^k - 1)$.

Then $\sigma(\mathbf{X})$ *satisfies the following stochastic differential equation:*

$$\begin{aligned} \sigma(\mathbf{X}(t)) = \sigma(\mathbf{X}(0)) &+ \int_0^t \mathbf{A}\sigma(\mathbf{X}(u))du + \sum_{k=1}^{K} \int_0^t \mathbf{diag}(\lambda^k - \mathbf{1})\sigma(\mathbf{X}(u-))dN^k(u) \\ &- \sum_{k=1}^{K} \int_0^{\tau^k \wedge t} \mathbf{diag}(\lambda^k - \mathbf{1})\sigma(\mathbf{X}(u))du . \end{aligned}$$

To simplify the filtering equation in Theorem 8.1, we consider a gauge transformation matrix $\Gamma(t)$ to be defined as follows:

For each $i = 1,2,\cdots,n$, we consider a scalar-valued process $\gamma_i := \{\gamma_i(t)|t \in \mathscr{T}\}$ defined by:

$$\gamma_i(t) := \exp\left[\sum_{k=1}^{K}(1-\lambda_i^k)(\tau^k \wedge t) + \sum_{k=1}^{K}\int_0^t \ln\lambda_i^k dN^k(u)\right], \quad t \in \mathscr{T}.$$

Write, for each $t \in \mathscr{T}$,

$$\Gamma(t) := \mathbf{diag}(\gamma_1(t), \gamma_2(t), \cdots, \gamma_n(t)),$$

where $\mathbf{diag}(\gamma_1(t), \gamma_2(t), \cdots, \gamma_n(t))$ is a diagonal matrix with diagonal elements being the vector $(\gamma_1(t), \gamma_2(t), \cdots, \gamma_n(t))$.

Write, for each $t \in \mathscr{T}$, $\Gamma^{-1}(t)$ for the inverse of $\Gamma(t)$. The existence of $\Gamma^{-1}(t)$ is ensured by the positivity of $\gamma_i(t)$, for each $i = 1,2,\cdots,n$. Then we have the following lemma.

Lemma 8.1. *For each $k = 1,2,\cdots,K$, define a diagonal matrix by:*

$$\mathbf{diag}\left(\frac{\mathbf{1}-\lambda^k}{\lambda^k}\right) := \mathbf{diag}\left(\frac{1-\lambda_1^k}{\lambda_1^k}, \frac{1-\lambda_2^k}{\lambda_2^k}, \cdots, \frac{1-\lambda_n^k}{\lambda_n^k}\right).$$

Then

$$d\Gamma^{-1}(t) = \sum_{k=1}^{K} I_{\{t\le\tau^k\}}\mathbf{diag}(\lambda^k - \mathbf{1})\Gamma^{-1}(t)dt + \sum_{k=1}^{K}\mathbf{diag}\left(\frac{\mathbf{1}-\lambda^k}{\lambda^k}\right)\Gamma^{-1}(t-)dN^k(t),$$

where $\Gamma(0) = \Gamma^{-1}(0) = \mathbf{I}$ and $\mathbf{I}$ is the $(n\times n)$-identity matrix.

Write, for each $t \in \mathscr{T}$,

$$\bar{\sigma}(\mathbf{X}(t)) := \Gamma^{-1}(t)\sigma(\mathbf{X}(t)).$$

This is a transformed unnormalized filter for $\mathbf{X}(t)$. Then the following theorem gives a filtering equation for $\bar{\sigma}(\mathbf{X}(t))$.

Theorem 8.2. $\bar{\sigma}(\mathbf{X}(t))$ *satisfies the following linear, vector-valued, ordinary differential equation:*

$$\frac{d\bar{\sigma}(\mathbf{X}(t))}{dt} = \Gamma^{-1}(t)\mathbf{A}\Gamma(t)\bar{\sigma}(\mathbf{X}(t)).$$

Suppose $\pi(\mathbf{X}(t))$ is a version of the expectation $\mathrm{E}[\mathbf{X}(t)|\mathscr{F}(t)]$. Note that $\sigma(\mathbf{X}(t)) := \Gamma(t)\bar{\sigma}(\mathbf{X}(t))$, so

$$\pi(\mathbf{X}(t)) = \frac{\Gamma(t)\bar{\sigma}(\mathbf{X}(t))}{\langle\Gamma(t)\bar{\sigma}(\mathbf{X}(t)), \mathbf{1}\rangle}.$$

8.4 A Robust Filter-Based EM Algorithm

In this section, we first give the (robust) filtering equations for some quantities related to the hidden Markov chain. Then these filtering equations are used to develop a (robust) filter-based EM algorithm for the online recursive estimates of model parameters. These model parameters include the default intensities of different constituents in the credit portfolio over different levels of the dynamic frailty factor and the transition intensities of the levels of the factor.

We start by defining the following quantities related to the hidden Markov chain $\mathbf{X}$:

1. $\mathscr{O}^i(t)$ is the occupation time of the chain $\mathbf{X}$ in state $\mathbf{e}_i$ up to time t, for each $i = 1,2,\cdots,n$ and each $t \in \mathscr{T}$. That is,

$$\mathscr{O}^i(t) := \int_0^t \langle \mathbf{X}(u), \mathbf{e}_i \rangle \, du \, .$$

2. $\mathscr{J}^{ji}(t)$ is the number of transitions of the chain from state $\mathbf{e}_i$ to state $\mathbf{e}_j$ up to time t, for each $i, j = 1,2,\cdots,n$ with $j \neq i$ and each $t \in \mathscr{T}$. That is,

$$\mathscr{J}^{ji}(t) := \int_0^t \langle \mathbf{X}(u-), \mathbf{e}_i \rangle \left\langle \mathbf{e}_j, d\mathbf{X}(u) \right\rangle \, .$$

3. $\mathscr{L}_i^k(t)$ is the level integral for the state $\mathbf{e}_i$ with respect to N^k, for each $i = 1,2,\cdots,n$, $k = 1,2,\cdots,K$ and $t \in \mathscr{T}$. That is,

$$\mathscr{L}_i^k(t) := \int_0^t \langle \mathbf{X}(u), \mathbf{e}_i \rangle \, dN^k(u) \, .$$

Write, for each $t \in \mathscr{T}$, $i, j = 1,2,\cdots,n$ and $k = 1,2,\cdots,K$,

$$\begin{aligned}
\sigma(\mathscr{O}^i(t)\mathbf{X}(t)) &:= \mathrm{E}^{\dagger}[Z(t)\mathscr{O}^i(t)\mathbf{X}(t)|\mathscr{F}(t)] \, , \\
\sigma(\mathscr{J}^{ji}(t)\mathbf{X}(t)) &:= \mathrm{E}^{\dagger}[Z(t)\mathscr{J}^{ji}(t)\mathbf{X}(t)|\mathscr{F}(t)] \, , \\
\sigma(\mathscr{L}_i^k(t)\mathbf{X}(t)) &:= \mathrm{E}^{\dagger}[Z(t)\mathscr{L}_i^k(t)\mathbf{X}(t)|\mathscr{F}(t)] \, .
\end{aligned}$$

The following theorem gives the filtering equations for $\sigma(\mathscr{O}^i(t)\mathbf{X}(t))$, $\sigma(\mathscr{J}^{ji}(t)\mathbf{X}(t))$ and $\sigma(\mathscr{L}_i^k(t)\mathbf{X}(t))$.

Theorem 8.3. *For each $i, j = 1,2,\cdots,n$ with $j \neq i$ and each $k = 1,2,\cdots,K$, $\sigma(\mathscr{O}^i(t)\mathbf{X}(t))$, $\sigma(\mathscr{J}^{ji}(t)\mathbf{X}(t))$ and $\sigma(\mathscr{L}_i^k(t)\mathbf{X}(t))$ satisfy the following stochastic differential equations:*

$$\begin{aligned}
&\sigma(\mathscr{L}_i^k(t)\mathbf{X}(t)) \\
&= \int_0^t \mathbf{A}\sigma(\mathscr{L}_i^k(u)\mathbf{X}(u))du + \int_0^t \langle \sigma(\mathbf{X}(u)), \mathbf{e}_i \rangle \, dN^k(u)\mathbf{e}_i \\
&\quad + \sum_{k=1}^{K} \int_0^t \mathbf{diag}(\lambda^k - \mathbf{1})\sigma(\mathscr{L}_i^k(u)\mathbf{X}(u))dN^k(u)
\end{aligned}$$

$$
\begin{aligned}
&- \sum_{k=1}^{K} \int_0^{\tau^k \wedge t} \mathbf{diag}(\lambda^k - \mathbf{1}) \sigma(\mathscr{L}_i^k(u)\mathbf{X}(u)) du \\
&+ \sum_{k=1}^{K} \int_0^t \langle \sigma(\mathbf{X}(u)), \mathbf{e}_i \rangle \left\langle \lambda^k - \mathbf{1}, \mathbf{e}_i \right\rangle dN^k(u) \mathbf{e}_i \ ,
\end{aligned}
$$

$$
\begin{aligned}
&\sigma(\mathscr{J}^{ji}(t)\mathbf{X}(t)) \\
&= \int_0^t \mathbf{A} \sigma(\mathscr{J}^{ji}(u)\mathbf{X}(u)) du + \int_0^t \langle \sigma(\mathbf{X}(u)), \mathbf{e}_i \rangle a_{ij} du \mathbf{e}_j \\
&+ \sum_{k=1}^{K} \int_0^t \mathbf{diag}(\lambda^k - \mathbf{1}) \sigma(\mathscr{J}^{ji}(u)\mathbf{X}(u)) dN^k(u) \\
&- \sum_{k=1}^{K} \int_0^{\tau^k \wedge t} \mathbf{diag}(\lambda^k - \mathbf{1}) \sigma(\mathscr{J}^{ji}(u)\mathbf{X}(u)) du \ ,
\end{aligned}
$$

and

$$
\begin{aligned}
&\sigma(\mathscr{O}^i(t)\mathbf{X}(t)) \\
&= \int_0^t \mathbf{A} \sigma(\mathscr{O}^i(u)\mathbf{X}(u)) du + \int_0^t \langle \sigma(\mathbf{X}(u)), \mathbf{e}_i \rangle du \mathbf{e}_i \\
&+ \sum_{k=1}^{K} \int_0^t \mathbf{diag}(\lambda^k - \mathbf{1}) \sigma(\mathscr{O}^i(u)\mathbf{X}(u)) dN^k(u) \\
&- \sum_{k=1}^{K} \int_0^{\tau^k \wedge t} \mathbf{diag}(\lambda^k - \mathbf{1}) \sigma(\mathscr{O}^i(u)\mathbf{X}(u)) du \ ,
\end{aligned}
$$

where $\sigma(\mathscr{L}_i^k(0)\mathbf{X}(0)) = \sigma(\mathscr{J}^{ji}(0)\mathbf{X}(0)) = \sigma(\mathscr{O}^i(0)\mathbf{X}(0)) = \mathbf{0} \in \Re^n$.

Again we shall simplify the filtering equations for the unnormalized filters $\sigma(\mathscr{L}_i^k(t)\mathbf{X}(t))$, $\sigma(\mathscr{J}^{ji}(t)\mathbf{X}(t))$ and $\sigma(\mathscr{O}^i(t)\mathbf{X}(t))$ using the transformation matrix $\Gamma(t)$.

Write, for each $t \in \mathscr{T}$, $k = 1, 2, \cdots, K$ and $i, j = 1, 2, \cdots, n$,

$$
\begin{aligned}
\bar{\sigma}(\mathscr{L}_i^k(t)\mathbf{X}(t)) &:= \Gamma^{-1}(t) \sigma(\mathscr{L}_i^k(t)\mathbf{X}(t)) \ , \\
\bar{\sigma}(\mathscr{J}^{ji}(t)\mathbf{X}(t)) &:= \Gamma^{-1}(t) \sigma(\mathscr{J}^{ji}(t)\mathbf{X}(t)) \ , \\
\bar{\sigma}(\mathscr{O}^i(t)\mathbf{X}(t)) &:= \Gamma^{-1}(t) \sigma(\mathscr{O}^i(t)\mathbf{X}(t)) \ .
\end{aligned}
$$

Then the following theorem gives the filtering equations for these transformed unnormalized filters $\bar{\sigma}(\mathscr{L}_i^k(t)\mathbf{X}(t))$, $\bar{\sigma}(\mathscr{J}^{ji}(t)\mathbf{X}(t))$ and $\bar{\sigma}(\mathscr{O}^i(t)\mathbf{X}(t))$.

Theorem 8.4. *$\bar{\sigma}(\mathscr{L}_i^k(t)\mathbf{X}(t))$, $\bar{\sigma}(\mathscr{J}^{ji}(t)\mathbf{X}(t))$ and $\bar{\sigma}(\mathscr{O}^i(t)\mathbf{X}(t))$ satisfy the following linear, vector-valued, filtering equations:*

$$
\begin{aligned}
&\bar{\sigma}(\mathscr{L}_i^k(t)\mathbf{X}(t)) \\
&= N^k(t) \langle \bar{\sigma}(\mathbf{X}(t)), \mathbf{e}_i \rangle \mathbf{e}_i - \int_0^t N^k(u) \langle d\bar{\sigma}(\mathbf{X}(u)), \mathbf{e}_i \rangle \mathbf{e}_i + \int_0^t \mathbf{A} \bar{\sigma}(\mathscr{L}_i^k(u)\mathbf{X}(u)) du
\end{aligned}
$$

$$+\sum_{k=1}^{K}\langle\bar{\sigma}(\mathbf{X}(t)),\mathbf{e}_i\rangle\left\langle\lambda^k-\mathbf{1},\mathbf{e}_i\right\rangle\mathbf{diag}\left(\frac{\mathbf{1}}{\lambda^k}\right)N^k(t)\mathbf{e}_i$$
$$-\sum_{k=1}^{K}\int_0^t N^k(u)\left\langle\lambda^k-\mathbf{1},\mathbf{e}_i\right\rangle\mathbf{diag}\left(\frac{\mathbf{1}}{\lambda^k}\right)\langle d\bar{\sigma}(\mathbf{X}(u)),\mathbf{e}_i\rangle\ ,$$

$$\begin{aligned}&\bar{\sigma}(\mathscr{J}^{ji}(t)\mathbf{X}(t))\\&=\int_0^t\langle\bar{\sigma}(\mathbf{X}(u)),\mathbf{e}_i\rangle\,a_{ij}du\mathbf{e}_j+\int_0^t\mathbf{A}\bar{\sigma}(\mathscr{J}^{ji}(u)\mathbf{X}(u))du\ ,\end{aligned}$$

and

$$\begin{aligned}&\bar{\sigma}(\mathscr{O}^i(t)\mathbf{X}(t))\\&=\int_0^t\langle\bar{\sigma}(\mathbf{X}(u)),\mathbf{e}_i\rangle\,dt\mathbf{e}_i+\int_0^t\mathbf{A}\bar{\sigma}(\mathscr{O}^i(u)\mathbf{X}(u))du\ ,\end{aligned}$$

where $\bar{\sigma}(\mathscr{L}_i^k(0)\mathbf{X}(0))=\bar{\sigma}(\mathscr{J}^{ji}(0)\mathbf{X}(0))=\bar{\sigma}(\mathscr{O}^i(0)\mathbf{X}(0))=\mathbf{0}\in\Re^n$.

Our goal is to estimate a_{ji} and λ_i^k, for each $i,j=1,2,\cdots,n$ and $k=1,2,\cdots,K$. The parameter estimates $\widehat{a}_{ji}$ and $\widehat{\lambda}_i^k$ of a_{ji} and λ_i^k, respectively, are given in Dembo and Zeitouni (1989) [8] as follows:

$$\begin{aligned}\widehat{a}_{ji}&=\frac{\mathrm{E}[\mathscr{J}^{ji}(t)|\mathscr{F}(t)]}{\mathrm{E}[\mathscr{O}^i(t)|\mathscr{F}(t)]}=\frac{\sigma(\mathscr{J}^{ji}(t))}{\sigma(\mathscr{O}^i(t))}\ ,\quad i\neq j\ ,\\\widehat{\lambda}_i^k&=\frac{\mathrm{E}[\mathscr{L}_i^k(t)|\mathscr{F}(t)]}{\mathrm{E}[\mathscr{O}^i(t)|\mathscr{F}(t)]}=\frac{\sigma(\mathscr{L}_i^k(t))}{\sigma(\mathscr{O}^i(t))}\ .\end{aligned}$$

It is not difficult to see that

$$\begin{aligned}\sigma(\mathscr{L}_i^k(t))&=\left\langle\sigma(\mathscr{L}_i^k(t)\mathbf{X}(t)),\mathbf{1}\right\rangle=\left\langle\Gamma(t)\bar{\sigma}(\mathscr{L}_i^k(t)\mathbf{X}(t)),\mathbf{1}\right\rangle\ ,\\\sigma(\mathscr{J}^{ji}(t))&=\left\langle\sigma(\mathscr{J}^{ji}(t)\mathbf{X}(t)),\mathbf{1}\right\rangle=\left\langle\Gamma(t)\bar{\sigma}(\mathscr{J}^{ji}(t)\mathbf{X}(t)),\mathbf{1}\right\rangle\ ,\\\sigma(\mathscr{O}^i(t))&=\left\langle\sigma(\mathscr{O}^i(t)\mathbf{X}(t)),\mathbf{1}\right\rangle=\left\langle\Gamma(t)\bar{\sigma}(\mathscr{O}^i(t)\mathbf{X}(t)),\mathbf{1}\right\rangle\ .\end{aligned}$$

Consequently, to evaluate the parameter estimates $\widehat{a}_{ji}$ and $\widehat{\lambda}_i^k$, we can implement a filter bank consisting of transformed unnormalized recursive filters given in Theorems 8.2 and 8.4. These parameter estimates can then be evaluated by following the three steps in the (robust) filter-based expectation maximization (EM) algorithm described as follows:

Step I: Choose the initial estimates $\widehat{a}_{ji}(0)$ and $\widehat{\lambda}_i^k(0)$.

Step II: Compute the maximum likelihood estimates $\widehat{a}_{ji}(l+1)$ and $\widehat{\lambda}_i^k(l+1)$ using $\widehat{a}_{ji}(l)$ and $\widehat{\lambda}_i^k(l)$ and the filter bank consisting of recursive filters given in Theorem 8.2 and Theorem 8.4, where l represents the lth run of the algorithm.

Step III: Stop when $|\widehat{a}_{ji}(l+1)-\widehat{a}_{ji}(l)|<\varepsilon_1$ and $|\widehat{\lambda}_i^k(l+1)-\widehat{\lambda}_i^k(l)|<\varepsilon_2$; otherwise, continue from Step II, where ε_1 and ε_2 are the desirable levels of accuracy and both of them are positive.

8.5 Variance Dynamics

We present the variance dynamics for the observation processes and the hidden state process in our HMM intensity-based credit risk model. Firstly, we recall that the observation processes and the hidden state process are, respectively, given by:

$$N^k(t) = \int_0^{\tau^k \wedge t} \left\langle \lambda^k, \mathbf{X}(u) \right\rangle du + \bar{N}^k(t) \ , \quad k = 1, 2, \cdots, K \ ,$$

and

$$\mathbf{X}(t) = \mathbf{X}(0) + \int_0^t \mathbf{A}\mathbf{X}(u-)du + \mathbf{M}(t) \ .$$

Note that for each $k = 1, 2, \cdots, K$, $\{\bar{N}^k(t) | t \in \mathscr{T}\}$ is the innovations process for the observation process $\{N^k(t) | t \in \mathscr{T}\}$, while $\{\mathbf{M}(t) | t \in \mathscr{T}\}$ represents the innovations process of the hidden state process $\{\mathbf{M}(t) | t \in \mathscr{T}\}$. From the modeling and estimation perspectives, one may be interested in the variance dynamics of the innovations processes $\{N^k(t) | t \in \mathscr{T}\}$, $k = 1, 2, \cdots, K$, and $\{\mathbf{M}(t) | t \in \mathscr{T}\}$. Since we are considering a continuous-time modeling setup here, the (predictable) quadratic variation processes of the innovations processes represent natural variance dynamics of the innovations processes. Consequently, we present the (predictable) quadratic variation processes of the innovations process $\{N^k(t) | t \in \mathscr{T}\}$, $k = 1, 2, \cdots, K$, and $\{\mathbf{M}(t) | t \in \mathscr{T}\}$ here.

Firstly, the (predictable) quadratic variation process $\{\langle \mathbf{M}, \mathbf{M} \rangle (t) | t \in \mathscr{T}\}$ of the innovations process $\{\mathbf{M}(t) | t \in \mathscr{T}\}$ was obtained in Elliott et al. (1995) [17] (see Lemma 1.3 in Appendix B therein). We state it here without proof.

$$\langle \mathbf{M}, \mathbf{M} \rangle (t) = \mathbf{diag}\left(\int_0^t \mathbf{A}\mathbf{X}(u-)du \right) - \int_0^t \mathbf{diag}(\mathbf{X}(u-))\mathbf{A}' du - \int_0^t \mathbf{A}\mathbf{diag}(\mathbf{X}(u-))du \ .$$

Note that $\{\langle \mathbf{M}, \mathbf{M} \rangle (t) | t \in \mathscr{T}\}$ may be related to a continuous-time version of the conditional mean-square-error process of the hidden state process $\{\mathbf{X}(t) | t \in \mathscr{T}\}$. Furthermore, $\langle \mathbf{M}, \mathbf{M} \rangle (t)$ depends on unknown parameters represented by the components in the rate matrix $\mathbf{A}$ as well as the hidden state process $\{\mathbf{X}(t) | t \in \mathscr{T}\}$. Consequently, we may provide an estimate $\widehat{\langle \mathbf{M}, \mathbf{M} \rangle}(t)$ of $\langle \mathbf{M}, \mathbf{M} \rangle (t)$ as follows:

$$\widehat{\langle \mathbf{M}, \mathbf{M} \rangle}(t) = \mathbf{diag}\left(\int_0^t \widehat{\mathbf{A}}\widehat{\mathbf{X}}(u-)du \right) - \int_0^t \mathbf{diag}(\widehat{\mathbf{X}}(u-))\widehat{\mathbf{A}}' du - \int_0^t \widehat{\mathbf{A}}\mathbf{diag}(\widehat{\mathbf{X}}(u-))du \ .$$

Here $\widehat{\mathbf{X}}(t)$ is the filtered estimate of $\mathbf{X}(t)$ presented in Sect. 8.3 and $\widehat{\mathbf{A}}$ is the filter-based EM estimate of $\mathbf{A}$ presented in Sect. 8.4. Note that $\{\widehat{\langle \mathbf{M}, \mathbf{M} \rangle}(t) | t \in \mathscr{T}\}$ may be related to a continuous-time version of the conditional standard error process of the hidden state process $\{\mathbf{X}(t) | t \in \mathscr{T}\}$, which is an estimate of the model estimation error frequently used in statistical analysis in discrete-time.

Similarly, we consider the (predictable) quadratic variation process $\{\langle \bar{N}^k, \bar{N}^k \rangle (t) | t \in \mathscr{T}\}$ of the innovations process $\{\bar{N}^k(t) | t \in \mathscr{T}\}$, for each $k = 1, 2, \cdots, K$, and its estimate $\{\widehat{\langle \bar{N}^k, \bar{N}^k \rangle}(t) | t \in \mathscr{T}\}$ in the sequel. Firstly, $\langle \bar{N}^k, \bar{N}^k \rangle (t)$ is given by:

$$\left\langle \bar{N}^k, \bar{N}^k \right\rangle (t) = \left\langle N^k, N^k \right\rangle (t) = \int_0^{\tau^k \wedge t} \left\langle \lambda^k, \mathbf{X}(u) \right\rangle du \,, \quad k = 1, 2, \cdots, K \,, \quad t \in \mathscr{T} \,.$$

This may be related to a continuous-time version of the conditional mean-square-error process of the observation process $\{N^k(t) | t \in \mathscr{T}\}$.

Again $\langle \bar{N}^k, \bar{N}^k \rangle (t)$ depends on unknown parameters represented by the components in the vector λ^k as well as the hidden state process $\{\mathbf{X}(t) | t \in \mathscr{T}\}$. Consequently, we may estimate $\langle \bar{N}^k, \bar{N}^k \rangle (t)$ by $\widehat{\langle \bar{N}^k, \bar{N}^k \rangle}(t)$ defined as follows:

$$\widehat{\langle \bar{N}^k, \bar{N}^k \rangle}(t) = \int_0^{\tau^k \wedge t} \left\langle \widehat{\lambda}^k, \widehat{\mathbf{X}}(u) \right\rangle du \,.$$

Here $\widehat{\mathbf{X}}(t)$ is the filtered estimate of $\mathbf{X}(t)$ presented in Sect. 8.3 and $\widehat{\lambda}^k$ is the filter-based EM estimate of λ^k presented in Sect. 8.4.

8.6 Default Probabilities

In this section, we present an analytical formula for the joint conditional distribution of defaults of K firms. The joint distribution is obtained using an analytical formula for the joint characteristic function of occupation times of the chain $\mathbf{X}$ together with the robust filter of the chain $\mathbf{X}$ derived in Sect. 8.3.

For each $i = 1, 2, \cdots, n$ and $t, t+h \in \mathscr{T}$ with $h > 0$, let $\mathscr{O}^i(t,h)$ be the occupation time of the chain $\mathbf{X}$ in state $\mathbf{e}_i$ over the time interval $[t, t+h]$. That is,

$$\mathscr{O}^i(t,h) := \int_t^{t+h} \langle \mathbf{X}(u), \mathbf{e}_i \rangle \, du \,.$$

Write

$$\mathbf{O}(t,h) := (\mathscr{O}^1(t,h), \mathscr{O}^2(t,h), \cdots, \mathscr{O}^n(t,h)) \in [t, t+h]^n \subset (\Re_+)^n \,.$$

Then it has been shown in Buffington and Elliott (2002) [4] that the conditional joint characteristic function of $\mathbf{O}(t,h)$ given $\mathscr{G}(t)$ evaluated at $\zeta \in \Re^n$ under P is:

$$\begin{aligned} \Phi_{\mathbf{O}(t,h)|\mathscr{G}(t)}(\zeta) &:= \mathrm{E}[e^{\sqrt{-1}\left\langle \zeta, \mathbf{O}(t,h) \right\rangle} | \mathscr{G}(t)] \\ &= \left\langle \exp[(\mathbf{A} + \sqrt{-1}\mathbf{diag}(\zeta))h] \mathbf{X}(t), \mathbf{1} \right\rangle \,. \end{aligned}$$

Let $\phi_{\mathbf{O}(t,h)|\mathscr{G}(t)}(\mathbf{u})$ be the conditional joint probability density function of $\mathbf{O}(t,h)$ given $\mathscr{G}(t)$ under P, where $\mathbf{u} := (u^1, u^2, \cdots, u^n) \in [t, t+h]^n$. Then by the inverse Fourier transform,

$$\begin{aligned}
&\phi_{\mathbf{O}(t,h)|\mathscr{G}(t)}(\mathbf{u}) \\
&= \frac{1}{(2\pi)^n} \int_{\Re^n} e^{-\sqrt{-1}\left\langle \zeta, \mathbf{u} \right\rangle} \Phi_{\mathbf{O}(t,h)|\mathscr{G}(t)}(\zeta) d\zeta_1 d\zeta_2 \cdots d\zeta_n \\
&= \frac{1}{(2\pi)^n} \int_{\Re^n} e^{-\sqrt{-1}\left\langle \zeta, \mathbf{u} \right\rangle} \left\langle \exp[(\mathbf{A} + \sqrt{-1}\mathbf{diag}(\zeta))h]\mathbf{X}(t), \mathbf{1} \right\rangle d\zeta_1 d\zeta_2 \cdots d\zeta_n \,.
\end{aligned}$$

We wish to evaluate the conditional joint default probability of the portfolio of firms in the time horizon given observed information up to time t, say $\mathscr{F}(t)$ and given that these firms have survived at time t. The result is presented in the following theorem.

Theorem 8.5. *For each $t, t+h \in \mathscr{T}$ with $h > 0$, with $\mathbf{u} \in \Re^n$,*

$$\begin{aligned}
&P(t < \tau^1 < t+h, t < \tau^2 < t+h, \cdots, t < \tau^K < t+h | \mathscr{F}(t)) \\
&= \frac{1}{(2\pi)^n} \int_{\Re^n} \int_{[t,t+h]^n} \frac{\left\langle \exp[(\mathbf{A} + \sqrt{-1}\mathbf{diag}(\zeta))h] \Gamma(t) \bar{\sigma}(\mathbf{X}(t)), \mathbf{1} \right\rangle}{\left\langle \Gamma(t) \bar{\sigma}(\mathbf{X}(t)), \mathbf{1} \right\rangle} \\
&\quad \times \left[\prod_{k=1}^{K} \left(1 - e^{-\left\langle \lambda^k, \mathbf{u} \right\rangle} \right) I_{\{\tau^k > t\}} \right] e^{-\sqrt{-1}\left\langle \zeta, \mathbf{u} \right\rangle} du_1 du_2 \cdots du_n d\zeta_1 d\zeta_2 \cdots d\zeta_n \,,
\end{aligned}$$

where, for each $k = 1, 2, \cdots, K$, $I_{\{\tau^k > t\}}$ is the indicator function of the event $\{\tau_k > t\}$; $\bar{\sigma}(\mathbf{X}(t))$ follows the linear, vector-valued, ordinary differential equation in Theorem 8.2.

8.7 Conclusion

An intensity-based model for the dependent default risk of constituents in a credit portfolio is considered, where the constituents are exposed to a common hidden dynamic frailty factor described by a continuous-time, finite-state, hidden Markov chain. Filtering equations of the hidden factor and an estimation algorithm of the model are given. An analytical formula for the joint default probability is obtained in Fourier transform space.

Acknowledgments The authors would like to thank the referee for helpful comments. We also wish to acknowledge the Discovery Grant from the Australian Research Council (ARC) (Project No.: DP1096243).

References

1. Arnsdorf, M., & Halperin, I. (2007). BSLP: Markovian bivariate spread-loss model for portfolio credit derivatives. Working Paper, Quantitative Research J.P. Morgan.
2. Black, F., & Scholes, M. S. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3), 637–654.
3. Brigo, D., Pallavicini, A., & Torresetti, R. (2006). Calibration of CDO tranches with the dynamical generalized-Poisson loss model. Working Paper, Banca IMI.
4. Buffington J., & Elliott, R. J. (2002). American options with regime switching. *International Journal of Theoretical and Applied Finance*, 5, 497–514.
5. Clark, J. M. (1978). The design of robust approximations to the stochastic differential equations for nonlinear filtering. In J. K. Skwirzynski (Ed.), *Communications systems and random process theory* (pp. 721–734). Amsterdam: Sijthoff and Noorhoff.
6. Cont, R., & Minca, A. (2008). Reconstructing portfolio default rates from CDO tranche spreads. Working Paper, Columbia University.
7. Cooper, I., & Martin, M. (1996). Default risk and derivative products. *Applied Mathematical Finance*, 3, 53–74.
8. Dembo, A., & Zeitouni, O. (1989). Parameter estimation of partially observed continuous time stochastic process via the EM algorithm. *Stochastic Processes and Applications*, 23, 91–113.
9. Das, S., Duffie, D., Kapadia, N., & Saita, L. (2007). Common failings: how corporate defaults are correlated. *Journal of Finance*, 62, 93–117.
10. Davis, M., & Lo, V. (2001). Modeling default correlation in bond portfolios. In C. Alexander (Ed.), *Mastering risk volume 2: applications* (pp. 141–151). Prentice Hall.
11. Ding, X., Giesecke, K., & Tomecek, P. (2006). Time-changed birth processes and multi-name credit derivatives. Working Paper, Stanford University.
12. Duffie, D., & Garleanu, N. (2001). Risk and valuation of collateralized debt obligations. *Financial Analysts Journal*, 57(1), 41–59.
13. Duffie, D., Saita, L., & Wang, K. (2006). Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, 83(3), 635–665.
14. Duffie, D., Eckner, A., Horel, G., & Saita, L. (2009). Frailty correlated default. *Journal of Finance*, 64(5), 2089–2123.
15. Eckner, A. (2007). Computational techniques for basic affine models of portfolio credit risk. Working Paper, Stanford University.
16. Elliott, R. J.: Stochastic Calculus and Applications. Springer-Verlag, Berlin (1982)
17. Elliott, R. J, Aggoun, L., & Moore, J. B.: Hidden Markov Models: Estimation and Control. Springer-Verlag, Berlin (1995)
18. Elliott, R. J., Jeanblanc, M., & Yor, M. (2000). On models of default risk. *Mathematical Finance*, 10(2), 179–195.
19. Elliott, R. J., & Siu, T. K. (2011). Filtering a hidden Markov-modulated reduced-form credit risk model and its application to credit VaR. Preprint. Haskayne School of Business, University of Calgary and Faculty of Business and Economics, Macquarie University.
20. Feldhütter, P. (2007). An empirical investigation of an intensity based model for pricing CDO tranches, Working Paper, Copenhagen Business School.
21. Giesecke, K., & Goldberg, L. (2004). Sequential defaults and incomplete information. *Journal of Risk*, 7(1), 1–26.
22. Giesecke, K., & Goldberg, L. (2005). A top down approach to multi-name credit. Working Paper, Stanford University.
23. Giesecke, K., & Weber, S. (2006). Credit contagion and aggregate loss. *Journal o Economic Dynamics and Control*, 30, 741–761.
24. Jarrow, R. A., & Turnbull, S.M. (1995). Pricing derivatives on financial securities subject to credit risk. *Journal of Finance*, 50, 53–86.
25. Jarrow, R. A., & Yu, F. (2001). Counterparty risk and the pricing of defaultable securities. *Journal of Finance*, 56(5), 555–576.

26. Kusuoka, S. (1999). A remark on default risk models. *Advances in Mathematical Economics*, 1, 69–82.
27. Longstaff, F., & Rajan, A. (2007). An empirical analysis of collateralized debt obligations. *Journal of Finance*, 63(2), 529–563.
28. Madan, D., & Unal, H. (1998). Pricing the risks of default. *Review of Derivatives Research*, 2(2–3), 121–160.
29. Merton, R. C. (1974). On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance*, 29(2), 449–470.
30. Mortensen, A. (2006). Semi-analytical valuation of basket credit derivatives in intensity-based models. *Journal of Derivatives*, 13, 8–26.
31. Schönbucher, P., & Schubert, D. (2001). Copula-dependent default risk in intensity models. Working paper, Universit at Bonn.
32. Schönbucher, P. (2004). Information-driven default contagion. Working Paper, ETH Zürich.
33. Yu, F. (2007). Correlated defaults in intensity based models. *Mathematical Finance*, 17, 155–173.

Chapter 9
Yield Curve Modelling Using a Multivariate Higher-Order HMM

Xiaojing Xi and Rogemar Mamon

9.1 Introduction

Interest rate modelling is a central consideration in financial market given its fundamental importance in pricing, risk management and investment. Classical short-term interest rate models, such as those proposed by Merton [1], Vasicek [2], Cox et al. [3] and Hull and White [4], assume deterministic parameters. However, we all know that the economy and market are subjected to dynamic and in some cases significant changes. Such changes have substantial impact on the evolution of interest rates. Research works in recent years focus on the development of appropriate quantitative models suited for time-varying model parameters to accurately capture the behaviour of various financial variables and economic indicators. The introduction of regime-switching models provided some ways of incorporating the impact of market and economic changes on interest rates. Pioneered by the work of Hamilton [5], the construction of regime-switching-based methods was explored in the modelling of non-stationary time series. In such models, values of the model parameters at a particular moment depends on the state of an underlying Markov chain at that moment. A study by Smith [6] found evidence that the volatility depends on the level of the short rate and supports Markov-switching model over a stochastic volatility model. Landén [7] developed a hidden Markov model (HMM) for the short-term interest rates, in which the mean and variance are governed by an underlying Markov process. In practice, the underlying state of the market and volatilities are unobservable and so parameter estimation for these

X. Xi
Department of Applied Mathematics, Western University, London, ON, Canada
e-mail: xxi2@uwo.ca

R. Mamon (✉)
Department of Statistical and Actuarial Sciences, Western University, 1151 Richmond Street, London, ON, Canada N6A 5B7
e-mail: rmamon@stats.uwo.ca

Y. Zeng and S. Wu (eds.), *State-Space Models: Applications in Economics and Finance*, Statistics and Econometrics for Finance 1, DOI 10.1007/978-1-4614-7789-1_9,

Markov-switching models presents some challenges from both the practical and the mathematical standpoints.

In a comprehensive work, Elliott et al. [8] provided recursive self-updating estimates for the Markov chain, the model drift and diffusion parameters modulated by the same Markov chain. Elliott et al. [9] proposed a multivariate HMM for the short rate process and HMM filtering techniques are employed in their implementation. Erlwein and Mamon [10] considered a Hull–White interest rate model in which the interest rate's volatility, mean-reverting level and speed of mean-reversion are all governed by a Markov chain in discrete time. The HMM filters are derived and implemented on a financial data set. Their analysis of prediction errors together with the aid of the Akaike information criterion shows that a two-regime model is sufficient to describe the interest rate dynamics in their study. More recent developments on regime-switching focus on extending various models. Hunt and Devolder [11] studied an extension of the Ho and Lee model under a semi-Markov regime-switching framework, and an application of their proposed extension to the pricing of European bond options is given. Zhou and Mamon [12] investigated the Vasicek, CIR and Black–Karasinski models with the parameters of the short rate being modulated by a finite-state Markov chain in discrete time. A quasi-maximum likelihood method is utilized to estimate the model parameters and implementation of their algorithms is carried out on the Canadian yield rates.

Some recent studies examine the integration of regime-switching models with other modelling approaches to obtain a richer methodology. A four-state model to capture rate dynamics in the US spot and forward rate markets was proposed by Guidolin and Timmermann [13]; their an out-of-sample forecasting exercise show evidence that, at short horizons, combining regime-switching forecasts with simpler univariate time-series forecasts can help reduce the root mean squared forecast error. Meligkotsidou and Dellaportas [14] adopted a Bayesian forecasting methodology of discrete-time finite state-space HMM with non-constant transition matrix in modelling monthly data on rates of return series; the results of their MCMC algorithms indicate that non-homogeneous HMMs improve the predictive ability of the model when compared against a standard homogeneous HMM.

Other papers on regime-switching models continue to develop new approaches in detecting further evidence of regime shifts in the market. Startz and Tsang [15] constructed an unobserved component model in which the short-term interest rate is composed of a stochastic trend and a stationary cycle; results from their model-based measures suggest that allowing for regime switching in shock variances improves model performance. Audrino and Mederos [16] proposed a smooth transition tree model that combines regression trees and GARCH models to describe regime switches in the short-term interest rate series; their empirical results provide evidence of the power of the model in forecasting the conditional mean and variance. Utilizing an adapted unit-root test, Holmes et al. [17] found evidence that Australian and New Zealand interest rates can switch between regimes characterized by differences in mean, variance and persistence.

For a review of models of term structure of interest rates under regime-switching setting, including earlier regime-switching models of short-term interest rate in

discrete time, and recent Markov-switching models in continuous-time, refer to Nieh et al. [18]. Whist the original HMM can reasonably model the impact of structural changes in the financial time series, there is a need to also develop quantitative models that are able to capture time series memory. Processes with long memory characteristics have stronger coupling between values at different times than that of short-memory processes, and they can be described by heavy-tailed distributions. Mandelbrot [19] demonstrated applications of stochastic processes with long-memory in economics and finance. Cajueiro and Tabak [20, 21] showed evidence of long-range dependence in the LIBOR and US interest rates. McCarthy et al. [22] investigated the presence of long memory in corporate bond yield spreads and found strong evidence that such presence exists. Numerous studies have developed stochastic models to capture the long-range dependence property in financial time series. Maheu [23] concluded that GARCH models can capture the long-memory property in volatility of financial prices under some circumstances. Dajcman [24] proposed an autoregressive fractionally integrated moving average (ARFIMA) model for eight European stock market returns. Duan and Jacobs [25] suggested that inclusion of long-range dependence in their model improves significantly model fitting performance on real interest rate data. It seems, however, that the existing literature on long-memory property of time series mainly concentrates on single-state stochastic models. This paper contributes to the widening of literature on the development of models that are able to capture not only regime-switching but also short- or long-term dependence in the HMM that modulates regime switches. We put forward a weak hidden Markov model (WHMM), also known as a higher-order HMM to model the movement of the term structure of interest rates. As Solberg [26] indicated, the real significance of WHMM is to rectify the weakness of the usual HMM. One may feel that HMM's memoryless property is unwarranted for many stochastic processes observed in real-life applications. WHMM generalizes HMM and therefore, the memoryless property implied by the Markov assumption is not really as restrictive as it first appears. By using WHMM, the probability of current state does not depend on just one prior time epoch but on any finite number of prior epochs, and so more information from the past are taken into account. The higher-order Markov chain and its applications in finance have been investigated by a number of authors, and these include, amongst others, Xi and Mamon [27] for returns of risky assets; Siu et al. [28] for risk measurement; Ching et al. [29] for exotic option pricing and Siu [30] for spot rates and credit ratings.

In this paper, we consider a multivariate WHMM for the evolution of the term structure of interest rates. Extending the formulation in [9], the short-term rate can be rewritten as a function of a discrete time weak Markov chain (WMC). We assume the drift and diffusion terms of the yield values are governed by a second-order multivariate Markov chain. The states of the WMC are associated with the states of the market, whose current behaviour depends on the behaviour at the previous two time steps. We utilize a transformation that converts a WHMM into a regular HMM allowing us to apply the usual HMM estimation algorithm.

The remainder of this paper is organized as follows. Section 9.2 describes the formulation of the multivariate modelling framework. The derivation of the filters

for the states of WMC and other related processes through a measure change is presented. The recursive estimations are obtained by applying the expectation–maximization (EM) algorithm. The implementation of this proposed model is given in Sect. 9.3. The data set involved in our numerical study consists of daily US Treasury yields. In Sect. 9.4, we provide a discussion on how to select the optimal number of states. Using some metrics and criteria, we conclude that a two-state WHMM is sufficient to capture the market dynamics of our data. We also present an analysis of h-day ahead forecasts under the one-, two-, three- and four-state settings. Forecasting errors generated under the WHMM are compared to those generated under the regular HMM. Our results demonstrate that by including a memory-capturing mechanism in the model, WHMM outperforms the HMM in terms of low forecasting error. We conclude with some remarks in Sect. 9.5.

9.2 Filtering and Parameter Estimation

We assume all processes in our modelling set-up are supported by a stochastic basis $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}, P)$, where P is a risk-neutral measure used in the pricing of the zero-coupon bond. Let $\{\mathbf{x}_t\}$, $t \geq 0$ be a continuous-time WMC with finite space $\mathscr{S} = \{s_1, s_2, \ldots, s_N\}$. Without loss of generality, we identify the points in $\mathscr{S}$ with the canonical basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_N\} \subset \mathbf{R}^N$, where $\mathbf{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)^\top$ and $\top$ denotes the transpose of a vector. The representation $\langle \mathbf{x}_t, \mathbf{e}_i \rangle$ refers to the event that the economy is in state i at time t. Here, $\langle \, , \, \rangle$ stands for the inner product in $\mathbf{R}^N$. We suppose the short rate process $\mathbf{r}_t$ is a function of the unobservable Markov chain $\mathbf{x}_t$, such that $\mathbf{r}_t = r(\mathbf{x}_t) = \langle \mathbf{r}, \mathbf{x}_t \rangle$ for some vector $\mathbf{r} \in \mathbf{R}^N$. At time t, a zero-coupon bond F^i, expiring at time $t + \tau_i$, $i = 1, \ldots, d$, has a price

$$F^i(\mathbf{x}_t, t) = E\left[\exp\left(-\int_t^{t+\tau_i} r(\mathbf{x}_s) ds\right) \Big| \mathscr{F}_t\right],$$

where the expectation is taken with respect to the risk-neutral measure.

It was shown in [28] that the yield values in discrete time can be expressed as $\langle \mathbf{f}^i, \mathbf{x}_k \rangle = -\frac{1}{\tau_i} \log F^i(\mathbf{x}_t, t)$, and $\mathbf{x}_k = \mathbf{x}_{t_k}$, as a discrete-time version of the state process $\mathbf{x}_t$. In practice, it has to be noted that at each time k, one yield curve is observed determined by the yields of T-bills (maturities of 30 days to 1 year), T-notes (maturities between 1 and 5 years), and T-bonds (maturities more than 5 years). Suppose that there are d maturities (and hence d fixed-income instruments trading) corresponding to the d number of yields at each time k. Let $\mathbf{y}_k = (y_k^1, y_k^2, \ldots, y_k^d)$ denote the d-dimensional yield process. We assume that the drift and volatility of each yield component y_k^i can switch between N regimes, i.e., each y_k^i, $1 \leq i \leq d$, is modulated by a Markov chain $\mathbf{x}_k$. In other words, the yield values have dynamics

$$y_{k+1}^i = f^i(\mathbf{x}_k) + \sigma^i(\mathbf{x}_k) z_{k+1}^i.$$

The sequence $\{z_k^i\}$, for $i = 1, 2, \ldots, d$, is a sequence of $N(0,1)$ independent, identically distributed (IID) random variables, which are independent of the $\mathbf{x}$-process. More specifically, the functions f^i and σ^i are given by the vectors $\mathbf{f}^i = (f_1^i, f_2^i, \ldots, f_N^i)^\top$ and $\sigma^i = (\sigma_1^i, \sigma_2^i, \ldots, \sigma_N^i)^\top$ in $\mathbf{R}^N$, $f^i(\mathbf{x}_k) = \langle \mathbf{f}^i, \mathbf{x}_k \rangle$ and $\sigma^i(\mathbf{x}_k) = \langle \sigma^i, \mathbf{x}_k \rangle$ represent the mean and variance of y_k^i, respectively. Note that all components of the vector observation process are modulated by the same underlying WMC.

In this paper, our attention will solely be on a WMC of order 2 to simplify the discussion and present a complete characterization of the parameter estimation. The probability of the next time step for the WMC given the previous information is

$$\begin{aligned} P\ (\mathbf{x}_{k+1} = x_{k+1} | \mathbf{x}_0 = x_0, \ldots, \mathbf{x}_{k-1} = x_{k-1}, \mathbf{x}_k = x_k) \\ = P(\mathbf{x}_{k+1} = x_{k+1} | \mathbf{x}_{k-1} = x_{k-1}, \mathbf{x}_k = x_k). \end{aligned}$$

Each entry of the transition probability matrix $\mathbf{A} := (a_{lmv}) \in \mathbf{R}^{N \times N^2}$, where $l, m, v \in 1, \ldots, N$, refers to the probability that the process enters state l given that the current and previous states were in states m and v, respectively. The salient idea in the filtering method for WHMM is that, a second-order Markov chain is transformed into a first-order Markov chain through a mapping ξ, and then we may apply the regular filtering method. The mapping ξ is defined by

$$\xi(\mathbf{e}_r, \mathbf{e}_s) = \mathbf{e}_{rs}, \text{ for } 1 \le r, s \le N,$$

where $\mathbf{e}_{rs}$ is an $\mathbf{R}^{N^2}$-unit vector with unity in its $((r-1)N+s)$th position. Note that

$$\langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{e}_{rs} \rangle = \langle \mathbf{x}_k, \mathbf{e}_r \rangle \langle \mathbf{x}_{k-1}, \mathbf{e}_s \rangle$$

indicates the identification of the new first-order Markov chain with the canonical basis. The new $N^2 \times N^2$ transition probability matrix $\mathbf{\Pi}$ of the new Markov chain is defined by

$$\pi_{ij} = \begin{cases} a_{lmv} & \text{if} i = (l-1)N + m,\ j = (m-1)N + v \\ 0 & \text{otherwise.} \end{cases}$$

Here, each non-zero element π_{ij} represents the probability

$$\pi_{ij} = a_{lmv} = P(\mathbf{x}_k = \mathbf{e}_l | \mathbf{x}_{k-1} = \mathbf{e}_m,\ \mathbf{x}_{k-2} = \mathbf{e}_v),$$

and each zero represents an impossible transition. It may be shown that the new Markov chain $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ has the semi-martingale representation

$$\xi(\mathbf{x}_k, \mathbf{x}_{k-1}) = \mathbf{\Pi} \xi(\mathbf{x}_{k-1}, \mathbf{x}_{k-2}) + \mathbf{v}_k, \tag{9.1}$$

where $\{\mathbf{v}_k\}_{k \ge 1}$ is a sequence of $\mathbf{R}^{N^2}$ martingale increments. We recognize that the above approach of transforming a higher-order Markov chain to a corresponding Markov chain of order one-lag lower is cumbersome to adopt for orders higher

than 3. However, it has to be noted that for financial time series models such as the GARCH, ARCH and EWMA, the lags employed do not exceed 2 anyway in majority of the applications and therefore concentrating on lower lags is a reasonable modelling assumption.

Under P, the underlying WMC is not known. Instead, the state $\mathbf{x}_k$ is contained in the noisy market observations $\mathbf{y}_k, k \geq 1$. We aim to "filter" the noise out of the observations to determine $\mathbf{x}_k$. However, the derivation of filters under P is complicated. By Kolmogorov's extension theorem, there exists a reference probability measure $\bar{P}$ under which the $\mathbf{y}_k$'s are $N(0,1)$ IID random variables and therefore $\bar{P}$ is an easier measure to work with. Now, we perform a measure change to construct the real-world measure P from the ideal-world measure $\bar{P}$ by invoking a discrete-time version of Girsanov's theorem. Let $\phi(z)$ denote the probability density function of a standard normal random variable Z. For each component i, write

$$\lambda_l^{\,i} := \frac{\phi(\sigma^i(\mathbf{x}_{l-1})^{-1}(y_l^i - f^i(\mathbf{x}_{l-1})))}{\sigma^i(\mathbf{x}_{l-1})\phi(y_l^i)}. \tag{9.2}$$

The Radon–Nikodým derivative of P with respect to $\bar{P}$, $\frac{dP}{d\bar{P}}|_{\mathscr{Y}_k} := \Lambda_k$, is defined by

$$\Lambda_k = \prod_{i=1}^{d}\prod_{l=1}^{k}\lambda_l^i,\ k \geq 1,\ \Lambda_0 = 1. \tag{9.3}$$

To obtain the estimates of $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ under the real world measure, we first perform all calculations under the reference probability measure $\bar{P}$. Calculations under the two measures are linked via the Bayes' theorem for conditional expectation.

Let us derive the conditional expectation of $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ given $\mathscr{Y}_k$ under P. Write

$$p_k^{ij} := P(\mathbf{x}_k = \mathbf{e}_i, \mathbf{x}_{k-1} = \mathbf{e}_j|\mathscr{Y}_k) = E[\langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{e}_{ij}\rangle|\mathscr{Y}_k] \tag{9.4}$$

with $\mathbf{p}_k = (p_k^{11}, \ldots, p_k^{ij}, \ldots, p_k^{NN}) \in \mathbf{R}^{N^2}$. Assuming that $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ is an integrable sequence of random variables and using Bayes' theorem, we have

$$\mathbf{p}_k = E[\xi(\mathbf{x}_k, \mathbf{x}_{k-1})|\mathscr{Y}_k] = \frac{\bar{E}[\Lambda_k \xi(\mathbf{x}_k, \mathbf{x}_{k-1})|\mathscr{Y}_k]}{\bar{E}[\Lambda_k|\mathscr{Y}_k]}. \tag{9.5}$$

Letting $\mathbf{q}_k = \bar{E}[\Lambda_k \xi(\mathbf{x}_k, \mathbf{x}_{k-1})|\mathscr{Y}_k]$ and $\mathbf{1} = (1, \ldots, 1)^\top \in \mathbf{R}^{N^2}$, we see that

$$\sum_{i,j=1}^{N} \langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{e}_{ij}\rangle = \langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{1}\rangle = 1,$$

so that

$$\langle \mathbf{q}_k, \mathbf{1}\rangle = \bar{E}[\Lambda_k \langle \xi(\mathbf{x}_k, \mathbf{x}_{k-1}), \mathbf{1}\rangle|\mathscr{Y}_k] = \bar{E}[\Lambda_k|\mathscr{Y}_k]. \tag{9.6}$$

From (9.5) and (9.6), we get the explicit form of the conditional distribution as

$$\mathbf{p}_k = \frac{\mathbf{q}_k}{\langle \mathbf{q}_k, \mathbf{1} \rangle}. \tag{9.7}$$

Now, we need a recursive filter for the process $\mathbf{q}_k$ in order to estimate the state process $\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$. Define the diagonal matrix $\mathbf{B}_k$ by

$$\mathbf{B}_k = \begin{pmatrix} b_k^1 & & & & & & \\ & \ddots & & & & & \\ & & b_k^1 & & & & \\ & & & \ddots & & & \\ & & & & b_k^N & & \\ & & & & & \ddots & \\ & & & & & & b_k^N \end{pmatrix}, \tag{9.8}$$

where

$$b_k^i = \prod_{g=1}^{d} \frac{\phi((y_k^g - f_i^g)/\sigma_i^g)}{\sigma_i^g \phi(y_k^g)}. \tag{9.9}$$

Notation: For any $\mathscr{Y}_k$-adapted and integrable process X_k, write $\hat{X}_k := E[X_k|\mathscr{Y}_k]$ and $\gamma(X)_k := \bar{E}[\Lambda_k X_k|\mathscr{Y}_k]$. Again invoking Bayes' theorem, we have

$$\hat{X}_k = \frac{\gamma(X)_k}{\bar{E}[\Lambda_k|\mathscr{Y}_k]}. \tag{9.10}$$

To estimate the parameters of the model, recursive filters will be derived for several quantities of interest. For $r,s,t = 1,\ldots,N$, let J^{rst} denote the number of jumps from $(\mathbf{e}_s, \mathbf{e}_t)$ to state $\mathbf{e}_r$ up to time k, that is,

$$J_k^{rst} = \sum_{l=1}^{k} \langle \mathbf{x}_l, \mathbf{e}_r \rangle \langle \mathbf{x}_{l-1}, \mathbf{e}_s \rangle \langle \mathbf{x}_{l-2}, \mathbf{e}_t \rangle;$$

O_k^{rs} represents the occupation time of the WMC spent in state $(\mathbf{e}_r, \mathbf{e}_s)$ up to time k, that is,

$$O_k^{rs} = \sum_{l=1}^{k} \langle \mathbf{x}_{l-1}, \mathbf{e}_r \rangle \langle \mathbf{x}_{l-2}, \mathbf{e}_s \rangle;$$

O_k^r denotes the occupation time spent by the WMC in state $\mathbf{e}_r$ up to time k, that is,

$$O_k^r = \sum_{l=1}^{k} \langle \mathbf{x}_{l-1}, \mathbf{e}_r \rangle;$$

$T_k^r(c)$ is the level sum for the state $\mathbf{e}_r$, that is,

$$T_k^r(c) = \sum_{l=1}^{k} c(y_l)\langle \mathbf{x}_{l-1}, \mathbf{e}_r \rangle.$$

Here, c is a function with the form $c(y) = y$ or $c(y) = y^2$.

The above quantities are needed in the above four related processes are needed in the estimation of model parameters as illustrated in Proposition 9.2 below. We shall take advantage of the semi-martingale representation in (9.1) and best estimate of an adapted process X in (9.10) to obtain recursive formulae for the vector quantities $J_k^{rst}\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$, $O_k^{rs}\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$, $O_k^r\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$ and $T_k^r(c)\xi(\mathbf{x}_k, \mathbf{x}_{k-1})$. The recursive relation of these vector processes and $\mathbf{q}_k$ under a multi-dimensional observation set-up are given in the following proposition.

Proposition 9.1. *Let $\mathbf{V}_r, 1 \le r \le N$ be an $N^2 \times N^2$ matrix such that the $((i-1)N + r)$th column of $\mathbf{V}_r$ is $\mathbf{e}_{ir}$ for $i = 1 \ldots N$ and zero elsewhere. If $\mathbf{B}$ is the diagonal matrix defined in (9.8), then*

$$\mathbf{q}_{k+1} = \mathbf{B}_{k+1}\mathbf{\Pi}\mathbf{q}_k \tag{9.11}$$

and

$$\begin{aligned}
\gamma(J^{rst}\xi(\mathbf{x}_{k+1}, \mathbf{x}_k))_{k+1} &= \mathbf{B}_{k+1}\mathbf{\Pi}\gamma(J^{rst}\xi(\mathbf{x}_k, \mathbf{x}_{k-1}))_k \\
&\quad + b_{k+1}^r\langle \mathbf{q}_k, \mathbf{e}_{st}\rangle \pi_{rst}\mathbf{e}_{rs}, \qquad (9.12)\\
\gamma(O^{rs}\xi(\mathbf{x}_{k+1}, \mathbf{x}_k))_{k+1} &= \mathbf{B}_{k+1}\mathbf{\Pi}\gamma(O^{rs}\xi(\mathbf{x}_k, \mathbf{x}_{k-1}))_k \\
&\quad + \mathbf{B}_{k+1}\mathbf{\Pi}\mathbf{e}_{rs}\langle \mathbf{q}_k, \mathbf{e}_{rs}\rangle, \qquad (9.13)\\
\gamma(O^{r}\xi(\mathbf{x}_{k+1}, \mathbf{x}_k))_{k+1} &= \mathbf{B}_{k+1}\mathbf{\Pi}\gamma(O^{r}\xi(\mathbf{x}_k, \mathbf{x}_{k-1}))_k \\
&\quad + \mathbf{V}_r\mathbf{B}_{k+1}\mathbf{\Pi}\mathbf{q}_k, \qquad (9.14)\\
\gamma(T^{r}(c)\xi(\mathbf{x}_{k+1}, \mathbf{x}_k))_{k+1} &= \mathbf{B}_{k+1}\mathbf{\Pi}\gamma(T^{r}(c)\xi(\mathbf{x}_k, \mathbf{x}_{k-1}))_k \\
&\quad + c(y_{k+1}^c)\mathbf{V}_r\mathbf{B}_{k+1}\mathbf{\Pi}\mathbf{q}_k, \qquad (9.15)
\end{aligned}$$

for $1 \le g \le d$.

Proof. See [27] for an analogous proof for each of the filters under the single observation setting. □

Similar to (9.7), we determine the normalized filter estimates of $\gamma(J^{rst})_k$, $\gamma(O^{rs})_k$, $\gamma(O^r)_k$ and $\gamma(T^r(c))_k$ by summing the components of the vector expressions given in (9.12)–(9.15).

We adopt the EM algorithm to estimate the optimal parameters. The calculation is similar to the technique for the single observation set-up. The estimates are expressed in terms of the recursions provided in (9.12)–(9.15) and given in the following proposition.

Proposition 9.2. *Suppose the observation is d-dimensional and the set of parameters $\{\hat{a}_{rst}, \hat{f}_r^g, \hat{\sigma}_r^g\}$ determines the dynamics of y_k^g, $k \ge 1$, $1 \le g \le d$. Then the EM estimates for these parameters are given by*

$$\hat{a}_{rst} = \frac{\widehat{J}_k^{rst}}{\widehat{O}_k^{st}} = \frac{\gamma(J^{rst})_k}{\gamma(O^{st})_k}, \ \forall pairs\,(r,s),\ r \neq s, \tag{9.16}$$

$$\hat{f}_r^g = \frac{\widehat{T}_k^r}{\widehat{O}_k^r} = \frac{\gamma(T^r(y^g))_k}{\gamma(O^r)_k}, \tag{9.17}$$

$$\hat{\sigma}_r^g = \sqrt{\frac{\widehat{T}^r((y^g)^2)_k - 2\hat{f}_r^g \widehat{T}^r(y^g)_k + (\hat{f}_r^g)^2 \widehat{O}_k^r}{\widehat{O}_k^r}}. \tag{9.18}$$

Proof. See [27] for an analogous proof of each estimate under the single observation setting. □

Given the observation up to time k, new parameters $\hat{a}_{rst}(k)$, $\hat{f}_r^g(k)$, $\hat{\sigma}_r^g(k)$, $1 \leq r,s,t \leq N$ are provided by (9.16)–(9.18). The recursive filters for the unobserved Markov chain and related processes in Proposition 9.1 can easily get updated every time new information arrives. Thus, the parameter estimation is self-calibrating.

9.3 Implementation

We implement the recursive filters derived in the previous section on yields of 3-month and 6-month US T-bills, 1- and 5-year US T-notes, and 20- and 30-year US bonds. The data set of yields, compiled by the Bank of Canada, contains 718 daily vector observations from 22 December 2008 to 31 October 2011. The evolution of yields underwent several regimes as evidenced by the changes in parameter values and the summary descriptive statistics (see Table 9.1) indicating that data are coming from a distribution with heavy tails relative to the normal distribution. In particular, we see that the values of excess kurtosis for the yield curves are higher than those from a normal distribution. Regime-switching models are designed to capture this type of behaviour of the data. Tables 9.1 and 9.2 display possible segregations of the actual data into either two or three states based on mean and volatility levels. This preliminary analysis of the actual data reveals that the yield volatilities are related to mean and maturity. When maturity is short, yield volatilities are higher with relatively high means; when maturity is long, yield volatilities are higher with lower means. The data set on yield values is a six-dimensional observation process, whose dynamics are given by

$$y_{k+1}^i = f^i(\mathbf{x}_k) + \sigma^i(\mathbf{x}_k) z_{k+1}^i,\ i = 1,\ldots,6,$$

where $\mathbf{f}^i = (f_1^i,\ldots,f_N^i) \in \mathbf{R}^N$ and $\sigma^i = (\sigma_1^i,\ldots,\sigma_N^i) \in \mathbf{R}^N$ are governed by the WHMM $\mathbf{x}$. The implementation procedure starts with choosing the initial values for $\mathbf{f}^i$ and σ^i, $i = 1,\ldots,6$. All non-zero entries in the transition matrix Π are set to $1/N$. The data are processed in 71 batches, and a batch consists of ten yield vectors. Each algorithm run through a batch of data is termed as a complete algorithm step. At the end of each step, new estimates for $\mathbf{f}$, σ and $\mathbf{A}$ are computed. From

Table 9.1 Descriptive summary statistics and data segregation into two states

	Overall			Dec/08–July/10		August/10–Nov/11	
Maturity	Mean	STD	Ex. Kurtosis	Mean	STD	Mean	STD
3-month	0.13	0.06	0.57	0.14	0.07	0.12	0.04
6-month	0.23	0.09	1.97	0.26	0.10	0.20	0.05
1-year	0.43	0.16	0.22	0.45	0.12	0.41	0.20
5-year	3.39	0.57	0.70	3.33	0.42	3.45	0.71
20-year	4.07	0.46	7.88	4.14	0.38	3.98	0.53
30-year	4.12	0.50	6.65	4.15	0.50	4.09	0.48

Table 9.2 Segregation of data into three states

	Dec/08–Feb/09		March/09–April/11		May/11–Nov/11	
Maturity	Mean	STD	Mean	STD	Mean	STD
3-month	0.17	0.07	0.13	0.04	0.09	0.04
6-month	0.34	0.08	0.20	0.04	0.18	0.06
1-year	0.54	0.11	0.38	0.14	0.44	0.19
5-year	3.11	0.47	3.41	0.57	3.71	0.53
20-year	3.97	0.43	4.14	0.44	4.01	0.50
30-year	3.86	0.55	4.30	0.38	4.01	0.50

the estimates of $\mathbf{A}$, we construct $\mathbf{\Pi}$. These new estimates are in turn used as initial parameter values in the succeeding batch data processing that employs the recursive filter equations. This self-turning algorithm allows a forthnightly update of the parameters. Figure 9.1 exhibits the plots for the evolution of the transition probabilities under the two-state setting. The plot in the top panel shows the probabilities of staying in the same regime as the previous step. The plot in the bottom panel shows the probabilities of switching to a different state from the previous step. Except for some jumps in the probability values around the 50th algorithm pass, the bond market is quite stable as demonstrated by the relatively smooth evolution of probabilities. The large changes correspond to yield fluctuations over a brief period of time, e.g., T-bill rate increases from 0.29 on 31 December 2010 to 0.61 on 03 January 2011, and T-bond rate increases from 4.13 on 31 December 31 2010 to 4.55 on 05 January 2011. These significant changes during a short time span constitute evidence of states switching that could be captured by WHMM. Additionally, these market changes are reflected in the dynamics of parameter estimates. Figures 9.2 and 9.3 show plots depicting the movement through time of the optimal parameter estimates for each yield vector under the two-state WHMM set-up. Furthermore, the values of $\mathbf{f}$ and σ are positively correlated with the yield maturity, i.e., the longer the maturity the higher the mean and volatility levels. The evolution of parameters for 1-, 5- and 20-year yields support our preliminary analysis that the 1- and 5-year yields have states characterized by high (low) means and high (low) volatilities. The 20-year yield series, however, has states characterized by low (high) means and high (low) volatilities. Such consistent behavioural mean-volatility relationship patterns are not necessarily present for yields of instruments that have either very short or very long maturities. It is worth mentioning that parameters appear to stablize after approximately seven steps through this online algorithm. The same patterns

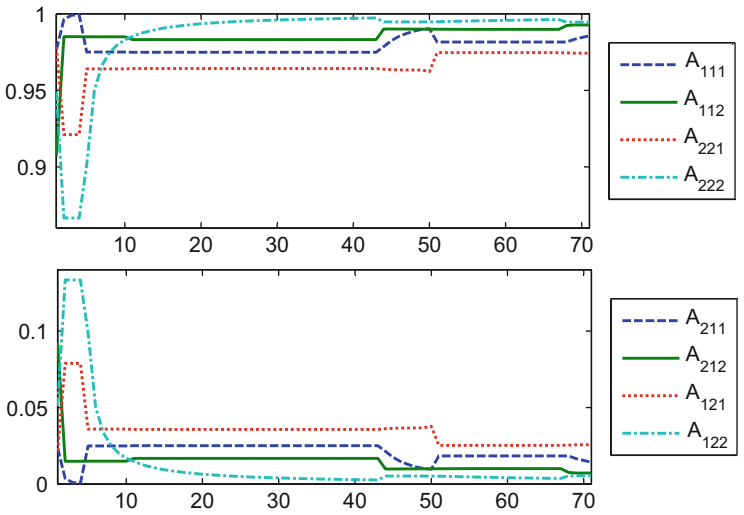

Fig. 9.1 Evolution of estimates for transition probabilities through algorithm steps under the two-state setting

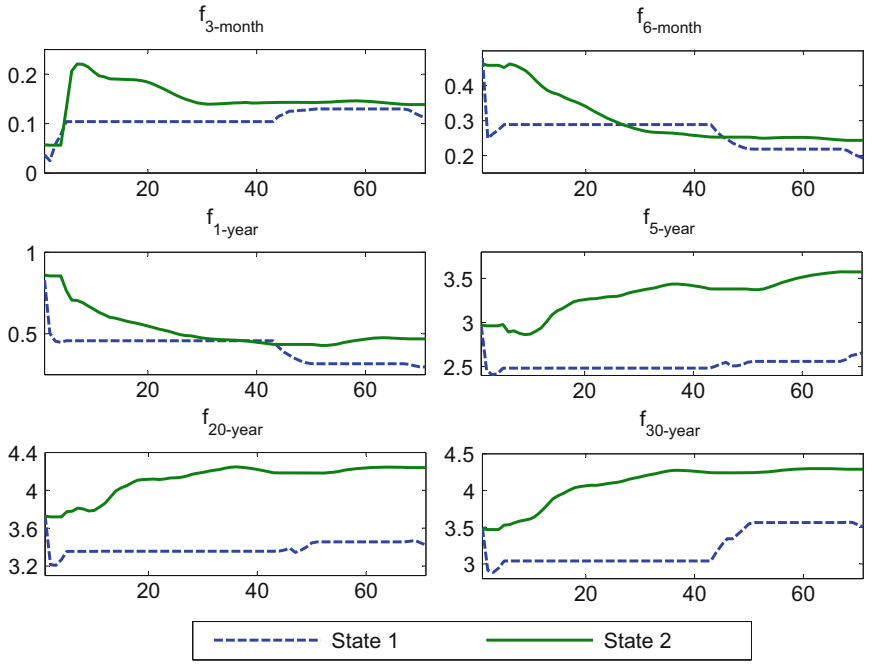

Fig. 9.2 Evolution of estimates for f through algorithm steps under the two-state setting

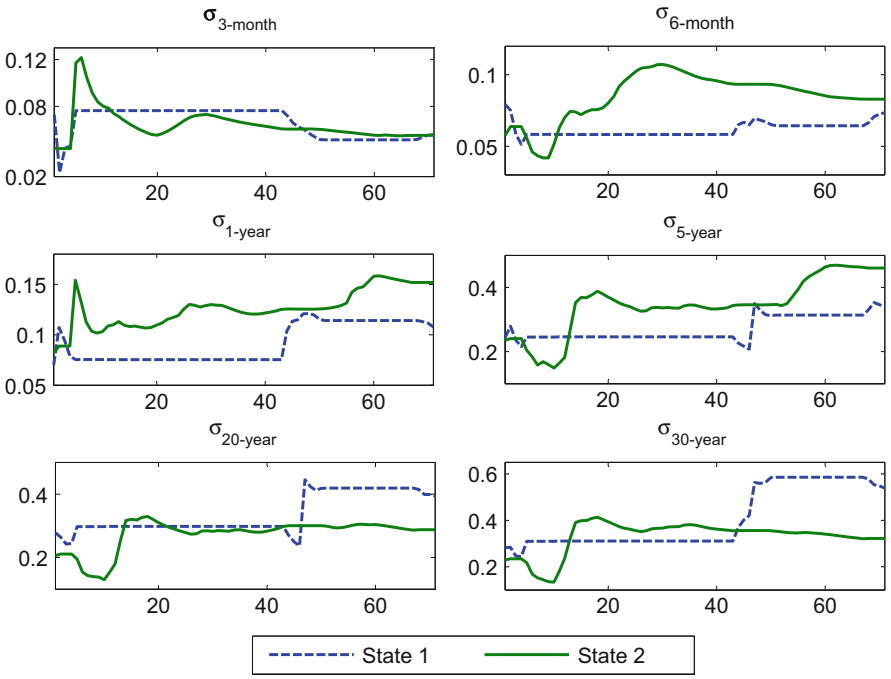

Fig. 9.3 Evolution of estimates for σ through algorithm steps under the two-state setting

Table 9.3 Parameter estimates at the end of final algorithm step for $N = 3$

Final estimation:

A matrix: $\begin{pmatrix} 0.818 & 0.848 & 0.836 & 0.029 & 0.000 & 0.005 & 0.041 & 0.007 & 0.010 \\ 0.091 & 0.076 & 0.082 & 0.942 & 1.000 & 0.989 & 0.041 & 0.008 & 0.010 \\ 0.091 & 0.076 & 0.082 & 0.029 & 0.000 & 0.006 & 0.918 & 0.985 & 0.980 \end{pmatrix}$

f matrix: $\begin{pmatrix} 0.06 & 0.13 & 0.07 \\ 0.53 & 0.23 & 0.50 \\ 0.87 & 0.43 & 0.91 \\ 3.04 & 3.39 & 3.12 \\ 3.79 & 4.08 & 3.85 \\ 3.56 & 4.14 & 3.62 \end{pmatrix}$ σ matrix: $\begin{pmatrix} 0.16 & 0.06 & 0.07 \\ 0.08 & 0.08 & 0.05 \\ 0.02 & 0.16 & 0.05 \\ 0.35 & 0.57 & 0.20 \\ 0.48 & 0.45 & 0.21 \\ 0.44 & 0.49 & 0.21 \end{pmatrix}$

are produced regardless of the choice of the initial values. The choice of the initial parameter values though can affect the speed of convergence. For the three-state setting, we report the final estimates of $\mathbf{A}$, $\mathbf{f}$ and σ after the final algorithm step in Table 9.3.

9.4 Forecasting and Error Analysis

In this section, we shall use the model parameter estimates to forecast yield values covering an h-day ahead horizon. The semi-martingale representation of $\mathbf{x}$ in (9.1) leads to

$$E[\xi(\mathbf{x}_{k+1},\mathbf{x}_k)|\mathscr{Y}_k] = \mathbf{\Pi}\xi(\mathbf{x}_k,\mathbf{x}_{k-1}) = \mathbf{\Pi}\mathbf{p}_k. \tag{9.19}$$

Furthermore, we have

$$E[\xi(\mathbf{x}_{k+h},\mathbf{x}_{k+h-1})|\mathscr{Y}_k] = \mathbf{\Pi}^h\mathbf{p}_k, \text{ for } h = 1,2,\ldots \tag{9.20}$$

Recall that $\mathbf{\Pi}$ is constructed from $\mathbf{A}$, which is defined by

$$a_{lmv} = P(\mathbf{x}_{k+1} = \mathbf{e}_l|\mathbf{x}_k = \mathbf{e}_m, \mathbf{x}_{k-1} = \mathbf{e}_v),$$

so that (9.19) gives

$$E[\mathbf{x}_{k+1}|\mathscr{Y}_k] = \mathbf{A}\mathbf{p}_k. \tag{9.21}$$

Hence, from (9.20) and (9.21),

$$E[\mathbf{x}_{k+h}|\mathscr{Y}_k] = \mathbf{A}\mathbf{p}_{k+h-1} = \mathbf{A}\mathbf{\Pi}^{h-1}\mathbf{p}_k. \tag{9.22}$$

Using (9.22), the best estimate of the h-step ahead predicted yields y^i_{k+h} given available information at time k is

$$\hat{y}^i_{k+h} = E[y^i_{k+h}|\mathscr{Y}_k] = \langle \mathbf{f}^i, \mathbf{A}\mathbf{\Pi}^{h-1}\mathbf{p}_k\rangle, \text{ for } 1 \le i \le d. \tag{9.23}$$

The conditional variance for the predicted yields are calculated using

$$\begin{aligned}\text{Var}[y^i_{k+h}|\mathscr{Y}_k] = {} & (\mathbf{f}^i)^\top \text{diag}(\mathbf{A}\mathbf{\Pi}^{h-1}\mathbf{p}_k)\mathbf{f}^i + (\sigma^i)^\top \text{diag}(\mathbf{A}\Pi^{h-1}\mathbf{p}_k)\sigma^i \\ & -\langle \mathbf{f}^i, \mathbf{A}\Pi^{h-1}\mathbf{p}_k\rangle^2,\end{aligned} \tag{9.24}$$

where $\text{diag}(\mathbf{A}\mathbf{\Pi}^{h-1}\mathbf{p}_k)$ is a diagonal matrix whose diagonal entries are the components of the vector $(\mathbf{A}\mathbf{\Pi}^{h-1}\mathbf{p}_k)$.

The determination of the optimal number of states given a particular data set is an important statistical inference problem. Hardy [31] and Erlwein and Mamon [10] applied the Akaike information criterion (AIC) to determine the optimal number of regimes in HMM-based models. The AIC is a measure of the relative goodness of fit of a statistical model. It offers a relative measure of lost information described by the trade-off between bias and variance in the model construction. The AIC is calculated as

$$\text{AIC} = 2s - 2\log(\mathscr{L}(\theta)),$$

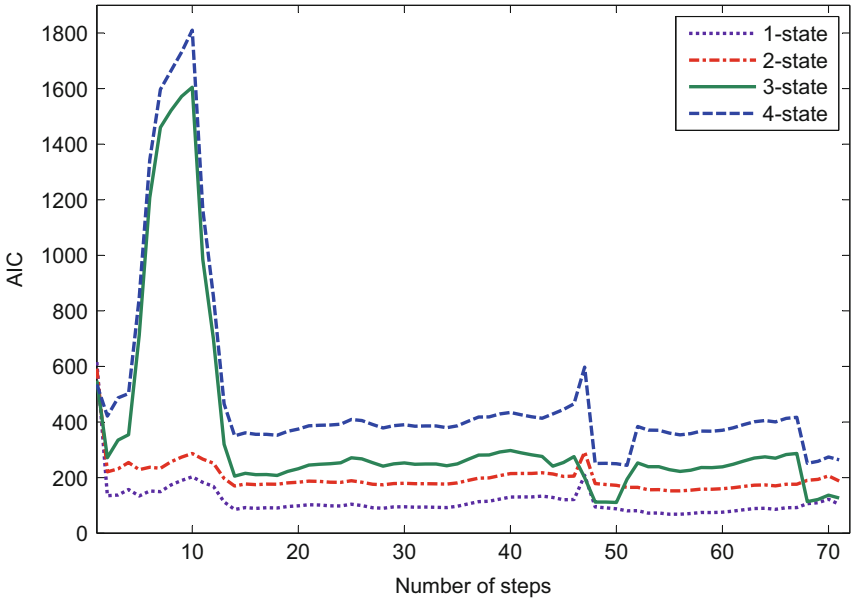

Fig. 9.4 AIC for the one-, two-, three- and four-state models

where s is the number of parameters and $\mathscr{L}(\theta)$ denotes the likelihood function of the model. The preferred model is the one that gives the minimal AIC value. For the vector observation process $\mathbf{y}_k$ in each pass, the log-likelihood of the parameter set θ is given by

$$\mathscr{L}(\theta) = \sum_{l=1}^{\#\text{ in batch}} \sum_{i=1}^{d} \left(-\frac{1}{2} \log\left(2\pi\sigma^i(\mathbf{x}_{l-1})^2\right) - \frac{\left(y_l^i - f^i(\mathbf{x}_{l-1})\right)^2}{2\sigma^i(\mathbf{x}_{l-1})^2} \right). \tag{9.25}$$

The calculated AIC values for the one-, two-, three- and four-state models after each algorithm step are presented in Fig. 9.4. Both one- and two-state models are reasonable in capturing the dynamics of our data gauging from this criterion with the one-state model producing the smallest AIC values. The results indicate that both one- and two-state models perform better than the three- and four-state models. The model we proposed requires the estimation of $(N^2-1)N+2mN$ parameters, where m is the number of securities. The number of needed estimations increases rapidly as N increases leading to higher AIC associated. The AIC, however, cannot measure how well a model fits the actual time series data. In order to assess the goodness of fit of the one-step ahead forecasts, we evaluate the root mean square error (RMSE) for the one-, two-, three- and four-state WHMM-based term structure models. The results of this error analysis are given in Table 9.4.

Clearly, the two-state model outperforms the model with no switching in terms of lower forecasting errors. The large improvement in the error indicates that the models with regime switching can generate better price forecasts. The comparison

Table 9.4 RMSE for one-step ahead predictions versus actual values

State setting	3-month	6-month	1-year	5-year	20-year	30-year
1	0.0558	0.0864	0.1631	0.4923	0.4363	0.4732
2	0.0539	0.0821	0.1458	0.3418	0.2994	0.3590
3	0.0558	0.0854	0.1619	0.4921	0.4357	0.4722
4	0.0524	0.0806	0.1426	0.3656	0.3194	0.3710

of error measure also shows that the four-state model is able to forecast the short maturity yields better than the two-state model. However, the improvement is not significant. Since a larger number of state increases the complexity of parameter estimation, a two-state model is sufficient to model the yield values.

Figure 9.5 exhibits the actual yields and one-step ahead forecasted yields for the 3- and 6-month T-bills, 1- and 5-year T-notes and 20- and 30-year T-bonds. The 99% confidence interval for the predicted yields is also displayed and was calculated using $E[y^i_{k+h}|\mathscr{Y}_k] \pm 2.575\sqrt{\text{Var}[y^i_{k+h}|\mathscr{Y}_k]}$. The resulting forecasts follow the actual data quite well. Empirical results confirm that the WHMM can capture most of the market dynamics.

In [27], the forecasting performance of the one-dimensional WHMM is compared with that of the regular HMM using the data set on S&P500 prices. The results suggest that the WHMM outperforms the HMM over a long forecasting horizon. In this empirical implementation, we also evaluate the goodness of fit of the h-day ahead forecasts using RMSE and absolute percentage error (APE) as benchmarks. The multi-dimensional WHMM-based term structure model is compared to the regular multi-dimensional HMM model using these two criteria. The RMSE for an h-day ahead prediction of y^i, $i = 1, \ldots, d$ is given by

$$\text{RMSE}(i,h) = \sqrt{\frac{1}{\text{M}-h}\sum_{k=1}^{\text{M}-h}(y^i_{k+h} - \hat{y}^i_{k+h})^2},$$

where M is the time horizon. Similarly, the APE for an h-day ahead prediction of y^i is defined by

$$\text{APE}(i,h) = \frac{1}{\text{M}-h}\sum_{k=1}^{\text{M}-h}\left|\frac{y^i_{k+h} - \hat{y}^i_{k+h}}{y^i_{k+h}}\right|.$$

Following the idea in Date et al. [32], we calculate the average RMSE (AvRMSE) and average APE (AvAPE) over six yields to measure the prediction performance. AvRMSE(h) is the average of RMSE(i,h) over yield values with different maturities, i.e.,

$$\text{AvRMSE}(h) = \frac{1}{d}\sum_{i=1}^{d}\text{RMSE}(i,h).$$

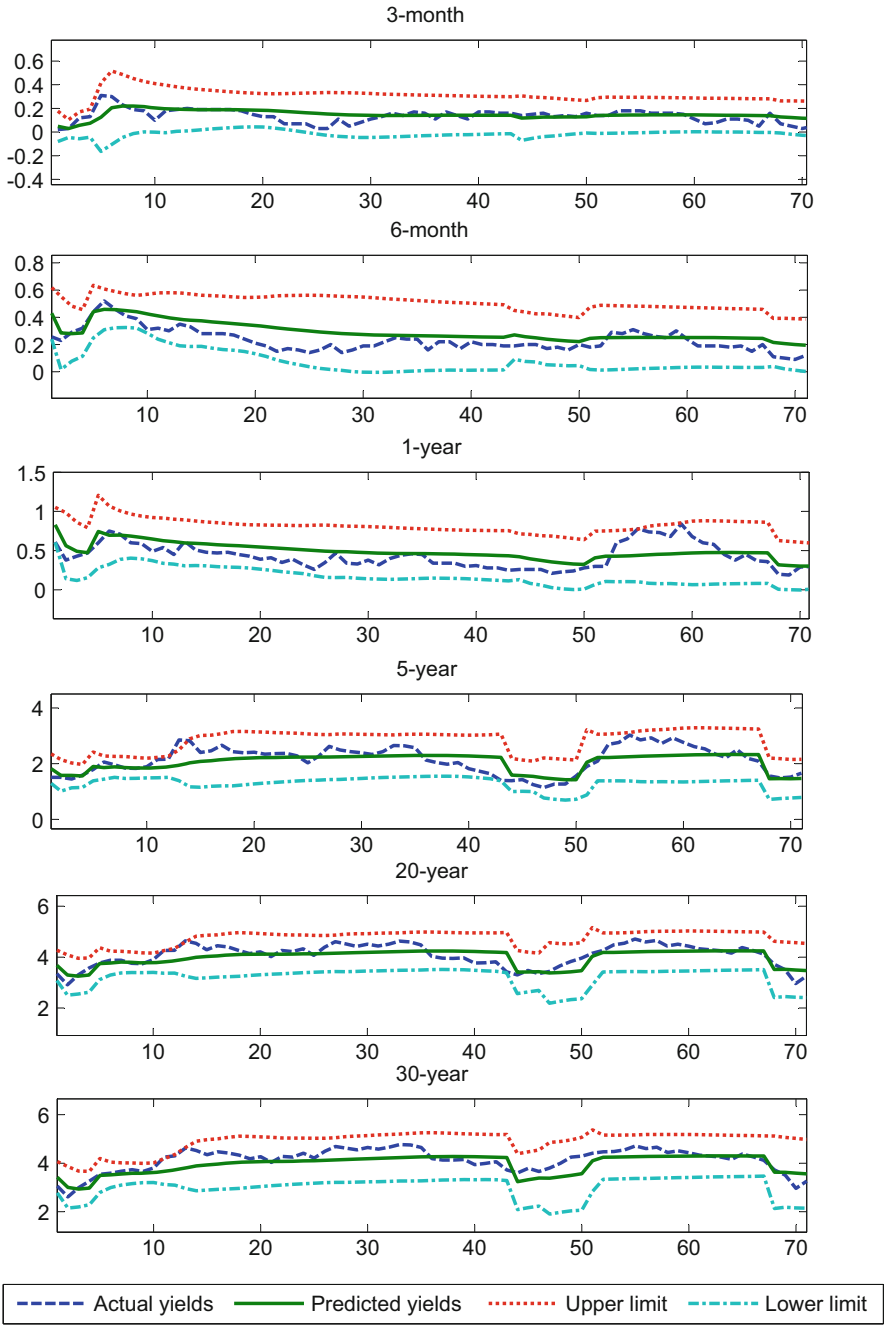

Fig. 9.5 One-step ahead predicted values (%) versus actual Treasury yields (%) under a two-state WHMM setting

AvAPE(h) denotes the average of APE(i,h) over yield values with different maturities, i.e.,

$$\text{AvAPE}(h) = \frac{1}{d}\sum_{i=1}^{d}\text{APE}(i,h).$$

These error analyses are displayed in Tables 9.5–9.8. In Table 9.5, the one-state WHMM coincides with the one-state HMM. Under the WHMM framework, memory is a property of the underlying market state process. A one-state Markov chain stays in only one state throughout the progression of time. That is, there is no memory of visiting other states in the previous steps. This is why WHMM collapses to the HMM set-up under the one-state setting. The two-state WHMM gives a better fit than the HMM in terms of lower forecasting errors with respect to both metrics. The differences of errors between the three-state WHMM and HMM models, shown in Table 9.7, are too small to make any practical significance. The four-state WHMM seems to outperform the regular HMM in the long-horizon forecasting under the RMSE but not for the APE metric.

Table 9.5 Error analysis of WHMM and HMM models under the one-state setting

	h-Day ahead						
	1	2	3	4	5	6	7
AvRMSE$_h$ of WHMM/HMM	0.2845	0.2862	0.2825	0.2883	0.2842	0.2828	0.2943
AvAPE$_h$ of WHMM/HMM	0.2843	0.2687	0.2690	0.2722	0.2817	0.3073	0.3018

Table 9.6 Error analysis of WHMM and HMM models under the two-state setting

	h-Day ahead						
	1	2	3	4	5	6	7
AvRMSE$_h$ of WHMM	0.2137	0.2175	0.2173	0.2263	0.2241	0.2812	0.2351
AvRMSE$_h$ of HMM	0.2738	0.2773	0.2742	0.2815	0.2794	0.3054	0.2897
AvAPE$_h$ of WHMM	0.2465	0.2352	0.2393	0.2436	0.2546	0.2812	0.2769
AvAPE$_h$ of HMM	0.2789	0.2643	0.2656	0.2694	0.2794	0.3054	0.2998

Table 9.7 Error analysis of WHMM and HMM models under the three-state setting

	h-Day ahead						
	1	2	3	4	5	6	7
AvRMSE$_h$ of WHMM	0.2839	0.2857	0.2819	0.2877	0.2838	0.2725	0.2851
AvRMSE$_h$ of HMM	0.2855	0.2841	0.2844	0.2864	0.2866	0.2328	0.2451
AvAPE$_h$ of WHMM	0.2828	0.2673	0.2676	0.2707	0.2804	0.2782	0.2763
AvAPE$_h$ of HMM	0.2855	0.2867	0.2670	0.2773	0.2875	0.2849	0.2805

Table 9.8 Error analysis of WHMM and HMM models under the four-state setting

	h-Day ahead						
	1	2	3	4	5	6	7
AvRMSE$_h$ of WHMM	0.2219	0.2252	0.2231	0.2320	0.2318	0.2307	0.2399
AvRMSE$_h$ of HMM	0.2098	0.2057	0.2182	0.2291	0.2368	0.2437	0.2514
AvAPE$_h$ of WHMM	0.2485	0.2364	0.2388	0.2416	0.2544	0.2805	0.2775
AvAPE$_h$ of HMM	0.2086	0.2113	0.2129	0.2172	0.2218	0.2346	0.2406

9.5 Conclusion

In this paper, we put forward a multivariate WHMM-driven term structure model where the means and volatilities of vector observations are governed by a second-order Markov chain in discrete time. The proposed model is tested on time series data of yields covering 3- and 6-month US T-bills, 1- and 5-year US T-notes and 20- and 30-year US T-bonds. A multivariate filtering technique along with the EM algorithm was employed in the optimal estimation of parameters. The algorithms were run in batches and parameters are updated when new information arrives thereby making the model self-tuning. The empirical results of the implementation of filters and parameter estimation demonstrate the adequacy of the proposed model in capturing market dynamics and regime changes in the data. We applied the Akaike information criterion to determine the optimal number of regimes and assessed the goodness of fit of the one-step ahead forecasts generated by the one-, two-, three- and four-state models. We found that within the data set examined, a two-state model is deemed sufficient to capture the term structure dynamics. An analysis of the h-day ahead predictions was also presented and results from WHMM were compared with those from the regular HMM. This paper manifests that WHMM outperforms the HMM in terms of low forecasting errors and we attribute such improved performance to building a model that takes into account both regime switching and memories in the data series.

References

1. Merton, R.C.: Theory of rational option pricing. Bell Journal of Economics. **4**, 141–183 (1973).
2. Vasicek, O.: An equilibrium characterisation of the term structure. Journal of Financial Economics. **5**, 177–188 (1977).
3. Cox, J.C., Ingersoll, J.E., Ross, S.A.: A theory of the term structure of interest rates. Econometrica. **53**, 385–407 (1985).
4. Hull, J., White, A.: Pricing interest-rate derivative securities. Review of Financial Studies. Stud. **3**, 573–592 (1990).
5. Hamilton, J.D.: A new approach to the economic analysis of nonstationary time series and business cycle. Econometrica. **57**, 357–384 (1989).
6. Smith, D.: Markov-switching and stochastic volatility diffusion models of short-term interest rates. Journal of Business and Economic Statistics. **20**, 183–197 (2002).
7. Landén, C.: Bond pricing in a hidden Markov model of the short rate. Finance Stochastics. **4**, 371–389 (2000).

8. Elliott, R.J., Aggoun, L., Moore, J.B.: Hidden Markov Models: Estimation and Control. Springer, New York (1995).
9. Elliott, R., Hunter, W., Jamieson, B.: Financial signal processing: a self calibrating model. International Journal of Theoretical and Applied Finance. **4**, 567–584 (2001).
10. Erlwein, C., Mamon, R.S.: An online estimation scheme for a Hull-White model with HMM-driven parameters. Statistical Methods and Applications. **18**, 87–107 (2009).
11. Hunt, J., Devolder, P.: Semi-Markov regime switching interest rate models and minimal entropy measure. Physica A. **390**: 3767–3781 (2011).
12. Zhou, N., Mamon, R.: An accessible implementation of interest rate models with Markov-switching. Expert Systems with Applications. **39**: 4679–4689 (2012).
13. Guidolin, M., Timmermann, A.:Forecasts of US short-term interest rates: A flexible forecast combination approach. Journal of Econometrics. **150**: 297–311 (2009).
14. Meligkotsidou, L., Dellaportas, P.: Forecasting with non-homogeneous hidden Markov models. Statistics and Computing. **21**: 439–449 (2011).
15. Startz, R., Tsang, K.P.: An unobserved components model of the yield curve. Journal of Money, Credit and Banking, **42**: 1613–1640 (2010).
16. Audrino, F., Mederos, M.C.: Modeling and forecasting short-term interest rates: The benefits of smooth regimes, macroeconomic variables, and bagging. Journal of Applied Econometrics. **26**: 999–1022 (2011).
17. Holmes, M.J, Dutu, R., Cui, X.: Real interest rates, inflation and the open economy: A regime-switching perspective on Australia and New Zealand. International Review of Economics and Finance. **18**: 351–360 (2009).
18. Nieh, C-C., Wu, S., Zeng, Y.: Regime shifts and the term structure of interest rates. In: Lee, C-F., Lee, J.: Handbook of Quantitative Finance and Risk Management, pp.1121–1134, Springer, (2010).
19. Mandelbrot, B.B: When can price be arbitraged efficiently? A limit to the validity of the random walk and martingale models. Review of Economics and Statistics. **53**, 225–236 (1971).
20. Cajueiro, D.O., Tabak, B.M.: Long-range dependence and multifractality in the term structure of LIBOR interest rates. Physica A. **373**, 603–614 (2007).
21. Cajueiro, D.O., Tabak, B.M.: Time-varying long-range dependence in US interest rates. Chaos, Solitons & Fractals. **34**, 360–367 (2007).
22. McCarthy, J., Pantalone, C., Li, H.C: Investigating long memory in yield spreads. Journal of Fixed Income. **19**, 73–81 (2009).
23. Maheu, J.: Can GARCH models capture long-range dependence? Studies in Nonlinear Dynamics. **9**, Article 1 (2005).
24. Dajcman, S.: Time-varying long-range dependence in stock market returns and financial market disruptions - a case of eight European countries. Applied Economics Letters. **19**, 953–957 (2012).
25. Duan, J-C., Jacobs, K.: A simple long-memory equilibrium interest rate model. Economics Letters. **53**: 317–321 (1996).
26. Solberg, J.: Modelling Random Processes for Engineers and Managers. Wiley & Sons, Inc., New Jersey (2009).
27. Xi, X., Mamon, R.: Parameter estimation of an asset price model driven by a weak hidden Markov chain. Economic Modelling. **28**, 36–46 (2011).
28. Siu, T.K., Ching, W.K., Fung, E., Ng, M., Li, X.: A high-order Markov-switching model for risk measurement. Computers & Mathematics with Applications. **58**, 1–10 (2009).
29. Ching, W. K., Siu, T. K., Li,L. M.: Pricing exotic options under a higher-order hidden Markov model. Journal of Applied Mathematics & Decision Sciences. v 18014: 1–15 (2007).
30. Siu, T.K., Ching, W.K., Fung, E.: Extracting information from spot interest rates and credit ratings using double higher-order hidden Markov models. Computational Economics. **26**, 251–284 (2005).
31. Hardy, M.: A regime-switching model of long-term stock returns. North American Actuarial Journal. **5**, 41–53 (2001).
32. Date, P., Jalen, L., Mamon, R.: A partially linearised sigma point filter for latent state estimation in nonlinear time series models. Journal of Computational and Applied Mathematics. **233**: 2675–2682 (2010).

Chapter 10
Numerical Methods for Optimal Annuity Purchasing and Dividend Optimization Strategies under Regime-Switching Models: Review of Recent Results

Zhuo Jin and George Yin

10.1 Introduction

In actuarial science, many stochastic control problems such as designing optimal risk controls, developing dividend payment policies, and hedging for contingent claims. for an insurance corporation have drawn increasing attention. The actual applications may vary, the problems of interest all involve stochastic processes in the problems that depend on controls. We are interested in determining the maximal value which is the so-called value function and the corresponding optimal control strategies. To obtain the optimal control and the value function, one basic approach is to use the dynamic programming methods to solve Hamilton–Jacobi–Bellman (HJB) equations. A comprehensive study of stochastic control problems in insurance can be found in [18].

Empirical studies indicate that traditional stochastic models often fail to capture more extreme movements. To better reflect the reality, one of the recent trends is to use regime-switching models. Hamilton (1989) in [9] introduced a regime-switching time series in the study of data analysis and time series. Recent work on regime-switching models and related issues in insurance applications can be found in [19, 22]. The models contain both continuous dynamics and discrete events and are more versatile. The discrete event describes, for example, market behaviors and other economics effects that cannot be modeled as a differential equation. The models that we are interested in this chapter can be thought of as a number of controlled diffusion processes modulated by a random switching device. They can be treated as two-component processes, in which one component delineates the diffusion behav-

Z. Jin
Centre for Actuarial Studies, Department of Economics, The University of Melbourne, Parkville, VIC 3010, Australia
e-mail: zjin@unimelb.edu.au

G. Yin (✉)
Department of Mathematics, Wayne State University, Detroit, MI 48202, USA
e-mail: gyin@math.wayne.edu

Y. Zeng and S. Wu (eds.), *State-Space Models: Applications in Economics and Finance*, Statistics and Econometrics for Finance 1, DOI 10.1007/978-1-4614-7789-1_10,

ior, and the other component describes the discrete events involved. For example, the market modes can be formulated as a finite-state Markov chain that takes values 1 and 2. We use 1 to represent the bullish (up-trend) market and 2 the bearish (down-trend) market. The switching between 1 and 2 describes the market changes. In this chapter, we assume that the Markov chain has a finite state space. For different discrete states, the rates of yield and volatility are different.

With the traditional dynamic programming approach, the HJB equations can be obtained. It would be ideal to find the explicit solutions under suitable assumptions. However, the underlying systems are normally highly nonlinear. In addition, for the regime-switching model, a coupled system of HJB equations instead of one HJB equation needs to be solved due to the Markov switching. To obtain closed-form solutions becomes very difficult. Furthermore, a large class of dividend payment problems will involve singular and impulse controls. The optimal value function will satisfy a coupled system of quasi-variational inequalities. It is virtually impossible to solve them analytically. Thus, a numerical approach for solving such problems is a viable alternative.

In this study, we survey some recent progress on numerical methods for stochastic control problems arising in insurance risk management. The basic models involve regime switching that delineates the coexistence of continuous dynamics and discrete events. In this chapter, we review our results obtained in [10] and [11]. Additional numerical examples different from that in the aforementioned papers are provided for demonstration. We extend the Markov chain approximation methodology developed by Kushner and Dupius (2001) in [13] to the regime-switching models in the problems of actuarial science and finance. In what follows, a generalized formulation of hybrid controlled diffusion model will be presented at first. Depending on if we are dealing with regular control or singular control, the dynamic programming principle will yield a coupled system of HJB equations or QVIs (if singular control is contained). To deal with the numerical algorithm of Markov chain method for solving the system of HJB equations or QVIs, we will construct a two-component discrete-time controlled Markov chain to approximate the diffusion process and the Markov switching term. With the piecewise constant interpolation and upwind discretization, the corresponding dynamic programming equation and transition probabilities will be obtained. To guarantee that the approximating Markov chain is in good alignment with the original diffusion process, which is technically known as consistency of the numerical approximation. The precise definition of local consistency will be introduced and verified. Under simple conditions, the convergence of the approximation sequence to the diffusion process and the convergence of the approximation to the value function will be confirmed. In the actual computation, we simply use the well-known policy iteration (or policy improvement) or value iteration method for implementation. It is also worth mentioning that the Markov chain approximation method requires little regularity of the value function and/or analytic properties of the associated systems of HJB equations and/or QVIs. In reality, for a great many cases, the actual analytic properties of the solutions are rarely known.

In this work, we examine optimal annuity purchasing strategies and dividend optimization problems. We concentrate on the numerical treatments of such problems.

The rest of the chapter is organized as follows. The optimal annuity-purchasing problem is considered in Sect. 10.2. The controlled hybrid wealth models are presented in Sect. 10.2.2. The constant hazard rate and the corresponding optimal control are considered in Sect. 10.2.3. Section 10.2.4 presents the more general hazard rate models and the associated numerical algorithms for finding the optimal strategies. Section 10.2.5 presents two examples with constant hazard rate and Gompertz hazard rate, respectively. Section 10.3 considers the problem of optimal dividend payment policies. The formulation and assumptions are presented in Sect. 10.3.2. It is also shown that the value functions are twice continuously differentiable with respect to the continuous component and satisfy the system of HJB equations. Section 10.3.3 deals with the Markov chain approximation method. The approximating Markov chain and the algorithm for solving the dynamic programming equation are presented. Section 10.3.4 deals with the convergence of the approximation scheme. Numerical examples are provided in Sect. 10.3.5 to illustrate the performance of the approximation procedure. In this chapter, we only present the conditions needed and the corresponding results, whereas the verbatim proofs of the results can be found in [10] and [11]. Finally, some additional remarks are provided in Sect. 10.4.

10.2 Optimal Annuity-Purchasing Strategies

10.2.1 Motivation

Due to the collapse of the housing market and the financial crisis, trillions of dollars have been lost. The retirement security may be one of the greatest casualties of the recent financial crisis. More and more people are concerned with their future financial safety after retirement. How to avoid financial ruin becomes a pressing issue. To secure the post-retirement life, purchasing annuities from insurance companies is of crucial importance. The recipient of the annuity could receive a continuous fixed payment throughout the life. This life stream income could guarantee the retiree a given level of consumption. On the other hand, since the Swedish actuary Filip Lundberg introduced the classical compound-Poisson risk model in 1903, probability of ruin has been among the prime quantities to measure the insurance risk. Therefore, to measure the financial risk of purchasing annuity and managing portfolio becomes a big issue.

According to the Transamerica Center for Retirement Studies, the number of U.S. workers who are confident in their ability to retire comfortably has declined significantly in the past year. Recently, Vanderhei and Copeland (2003) in [20] reported that American retirees would have at least $45 billion less in retirement income in 2030 than needed to cover their expenses. This shortfall highlights the pressing need of better strategies to manage the wealth and to avoid financial ruin. Annuities can be a very good way of saving money and securing post retirement benefits. A fixed-payout life annuity is a financial instrument that pays a fixed amount

periodically throughout the life of the recipient. It is crucial to choose the right plan and the right time to invest in annuity products. Therefore, motivated by the desire to apply probability optimization to problems faced by retirees, we find the optimal annuity-purchasing strategy for an individual who seeks to minimize the probability of wealth running out zero while the individual is still alive.

In the literature, Yaari (1965) in [23] proved that in the absence of bequest motives and in a deterministic financial economy, consumers would annuitize all of their liquid wealth. This result was generalized to a stochastic model by Richard (1975) in [17]. Recently, Davidoff et al. (2005) in [5] demonstrated the robustness of Yaari's result. Similarly, Kapur and Orszag (1999) in [12] and Brown (2001) in [3] provide theoretical and empirical guidance on the optimal time to annuitize under various market structures. Optimal investment strategy to minimize the probability of lifetime ruin was considered by Milevsky et al. (2006) in [15], in which they provided the annuity-purchasing strategies to minimize the probability of lifetime ruin.

Along another line, recent applications in financial engineering demand the consideration of systems that better describe the random environment. Unlike the previous work, where wealth is described as Brownian motion risk model (see [7]) or compound Poisson model (see [8]), the wealth in our chapter is modeled as a regime-switching diffusion modulated by a continuous-time Markov chain. Based on Markov chain approximation techniques, an approximation procedure to find optimal annuity-purchasing strategies for minimizing the probability of lifetime ruin was constructed. Several interesting results that are consistent with the economics intuition were obtained.

10.2.2 Formulation

We use a controlled hybrid diffusion model to represent the wealth. For simplicity, assume the system to be one dimensional. Let $(\Omega, \mathscr{F}, P)$ be a probability space and $\{\mathscr{F}_t\}$ be a filtration defined on it. Suppose that the discrete event process $\alpha(\cdot)$ is a continuous-time Markov chain having state space $\mathscr{M} = \{1, \ldots, m\}$ and generator $Q = (q_{\iota\ell})$. Let $\omega(\cdot)$ be a standard $\mathscr{F}_t$-Wiener process, and $u(\cdot)$ be an $\mathscr{F}_t$-adapted control, taking value in a compact set U. Such controls are said to be *admissible*.

First, for each $\iota \in \mathscr{M}$, define the excess consumption $Z(s,\iota) = c(\iota) - A(s,\iota)$, where $c(\iota)$ denotes a constant rate that the individual consumes and $A(s,\iota)$ is a nonnegative income rate at time s after any annuity purchases at that time. Then, $Z(s,\iota)$ is the net income the decision maker requires. Then the dynamic system can be written as

$$\begin{cases} dW(s,\alpha(s)) = r(\alpha(s)W(s,\alpha(s)) + [\mu(s,\alpha(s)) - r(\alpha(s))u(s) - Z(s,\alpha(s))]ds \\ \qquad\qquad\qquad\qquad +\sigma(s,\alpha(s))u(s)d\omega + a(s)dZ(s), \\ W(t,\alpha(t)) = w \geq 0, \\ Z(t,\alpha(t)) = z \geq 0, \end{cases} \tag{10.1}$$

where for each $\iota \in \mathscr{M}$, $W(s,\iota)$ denotes the wealth of the individual at time s. Let $u(s)$ be the amount that the decision maker invests in the risky asset at time s, and $0 \le u \le W$. We assume that the interest rate at time s is given by $r(\alpha(s))$. The individual can invest in a riskless asset with the yield rate $r(\iota)$ for each $\iota \in \mathscr{M}$, and a risky asset with return rate $\mu(s,\iota) > r(\iota)$ and volatility $\sigma(s,\iota) > 0$ for all $\iota \in \mathscr{M}$. We use $\lambda(s)$ to denote the hazard rate at age s. The actuarial present value of perpetuity with the life stream payment of one dollar per year by the interest rate $r(\alpha(t))$ and the hazard rate λ with the discount is

$$a(t) = \int_0^\infty \exp(-r(\alpha(t))s) \exp\left(-\int_t^{t+s} \lambda(v)dv\right) ds. \tag{10.2}$$

Since if $w \ge za(t)$, the individual can purchase the annuity immediately to guarantee a net income of z to avoid the lifetime ruin. Let τ_0 be the time when the wealth reaches zero and τ_d be the random time of death of the individual. Then the probability of lifetime ruin ψ at time t can be represented on the domain $D = \{(w,z,t,\alpha(t)) : 0 \le w \le za(t), z \ge 0, t \ge 0, \alpha(t) \in \mathscr{M}\}$ as

$$\begin{aligned} &\psi(w,z,t,\alpha(t)) \\ &= \inf_{\{u,Z\}} P[\tau_0 < \tau_d | W(t,\alpha(t)) = w, Z(t,\alpha(t)) = z, \tau_0 > t, \tau_d > t]. \end{aligned} \tag{10.3}$$

Note that $\tau_0 = \tau_0(x,u)$. That is, it depends on x as well on the control u. However, for notational simplicity, in what follows, we suppress the (x,u) dependence. Thus, $\psi(w,z,t,\alpha(t)) = 0$ when $w \ge za(t)$. Denote the cost function $P[\cdot]$ by $\varphi(w,z,t,\alpha(t),u)$. We can prove the following results. The proofs are omitted for brevity.

Proposition 10.1. *The probability of lifetime ruin is a constrained viscosity solution of the system of HJB variational inequalities*

$$\begin{aligned} \max\Big[\lambda(t)\psi - \psi_t - (rw-z)\psi_w - \min_u[(\mu-r)u\psi_w + \frac{1}{2}\sigma^2u^2\psi_{ww}] \\ + Q\psi(w,z,t,\cdot)(\iota), a(t)\psi_w + \psi_z\Big] = 0, \ \ \iota \in \mathscr{M}. \end{aligned} \tag{10.4}$$

Define $V(x,t,\iota) = \psi(x,1,t,\iota)$, where $x = w/z \in G$ and G denotes the range of x. Since the probability of lifetime ruin depends only on the ratio of wealth and income, which is $\psi(x,1,t,\iota) = \psi(w,z,t,\iota)$. We aim to find the optimal investment proportion in risky asset $u(t)$ to minimize the ruin probability. Then the value function is

$$V(x,t,\iota) = \inf_{u \in U} \varphi(w,z,t,\iota,u), \ \iota \in \mathscr{M}. \tag{10.5}$$

For an arbitrary $u \in U$, $\iota = \alpha(t) \in \mathscr{M}$, define an operator $\mathscr{L}_t^u$ by

$$\mathscr{L}_t^u V(x,t,\iota) = V_t + V_x(x,t,\iota)b(x,t,\iota,u) + \frac{1}{2}V_{xx}(x,t,\iota)\rho^2(t,\iota,u) + QV(x,t,\cdot)(\iota), \tag{10.6}$$

where V_x and V_{xx} denote the first and second derivatives with respect to x, V_t is the derivative with respect to t, and

$$\begin{aligned} &b(x,t,\iota,u) = r(\iota)x - 1 + (\mu(t,\iota) - r(\iota))u, \\ &\rho(t,\iota,u) = \sigma(t,\iota)u, \\ &QV(x,t,\cdot)(\iota) = \sum_{\ell \neq \iota} q_{\iota\ell}(V(x,t,\ell) - V(x,t,\iota)),\ \iota \in \mathcal{M}. \end{aligned}$$

Let $\mathcal{U}$ be the collection of admissible controls, the value functions have the following properties.

Proposition 10.2. *The probability of lifetime ruin can be written as*

$$\lambda(t)V(x,t,\iota) - \inf_{u\in U} \mathcal{L}_t^u V(x,t,\iota) = 0,\ \iota \in \mathcal{M}, \tag{10.7}$$

for $x < a(t)$ with boundary conditions $V(0,t,\iota) = 1$ and $V(a(t),t,\iota) = 0$ with the transversality condition

$$\lim_{s\to\infty} \exp(-\int_t^s \lambda(v)dv) E[V(X_s^*,s,\iota)|X_t = x] = 0,$$

in which X_s^* is the optimally controlled X_s.

10.2.3 Constant Hazard Rate

In this section, we assume the forces of mortality to be a constant. That is, $\lambda(t) = \lambda$ for all $t \geq 0$. Define an operator $\mathcal{L}^u$

$$\mathcal{L}^u V(x,\iota) = V_x(x,\iota)b(x,\iota,u) + \frac{1}{2}V_{xx}(x,\iota,u)\rho^2(\iota,u) + QV(x,\cdot)(\iota),\ \iota \in \mathcal{M}. \tag{10.8}$$

Using (8), (7) becomes

$$\lambda V(x,\iota) - \inf_{u\in U} \mathcal{L}^u V(x,\iota) = 0, \tag{10.9}$$

and the boundary conditions are $V(0,\iota) = 1$ and $V(1/(\min_i r(i) + \lambda),\iota) = 0, i \in \mathcal{M}$ Next, we present the local consistency for our approximating Markov chain.

Lemma 10.1. *The Markov Chain $\{\xi_n^h, \alpha_n^h\}$ with proper transition probabilities $(p^h(\cdot))$ is locally consistent with the stochastic differential equation in* (1).

To proceed, we use the Markov chain approximation method to construct the sequence, then piecewise constant interpolation is obtained here with appropriately chosen interpolation intervals. Using (ξ_n^h, α_n^h) to approximate the continuous-time process $(x(\cdot), \alpha(\cdot))$, we defined the continuous-time interpolation $(\xi^h(\cdot), \alpha^h(\cdot))$, $u^h(\cdot)$ and $\eta^h(t)$. Define $\mathcal{D}_t^h$ as the smallest σ-algebra of $\{\xi^h(s), \alpha^h(s), u^h(s), \eta^h(s), s \leq$

$t\}$. In addition, $\mathscr{U}^h$ is equivalent to the collection of all piecewise constant admissible controls with respect to $\mathscr{D}_t^h$.

Lemma 10.2. *The interpolated process of the constructed Markov chain* $\{\alpha^h(\cdot)\}$ *converges weakly to* $\alpha(\cdot)$*, the Markov chain with generator* $Q = (q_{\iota\ell})$.

We omit the details of proof here. The lemma above can be proved using the idea of two-time-scale Markov chains worked in [25]. To continue, we need the following assumptions.

(A1) For each $\iota \in \mathscr{M}$ and each $u \in U$, the function $b(\cdot,\iota,u)$ and $\sigma(\cdot,\iota)$ are continuous.
(A2) For each $\iota \in \mathscr{M}$, $\sigma(s,\iota) > 0, \forall s > 0$.
(A3) Let $u(\cdot)$ be an admissible ordinary control with respect to $\omega(\cdot)$ and $\alpha(\cdot)$, and suppose that $u(\cdot)$ is piecewise constant and takes only a finite number of values. Then for each initial condition, there exists a solution to the dynamic system where $m(\cdot)$ is the relaxed control representation of $u(\cdot)$. This solution is unique in the weak sense.

Theorem 10.1. *Assume* (A1) *and* (A2). *Let* $\{\xi_n^h, \alpha_n^h, n < \infty\}$*, the approximating chain, be constructed with transition probabilities,* $\{u_n^h, n < \infty\}$ *be a sequence of admissible controls,* $(\xi^h(\cdot), \alpha^h(\cdot))$ *be the continuous-time interpolation,* $m^h(\cdot)$ *be the relaxed control representation of* $\{u_n^h, n < \infty\}$*. Then* $(\xi^h(\cdot), \alpha^h(\cdot), m^h(\cdot), \omega^h(\cdot))$ *is tight. Denote the limit of weakly convergent subsequence by* $(\xi(\cdot), \alpha(\cdot), m(\cdot), \omega(\cdot))$ *and denote by* $\mathscr{F}_t$ *the* σ*-algebra generated by* $\{x(s), \alpha(s), m(s), \omega(s), s \le t\}$*. Then* $\omega(\cdot)$ *is a standard* $\mathscr{F}_t$*-Wiener process, and* $m(\cdot)$ *is an admissible control.*

By using the Skorohod representation, we can obtain the convergence of the cost function. As $h \to 0$, $\varphi^h(x,\iota,m^h) \to \varphi(x,\iota,m)$. We further obtain the following result.

Theorem 10.2. *Assume* (A1)–(A4). $V(x,\iota)$ *and* $V^h(x,\iota)$ *are value functions defined in* (5) *and the corresponding approximation sequence, respectively. Then* $V^h(x,\iota) \to V(x,\iota)$ *as* $h \to 0$.

10.2.4 General Hazard Rate

In this section, we assume the forces of mortality are not constant, but a continuous function with respect to t for all $t \ge 0$. Define another function $\hat{g}(x,T)$ to approximate the transversality condition of (10.2), and $\hat{g}(x,T) \to 0$ as $T \to \infty$. Under this condition, (7) becomes

$$\lambda(t)V(x,t,\iota) - \inf_{u\in U} \mathscr{L}_t^u V(x,t,\iota) = 0 \tag{10.10}$$

with the boundary condition $V(0,t,\iota) = 1$ and $V(a(t),t,\iota) = 0$, and terminal condition as $V(x,T,\iota) = \hat{g}(x,T)$.

Theorem 10.3. *Assume* (A1) *and* (A2). *Let* $\{\xi_n^{h,\delta}, \alpha_n^{h,\delta}, n < \infty\}$, *the approximating chain, be constructed with transition probabilities,* $\{u_n^{h,\delta}, n < \infty\}$ *be a sequence of admissible controls,* $(\xi^{h,\delta}(\cdot), \alpha^{h,\delta}(\cdot), \lambda^{h,\delta}(\cdot))$ *be the continuous-time interpolation,* $m^{h,\delta}(\cdot)$ *be the relaxed control representation of* $\{u_n^{h,\delta}, n < \infty\}$. *Then* $(\xi^{h,\delta}(\cdot), \alpha^{h,\delta}(\cdot), m^{h,\delta}(\cdot), \omega^{h,\delta}(\cdot), \lambda^{h,\delta}(\cdot))$ *is tight. Denote by* $\mathscr{F}_t$ *the limit of weakly convergent subsequence by* $(\xi(\cdot), \alpha(\cdot), m(\cdot), \omega(\cdot), \lambda(\cdot))$ *and denote the* σ*-algebra generated by* $\{x(s), \alpha(s), m(s), \omega(s), \lambda(s), s \le t\}$. *Then* $\omega(\cdot)$ *is a standard* $\mathscr{F}_t$*-Wiener process, and* $m(\cdot)$ *is an admissible control.*

Similar to the constant hazard rate case, it is not hard to get the convergence of the cost function with the general hazard rate by virtue of the Skorohod representation. Convergence of the value function can also be obtained.

Theorem 10.4. *Assume* (A1)–(A4). $V(x,t,\iota)$ *and* $V^{h,\delta}(x,t,\iota)$ *are value functions defined in* (5) *and corresponding approximation sequence, respectively. Then* $V^{h,\delta}(x,t,\iota) \to V(x,t,\iota)$ *as* $h, \delta \to 0$.

10.2.5 Examples

In this section, we consider a couple of examples with constant and more general hazard rates with two regimes, respectively. For simplicity, we deal with systems that are linear in the wealth. The dynamic system becomes

$$\begin{aligned} dW(s,\alpha(s)) = {} & rA(\alpha(s))W(s) + B(\alpha(s))(\mu(s) - r))u(s) - Z(s))ds \\ & + C(\alpha(s))\sigma(s)u(s)d\omega + a(s)dZ(s). \end{aligned} \tag{10.11}$$

Suppose $r = 0.01$ (the yield rate of riskless asset), $\mu = 0.04$ (the yield rate of risky asset), $\sigma = 0.1$ (the volatility of the risky asset), $z = 1$ (the individual consumes one unit wealth per year).

Example 10.1. Take $\lambda = 0.03$, the hazard rate is 0.03 such that the expected future lifetime of individual is 33.3 years. The Markov Chain $\alpha(\cdot) \in \mathscr{M}$ with $\mathscr{M} = \{1,2\}$ and generator Q, and

$$Q = \begin{pmatrix} -0.2 & 0.2 \\ 0.8 & -0.8 \end{pmatrix}, \quad \begin{cases} A(1) = 1 \\ A(2) = 2, \end{cases} \quad \begin{cases} B(1) = 4 \\ B(2) = 1, \end{cases} \quad \begin{cases} C(1) = 2 \\ C(2) = 1. \end{cases} \tag{10.12}$$

We use the value iteration to numerically solve the optimal control problem, then we obtain the relationship between wealth and the probability of lifetime ruin as in Fig. 10.1. In addition, we compute the probability of lifetime ruin under the assumption of exponential future lifetime for an individual with wealth \$1 who invests in the riskless asset only with constant interest rate and self-annuitizes. The probability of lifetime ruin will be

$$P[\tau_0 < \tau_d] = \exp(-r\tau_d) = (1 + r/\lambda)^{-\frac{\lambda}{r}} = 0.422.$$

Moreover, Fig. 10.1 shows that the probability of ruin with life annuity purchase is less than 0.444 when the initial wealth $w \in (0.5, 1)$. Comparing to the probability of lifetime ruin without life annuity purchasing and the consumption $z = r + \lambda = 0.04$, if the individual buys the life annuity as in (1), the individual will have less probability of financial ruin even with lower wealth than the individual with self-annuitization to maintain the same consumption.

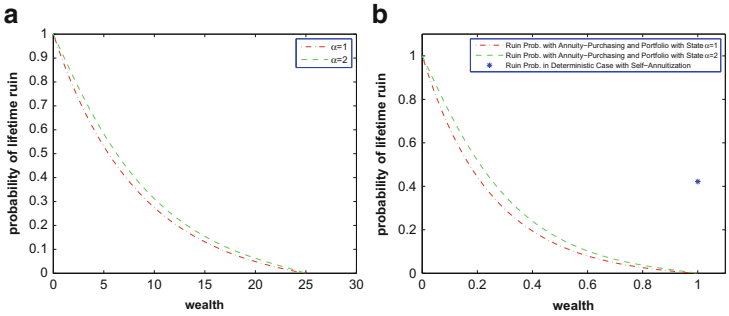

Fig. 10.1 Constant hazard rate with two regimes. (**a**) Probability of lifetime ruin versus wealth. (**b**) Comparison of ruin probability between annuity-purchasing and self-annuitization.

Example 10.2. In this example, we consider Gompertz hazard rate $\lambda(t) = \exp(\frac{t-\bar{m}}{b})/b$, where $\bar{m}$ is a model value and b is a scale parameter, we choose $\bar{m} = 90$ and $b = 9$. We also consider the terminal condition to be exponentially decay as $\hat{g}(x, T) = \exp(-xT)$. We consider the same Markov chain as in the last example. To illustrate the impact of ages of the investors on the probability of lifetime ruin, three age levels are presented as $t = 30, t = 50, t = 70$. From Fig. 10.2, we can see that the individual with the same wealth but younger age will more likely to outlive his or her wealth.

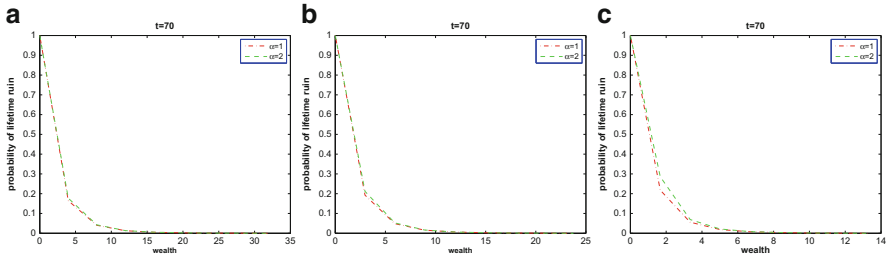

Fig. 10.2 Proportion of assets vs. time. (**a**) Probability of lifetime ruin versus wealth with age 30, (**b**) probability of lifetime ruin versus wealth with age 50, (**c**) probability of lifetime ruin versus wealth with age 70

10.3 Optimal Dividend Payment Policies

10.3.1 Motivation

Designing dividend payment policies has long been an interesting and important research issue in actuarial science and finance literature. The dividend decision is crucial because not only does it represent an important signal about a firm's future growth opportunities and profitability but also may influence the investment and financing decisions of firms and the wealth of the policyholders. For insurance companies, because of the nature of their product, insurers tend to accumulate relatively large amounts of cash, cash equivalents, and investments in order to pay future claims and avoid financial ruin. The study of insurance companies' dividend decisions is thus desirable because the payment of dividends to shareholders may reduce an insurer's ability to survive adverse investment and underwriting experience. Recently, the financial crisis has led to the controversial discussion on the dividend policy of European insurance industry, see [16].

In actuarial science, stochastic optimization problems such as designing optimal risk controls for an insurance corporation have drawn increasing attention since the introduction of the classical collective risk model in [14], where the probability of ruin was used to measure the risk. De Finetti (1957) in [6] proposed a dividend optimization problem after realizing that the surplus is not realistic in practice to reach arbitrarily high and exceed any finite level. Instead of considering the safety aspect (ruin probability), aiming at maximizing the expected discounted total dividends until ruin by assuming the surplus process follows a simple random walk, he showed that the optimal dividend strategy is a barrier strategy. Since then, many researchers have analyzed this optimization problem under more realistic assumptions and extended its range of applications. Some recent work can be found in [1, 2, 4, 7] and references therein.

In particular, dividend optimization has been widely studied using regime-switching models such as optimal dividend payment policy with ruin constraint, reinsurance strategy, and investment portfolio allocation, see [21, 24]. In this part, the surplus process is modulated by a jump diffusion with regime-switching process to study the maximization of the expected discounted total dividends until ruin. The process describing the regime-switching is assumed to be a continuous-time Markov chain representing the random environment. As mentioned above, this model appears to be more versatile and more realistic than the classical compound Poisson and diffusion models. Thus, we solve a system of HJB partial differential equations instead of a single HJB equation under this model, which is very difficult to solve analytically. We aim to construct feasible numerical approximation schemes for finding a good approximation to the underlying problems.

10.3.2 Formulation

To delineate the random environment and other random factors, we use a continuous-time Markov chain $\alpha(t)$ whose generator is $Q = (q_{\iota\ell}) \in \mathbb{R}^{m\times m}$ and state space is $\mathscr{M} = \{1,\ldots,m\}$. Let ν_n be the arrival time of the nth claim. Corresponding to each $\iota \in \mathscr{M}$, $N_\iota(t) = \max\{n \in \mathbb{N} : \nu_n \le t\}$ is the number of claims up to time t, which is a Poisson counting process.

The surplus process under consideration is a regime-switching jump diffusion. For each $\iota \in \mathscr{M}$, the premium rate is $c(\iota) > 0$ and the volatility is $\sigma(\iota) > 0$. Let $R_\iota(t)$ for each $\iota \in \mathscr{M}$ be a jump process representing claims with arrival rate λ_ι, claim size distribution F_ι, and zero initial surplus. The function $q(x,\iota,\rho)$ is assumed to be the magnitude of claim size, where ρ have the distribution $\Pi(\cdot)$. Then the Poisson measure $N_\iota(\cdot)$ has intensity $\lambda_\iota dt \times \Pi_\iota(d\rho)$ where $\Pi_\iota(d\rho) = f_\iota(\rho)d\rho$. The surplus process before dividend payment is given by

$$\begin{aligned} d\widetilde{x}(t) &= \sum_{\iota\in\mathscr{M}} I_{\{\alpha(t)=i\}}(c(\iota)dt + \sigma(\iota)dw(t) - dR_\iota(t)) \\ &= \Big[c(\alpha(t))dt + \sigma(\alpha(t))dw(t)\Big] - \int_{\mathbb{R}_+} q(x(t^-),\alpha(t),\rho)N_{\alpha(t)}(dt,d\rho), \end{aligned} \tag{10.13}$$

where I_A is the indicator function of the set A, $c(\iota) > 0$ and $\sigma(\iota) > 0$ for each $\iota \in \mathscr{M}$, and $w(t)$ is a standard Brownian motion. Assume that $q(\cdot,\iota,\rho)$ is continuous for each ρ and each $\iota \in \mathscr{M}$. We are working on a filtered probability space $(\Omega,\mathscr{F},\{\mathscr{F}_t\},P)$, where $\mathscr{F}_t$ is the σ-algebra generated by $\{\alpha(s),w(s),N_\iota(s) : 0 \le s \le t,\ \iota \in \mathscr{M}\}$.

Note that the drift c describes the premium magnitude collected by the insurance company, and is modulated by a finite Markov Chain $\alpha(t)$, which represents the market mode and other economic conditions. It is used to determine the amount charged by the insurer and mainly depends on the insurance coverage, not surplus. The volatility σ refers to measures of risk in the market here. Like the drift c, it is mainly affected by the market mode. From a numerical approximation point of view, making c and σ x-dependent will not introduce any essential difficulty.

A dividend strategy $D(\cdot)$ is an $\mathscr{F}_t$-adapted process $\{D(t) : t \ge 0\}$ corresponding to the accumulated amount of dividends paid up to time t such that $D(t)$ is a nonnegative and nondecreasing stochastic process that is right continuous and have left limits with $D(0^-) = 0$. In general, a dividend process is not necessarily an absolutely continuous process. In this chapter, we consider the optimal dividend strategy, which is either a barrier strategy or a band strategy. In both cases, the dividend rate is the same as the premium rate. As a result, $D(t)$ is absolutely continuous. Denote $\Gamma = [0,C]$. Since the optimal dividends policy is either a barrier or a band strategy, $D(t)$ is an absolutely continuous process. We write $D(t)$ as

$$dD(t) = u(t)dt,\ \ 0 \le u(t) \le C, \tag{10.14}$$

where $u(t)$ is an $\mathscr{F}_t$-adapted process and $0 < C < \infty$. Note that if $C < c(\iota)$ for some $\iota \in \mathscr{M}$, this formulation will lead to a threshold strategy. If $C \geq c(\iota)$ for all $\iota \in \mathscr{M}$, the optimal strategy is either a barrier or band strategy. Then the surplus process in the presence of dividend payments is given by

$$dx(t) = d\tilde{x}(t) - dD(t),\ x(0) = x \geq 0 \tag{10.15}$$

for all $t < \tau$ and we impose $x(t) = 0$ for all $t > \tau$, where $\tau = \inf\{t \geq 0 : x(t) \leq 0\}$ represents the time of ruin. Denote $\Gamma = [0, C]$, $0 < C < \infty$. Suppose the dividend is paid at a rate $u(t)$, where $u(t)$ is an $\mathscr{F}_t$-adapted process, and the optimal payout strategy is applied subsequently. Then the expected discounted dividend until ruin is given by

$$J(x, \iota, u(\cdot)) = E_{x,\iota}\Big[\int_0^\tau e^{-rt}u(t)dt\Big],\ \iota \in \mathscr{M}, \tag{10.16}$$

where $E_{x,\iota}$ denotes the expectation conditioned on $x(0) = x$ and $\alpha(0) = \iota$.

Combining (13) and (15), we can rewrite the surplus process with the dividend payment as

$$\begin{aligned} dx(t) &= \Big[c(\alpha(t)) - u(t)\Big]dt + \sigma(\alpha(t))dw(t) - dR(t), \\ R(t) &= \sum_{\iota \in \mathscr{M}} I_{\{\alpha(t)=\iota\}} R_\iota(t) = \int_0^t \int_{\mathbb{R}_+} q(x(t^-), \alpha(t), \rho) N_{\alpha(t)}(dt, d\rho), \\ x(0) &= x. \end{aligned} \tag{10.17}$$

Admissible Strategies. A strategy $u(\cdot) = \{u(t) : t \geq 0\}$ satisfying $u(t) \in \Gamma$ being progressively measurable with respect to $\sigma\{\alpha(s), w(s), N_\iota(s) : 0 \leq s \leq t,\ \iota \in \mathscr{M}\}$ is called an admissible strategy. Denote the collection of all admissible strategies or admissible controls by $\mathscr{A}$. A Borel measurable function $u(x, \alpha)$ is an admissible feedback strategy or feedback control if (17) has a unique solution.

We are interested in finding the optimal dividend rate $u(t)$ that is bounded and is a function of x and α to maximize the expected utility function $J(x, \iota, u(\cdot))$. Define $V(x, \iota)$ as the optimal value of the corresponding problem. That is,

$$V(x, \iota) = \sup_{u(\cdot) \in \mathscr{A}} J(x, \iota, u(\cdot)). \tag{10.18}$$

Setting $u(t)$ to any quantity such that it does not change the value of $V(x(\tau), \alpha(\tau))$ for $t \geq \tau$, that is, $u(t) = 0$ for $t \geq \tau$, Therefore, (16) can be rewritten as

$$J(x, \iota, u(\cdot)) = E_{x,\iota}\Big[\int_0^\infty e^{-rt}u(t)dt\Big]. \tag{10.19}$$

The optimal dividend problem is formulated as

$$\begin{cases} \text{maximize}: J(x,\iota,u(\cdot)) = E_{x,\iota}\displaystyle\int_0^\infty e^{-rt}u(t)dt, \\ \text{subject to}: dx(t) = [c(\alpha(t)) - u(t)]dt + \sigma(\alpha(t))dw(t) \\ \qquad\qquad - \displaystyle\int_{\mathbb{R}_+} q(x(t^-),\alpha(t),\rho)N_{\alpha(t)}(dt,d\rho), \\ \qquad\qquad x(0) = x, \;\; \alpha(0) = \iota, \;\; u(\cdot) \in \mathscr{A}, \\ \text{value function}: V(x,\iota) = \displaystyle\sup_{u(\cdot)\in\mathscr{A}} J(x,\iota,u(\cdot)), \quad \text{for each } \iota \in \mathscr{M}. \end{cases} \tag{10.20}$$

For an arbitrary $u \in \mathscr{A}$, $\iota = \alpha(t) \in \mathscr{M}$, and $V(\cdot,\iota) \in C^2(\mathbb{R})$, define an operator $\mathscr{L}^u$ by

$$\begin{aligned} \mathscr{L}^u V(x,\iota) = {} & V_x(x,\iota)(c(\iota) - u) + \frac{1}{2}\sigma(\iota)^2 V_{xx}(x,\iota) + QV(x,\cdot)(\iota) \\ & + \lambda_\iota \int_0^x [V(x - q(x,\iota,\rho),\iota) - V(x,\iota)] f_\iota(\rho)d\rho, \end{aligned} \tag{10.21}$$

where V_x and V_{xx} denote the first and second derivatives with respect to x, and

$$QV(x,\cdot)(\iota) = \sum_{\ell\neq\iota} q_{\iota\ell}(V(x,\ell) - V(x,\iota)).$$

Note that

$$J(x,\iota,u) = E_{x,\iota}\left[\int_0^\infty e^{-rt}u(t)dt\right] \le E_{x,\iota}\left[\int_0^\infty e^{-rt}Cdt\right] \le \frac{C}{r}.$$

Taking $\sup_u$ in the above inequality leads to that $V(x,\iota)$ is bounded. Furthermore, by the concavity of $V(x,\iota)$ and monotonicity (nondecreasing) of $V_x(x,\iota)$ (see [19]), we have

$$\lim_{x\to\infty} V_x(x,\iota) = 0.$$

Formally, the value function (18) satisfies the HJB equations

$$\begin{cases} \displaystyle\max_{u\in[0,C]} \{\mathscr{L}^u V(x,\iota) - rV(x,\iota) + u\} = 0, \;\; \forall \iota \in \mathscr{M}, \\ V(0,\iota) = 0, \;\; \forall \iota \in \mathscr{M}. \end{cases} \tag{10.22}$$

10.3.3 Algorithm

To construct a locally consistent Markov chain approximation for the jump diffusion model with regime-switching, we consider an equivalent way to define the process (17) by working with the claim times and values. To do this, set $\nu_0 = 0$, and let ν_n, $n \ge 1$, denote the time of the nth claim, and $q(\cdot,\cdot,\rho_n)$ is the corresponding claim intensity with a suitable function of $q(\cdot)$. Let $\{\nu_{n+1} - \nu_n, \rho_n, n < \infty\}$ be mutually independent random variables with $\nu_{n+1} - \nu_n$ being exponentially distributed with mean $1/\lambda$, and let ρ_n have a distribution $\Pi(\cdot)$. Furthermore, let $\{\nu_{k+1} - \nu_k, \rho_k, k \ge n\}$ be

independent of $\{x(s),\alpha(s),s<\nu_n,\nu_{k+1}-\nu_k,\rho_k,k<n\}$, then the nth claim term is $q(x(\nu_n^-),\alpha(\nu_n),\rho_n)$, and the claim amount $R(t)$ can be written as

$$R(t)=\sum_{\nu_n\le t} q(x(\nu_n^-),\alpha(\nu_n),\rho_n).$$

Because $\nu_{n+1}-\nu_n$ is exponentially distributed, we can write

$$P\{\text{claim occurs on } [t,t+\Delta)|x(s),\alpha(s),w(s),N(s,\cdot),s\le t\}=\lambda\Delta+o(\Delta). \quad (10.23)$$

By the independence and the definition of ρ_n, for any $H\in\mathscr{B}(\mathbb{R}_+)$, we have

$$\begin{aligned} &P\{x(t)-x(t^-)\in H|t=\nu_n \text{ for some n};w(s),x(s),\alpha(s),N(s,\cdot),s<t;\\ &\qquad\qquad x(t^-)=x,\alpha(t)=\alpha\} \qquad (10.24)\\ &=\Pi(\rho: q(x(t^-),\alpha(t),\rho)\in H). \end{aligned}$$

It is implied by the above discussion that $x(\cdot)$ satisfying (17) can be viewed as a process that involves regime-switching diffusion with claims according to the claim rate defined by (23). Given that the nth claim occurs at time ν_n, we construct the values according to the conditional probability law (24) or, equivalently, write it as $q(x(\nu_n^-),\alpha(\nu_n),\rho_n)$. Then the process given in (17) is a switching diffusion process until the time of the next claim.

To begin, we construct a discrete-time, finite-state, controlled Markov chain to approximate the controlled diffusion process with regime-switching in the finite state space G_h. Then we obtain the corresponding transition probabilities

$$\begin{aligned} &p_D^h((x,\iota),(x+h,\iota)|u)=\frac{(\sigma^2(\iota)/2)+h(c(\iota)-u)^+}{D-rh^2},\\ &p_D^h((x,\iota),(x-h,\iota)|u)=\frac{(\sigma^2(\iota)/2)+h(c(\iota)-u)^-}{D-rh^2},\\ &p_D^h((x,\iota),(x,\ell)|u)=\frac{h^2}{D-rh^2}q_{\iota\ell},\ \text{ for } \ell\ne\iota,\\ &p_D^h(\cdot)=0,\ \text{ otherwise},\\ &\Delta t^h(x,\iota,u)=\frac{h^2}{D}, \end{aligned} \quad (10.25)$$

with

$$D=\sigma^2(\iota)+h|c(\iota)-u|+h^2(r-q_{\iota\iota})$$

being well defined. Suppose that the current state is $\xi_n^h=x$, $\alpha_n^h=\iota$, and control is $u_n^h=u$. The next interpolation interval $\Delta t^h(x,\iota,u)$ is determined by (25). To present the claim terms, we determine the next state $(\xi_{n+1}^h,\alpha_{n+1}^h)$ by noting:

1. With probability $(1-\lambda\Delta t^h(x,\iota,u)+o(\Delta t^h(x,\iota,u)))$, no claims are in $[t_n^h,t_{n+1}^h)$; we determine $(\xi_{n+1}^h,\alpha_{n+1}^h)$ by transition probability $p_D^h(\cdot)$ as in (25).
2. With probability $\lambda\Delta t^h(x,\iota,u)+o(\Delta t^h(x,\iota,u)))$, there is a claim in $[t_n^h,t_{n+1}^h)$; we determine $(\xi_{n+1}^h,\alpha_{n+1}^h)$ by

$$\xi_{n+1}^h = \xi_n^h - q_h(x,\iota,\rho), \alpha_{n+1}^h = \alpha_n^h,$$

where $\rho \sim \Pi(\cdot)$, and $q_h(x,\iota,\rho) \in S_h \subseteq \mathbb{R}_+$ such that $q_h(x,\iota,\rho)$ is the nearest value of $q(x,\iota,\rho)$ so that $\xi_{n+1}^h \in S_h$. Then $|q_h(x,\iota,\rho) - q(x,\iota,\rho)| \to 0$ as $h \to 0$, uniformly in x.

Let H_n^h denote the event that $(\xi_{n+1}^h, \alpha_{n+1}^h)$ is determined by the first alternative above and use T_n^h to denote the event of the second case. Let $I_{H_n^h}$ and $I_{T_n^h}$ be corresponding indicator functions, respectively. Then $I_{H_n^h} + I_{T_n^h} = 1$. Then we need a new definition of the local consistency for Markov chain approximation of compound Poisson process with diffusion and regime-switching.

Definition 10.1. A controlled Markov chain $\{(\xi_n^h, \alpha_n^h), n < \infty\}$ is said to be locally consistent with (17), if there is an interpolation interval $\Delta t^h(x,\iota,u) \to 0$ as $h \to 0$ uniformly in x,ι, and u such that

1. there is a transition probability $p_D^h(\cdot)$ that is locally consistent with the diffusion process without jumps.
2. there is a $\delta^h(x,\iota,u) = o(\Delta t^h(x,\iota,u))$ such that the one-step transition probability $\{p^h((x,\iota),(y,\ell))|u\}$ is given by

$$\begin{aligned} p^h(((x,\iota),(y,\ell))|u) &= (1-\lambda\Delta t^h(x,\iota,u) + \delta^h(x,\iota,u))p_D^h((x,\iota),(y,\ell)) \\ &\quad +(\lambda\Delta t^h(x,\iota,u) + \delta^h(x,\iota,u))\Pi\{\rho : q_h(x,\iota,\rho) = x-y\}. \end{aligned} \tag{10.26}$$

Furthermore, the system of dynamic programming equations is

$$V^h(x,\iota) = \begin{cases} \max\limits_{u\in U}\Big[(1-\lambda\Delta t^h(x,\iota,u) + \delta^h(x,\iota,u))e^{-r\Delta t^h(x,\iota,u)} \\ \quad \times \sum\limits_{y,\ell}(p_D^h((x,\iota),(y,\ell))|u)V^h(y,\ell) \\ \quad +(\lambda\Delta t^h(x,\iota,u) + \delta^h(x,\iota,u))e^{-r\Delta t^h(x,\iota,u)} \\ \quad \times \int_0^x V^h(x-q_h(x,\iota,\rho),\iota)\Pi(d\rho) + u\Delta t^h(x,\iota,u)\Big], \text{ for } x \in G_h, \\ 0, \text{ for } x = 0. \end{cases} \tag{10.27}$$

10.3.4 Convergence

Lemma 10.3. *The Markov chain $\{\xi_n^h, \alpha_n^h\}$ with transition probabilities $(p_D^h(\cdot))$ defined in* (25) *is locally consistent with the stochastic differential equation in* (17).

We need one more assumption.

(B1) Let $\hat{\tau}(\phi) = \infty$, if $\phi(t) \in G^o$, for all $t < \infty$, otherwise, define $\hat{\tau}(\phi) = \inf\{t : \phi \notin G^o\}$. The function $\hat{\tau}(\cdot)$ is continuous (as a map from $D[0,\infty)$, the space of functions that are right continuous and have left limits endowed with the

Skorohod topology to the interval $[0,\infty]$ (the extended and compactified positive real numbers)) with probability one relative to the measure with initial condition (x,α).

Lemma 10.4. *The interpolated process of the constructed Markov chain* $\{\alpha^h(\cdot)\}$ *converges weakly to* $\alpha(\cdot)$*, the Markov chain with generator* $Q=(q_{\iota\ell})$.

Theorem 10.5. *Let the approximating chain* $\{\xi_n^h,\alpha_n^h,n<\infty\}$ *constructed with transition probabilities defined in* (25) *be locally consistent with* (17)*,* $\{u_n^h,n<\infty\}$ *be a sequence of admissible controls, and* $(\xi^h(\cdot),\alpha^h(\cdot))$ *be the continuous-time interpolation. Let* $\{\widetilde{\tau}_h\}$ *be a sequence of* $\mathscr{F}_t^h$*-stopping times. Then* $\{\xi^h(\cdot),\alpha^h(\cdot),u^h(\cdot),w^h(\cdot),N^h(\cdot),\widetilde{\tau}_h\}$ *is tight.*

Theorem 10.6. *Let the limit of the weakly convergent subsequence be denoted by* $(\xi(\cdot),\alpha(\cdot),u(\cdot),w(\cdot),N(\cdot),\widetilde{\tau})$ *and* $\mathscr{F}_t$ *the* σ*-algebra generated by* $\{x(s),\alpha(s),u(s),w(s),N(s),s\le t,\widetilde{\tau}I_{\{\widetilde{\tau}<t\}}\}$*. Then* $w(\cdot)$ *and* $N(\cdot)$ *are a standard* $\mathscr{F}_t$*-Wiener process and Poisson measure, respectively, and* $\widetilde{\tau}$ *is an* $\mathscr{F}_t$*-stopping time and* $u(\cdot)$ *is an admissible control. Let the claim times and claim sizes of* $N(\cdot)$ *be denoted by* ν_n,ρ_n*. Then, the limit satisfies the jump diffusion model with regime switching.*

Theorem 10.7. *Assume* (B1)*.* $V(x,\iota)$ *and* $V^h(x,\iota)$ *are value functions and corresponding approximation sequence, respectively. Then* $V^h(x,\iota)\to V(x,\iota)$ *as* $h\to 0$.

10.3.5 Examples

This section is devoted to a couple of examples. For simplicity, we consider the case the discrete event has two states. That is, the continuous-time Markov chain has two states.

Example 10.3. The Markov chain $\alpha(t)$ representing the discrete event state has generator Q

$$Q=\begin{pmatrix}-0.5 & 0.5\\ 0.5 & -0.5\end{pmatrix},$$

and takes values in $\mathscr{M}=\{1,2\}$. The premium size depends on the discrete state with $c(1)=2$ and $c(2)=3$. The dividend rate $u(t)$ taking its value in $[0,2]$ is a control parameter, $\sigma(\alpha(t))dw(t)$ is interpreted as small claim fluctuation and/or fluctuations due to premium incomes with $\sigma(1)=0.2$ and $\sigma(2)=2$, and $R(t)$ is a Poisson process interpreted as claims with $R(t)=\sum_{\nu_n\le t}\rho_n$, where $\rho_n\in\{0.01,0.015\}$, with distribution $\Pi(0.01)=0.7,\Pi(0.02)=0.3$. Let $\lambda_\iota=4$, for $\iota=1,2$. Then $\{\nu_{n+1}-\nu_n\}$ is a sequence of exponentially distributed random variables with mean 1/4. Furthermore, the initial surplus x is supposed to have the maximum 100 and the minimum 0. We use policy iteration methods to numerically solve the optimal control problems. This provides us with the advantage that we trace out the optimal policy for the portfolio selection. We obtain the computation results depicted in Figs. 10.3 and 10.4 as follows.

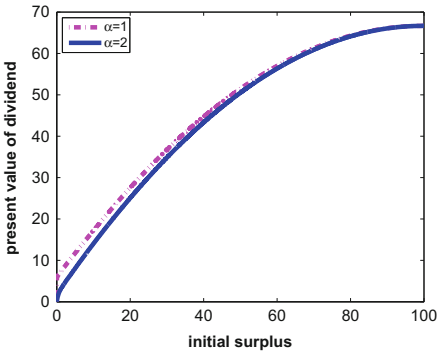

Fig. 10.3 Maximal expected present value of dividend versus initial surplus

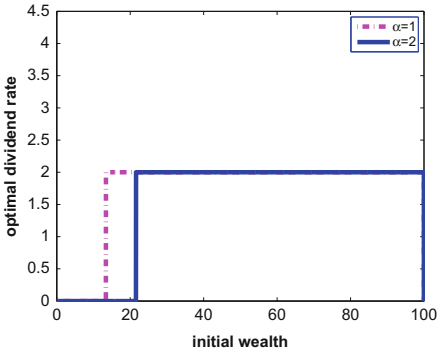

Fig. 10.4 Optimal dividend rate versus initial surplus

Example 10.4. Comparing with Example 10.3, we consider the case that the dividend rate is more than the premium rate. Use data exactly the same as above, but change the range of dividend rate to $[0,4]$. Then we obtain the computation results depicted in Figs. 10.5 and 10.6 as follows.

Example 10.5. In this example, we assume the difference of the volatilities in the two regimes is bigger comparing to Example 10.3. That is, taking $\sigma(1) = 0.1$ and $\sigma(2) = 3$. Then we obtain the computation results given in Figs. 10.7 and 10.8 as follows.

Figures 10.3–10.8 show that the dividend strategy is a threshold strategy (Example 10.3), or a band strategy (Example 10.4). The dividend is paid when $V_x(x,\iota) < 1$, in which case the company is "inefficient" and cash surplus is high, otherwise, the company is considered "efficient" when $V_x(x,\iota) > 1$. It is best to pay no dividend when the company is efficient and the cash surplus is low, then funds should be left to company for growth. Furthermore, when the dividend payment is executed, the pay out is supposed to be as much as allowable. That is, the cap of the dividend payment rate could be achieved. Such a dividend payment policy is the so-called bang–bang strategy.

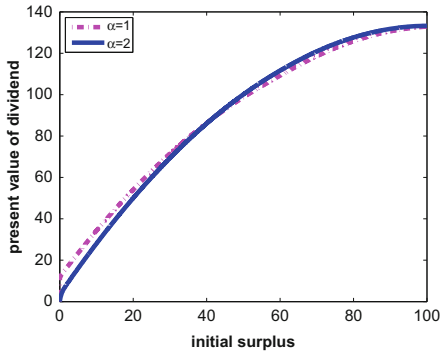

Fig. 10.5 Maximal expected present value of dividend versus initial surplus

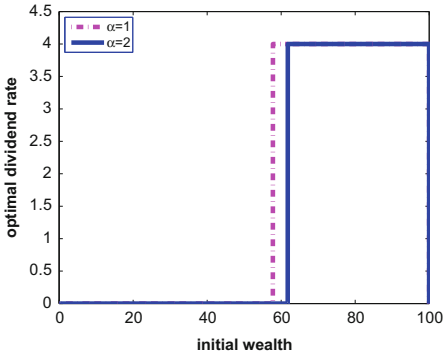

Fig. 10.6 Optimal dividend rate versus initial surplus

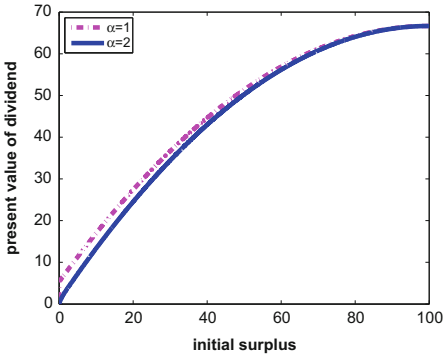

Fig. 10.7 Maximal expected present value of dividend versus initial surplus

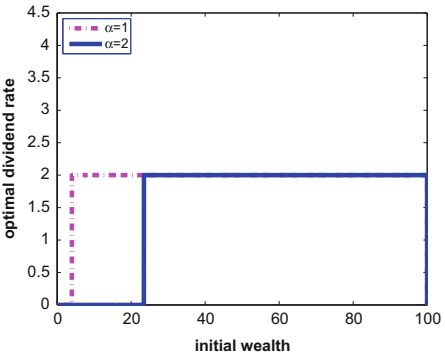

Fig. 10.8 Optimal dividend rate versus initial surplus

By examining the graphs, the following observations are in order. Figures 10.3–10.6 show that the dividend payment rates reach the thresholds depending on the sign of $V_x(x,t)-1$ no matter whether the ceiling of the dividend payment rate is greater than the premium rate or not. However, when the cap of dividend rate is larger in Example 10.4, the dividend is paid until the surplus reaches much higher level than the one in Example 10.3. This is a kind band strategy, when the initial surplus is greater than 60 (the surplus – 60) should be paid as dividend. But since we have a upper bound for the dividend rate, the payment rate becomes the upper bound. It is consistent with the theoretical result.

In addition, the difference of volatilities in Example 10.3 is 1.8 and the difference of volatilities in Example 10.5 is 3.9. From Figs. 10.7 and 10.8, we can see that the difference of the dividend payment strategies is bigger comparing to Figs. 10.3 and 10.4, in which case the difference of the volatilities is smaller. So the optimal dividend strategies are sensitive to the market regimes. This indicates that the regime-switching models are appropriate for the intended modeling and optimization.

10.4 Concluding Remarks

In this work, we reviewed some of our recent work on numerical approximation schemes to annuity purchasing and dividend optimization problems arising in insurance risk controls. The models under consideration involve regime switching, in which the switching process is represented by a Markov chain. Although detailed proofs are not provided and referred to our recent papers [10] and [11] for further reading, the statements of the results as well as conditions needed are spelled out. Although one could derive the associate system of HJB equations by using the usual dynamic programming approach together with the use of properties of switching jump diffusions, solving them analytically is very difficult. As an alternative, one may try to discretize the system of HJB equations directly, but this relies on the

properties of the HJB equations. We present a viable alternative. Our Markov chain approximation method uses mainly probabilistic methods that do not need any analytic properties of the solutions of the system of HJB equations. In the actual computation, the optimal control can be obtained by using the value or policy iteration methods.

Acknowledgments The research of Zhuo Jin was supported in part by Faculty Research Grant of the University of Melbourne. The research of George Yin was supported in part by the National Science Foundation under DMS-1207667.

References

1. Asmussen, S., Høgaard, B., and Taksar, M. (2000). Optimal risk control and dividend distribution policies. Example of excess-of loss reinsurance for an insurance corporation. *Finance and Stochastics*, 4: 299–324.
2. Asmussen, S. and Taksar, M. (1997). Controlled diffusion models for optimal dividend pay-Out. *Insurance: Math. and Economics*, 20: 1–15.
3. Brown, J. R., (2001). Private pensions, mortality risk, and the decision to annuitize, *Journal of Public Economics*, 82: 29–62.
4. Choulli, T., Taksar, M., and Zhou, X. Y. (2001). Excess-of-loss reinsurance for a company with debt liability and constraints on risk reduction. *Quant. Finance* 1: 573–96.
5. Davidoff, T., Brown, J. and Diamond, P. (2005). Annuities and individual swelfare, M.I.T. Department of Economics Working Paper Series, *The American Economic Review*, 95: 1573–1590.
6. De Finetti, B. (1957). Su unimpostazione alternativa della teoria collettiva del rischio. *Transactions of the XVth International Congress of Actuaries* 2: 433–443.
7. Gerber, H. and Shiu, E. (2004). Optimal dividends: analysis with Brownian motion. *North American Actuarial Journal*, 8: 1–20.
8. Gerber, H. and Shiu, E. (2006). On optimal dividend strategies in the compound Poisson model. *North American Actuarial Journal*, 10: 76–93.
9. Hamilton, J. (1989). A new approach to the economic analysis of non-stationary time series. *Econometrica*, 57: 357–384.
10. Jin, Z. and Yin, G. (2011). A numerical method for annuity-purchasing decision making to minimize the probability of financial ruin for regime-switching wealth models, *Internat. J. Computer Math.*, 88:1256–1282.
11. Jin, Z., Yin, G., and Yang, H.L. (2011). Numerical methods for dividend soptimization using regime-switching jump-diffusion models, *Math. Control Related Fields*, 1: 21–40.
12. Kapur, S. and Orszag, M., (1999). A portfolio approach to investment and annuitization during retirement. Working Paper. Birkbeck College, London.
13. Kushner, H. and Dupuis, P. (2001). *Numerical Methods for Stochstic Control Problems in Continuous Time*, volume 24 of *Stochastic Modelling and Applied Probability*. Springer, New York, second edition.
14. Lundberg, F. (1903). *I. Approximerad framstallning af sannolikhetsfunktionen: II. Aterforsakring af kollektivrisker*, Uppsala.
15. Milevsky, M., Moore, K., Young, V. R., (2006). Asset allocation and annuity-purchase strategies to minimize the probability of financial ruin, *Mathematical Finance*, 16: 647–671.
16. Reddemann, S., Basse, T., von der Schulenburg, and Johann-Matthias G. (2010). On the Impact of the Financial Crisis on the Dividend Policy of the European Insurance Industry *Geneva Papers on Risk and Insurance: Issues and Practice*, 35(1): 53–62.

17. Richard, S., (1975). Optimal Consumption, portfolio and life insurance rules for an uncertain lived individual in a continuous time model, *Journal of Financial Economics*, 2: 187–203.
18. Schmidli, H., (2008). *Stochastic Control in Insurance*, Springer Verlag.
19. Sotomayor, L. and Cadenillas, A. (2011). Classical and Singular Stochastic Control for the Optimal Dividend Policy When There Is Regime Switching. *Insurance: Mathematics and Economics*, 48(3), 344–54.
20. Vanderhei, J. and Copeland, C., (2003). *Can America afford tomorrows retirees: Results from the EBRI-ERF retirement security projection model*, Issue Brief of the Employee Benefit Research Institute, www.ebri.org.
21. Wei, J., Yang, H. and Wang, R. (2010). Classical and impulse control for the optimization of dividend and proportional reinsurance policies with regime switching. *Journal of Optimization Theory and Applications*, 147(2).
22. Yang, H. and Yin, G. (2004). Ruin probability for a model under Markovian switching regime, In T.L. Lai, H. Yang, and S.P. Yung, editors, *Probability, Finance and Insurance*, pages 206–217. World Scientific, River Edge, NJ.
23. Yaari,M. (1965). Uncertain lifetime, life insurance and the theory of the consumer, *Review of Economic Studies*, 32: 137–150.
24. Yin, G., Jin, Z., and Yang, H. (2010). Asymptotically optimal dividend policy for regime-switching compound Poisson models *Acta Mathematicae Applicatae Sinica*, 26:529–542.
25. Yin, G. and Zhang, Q. (2005). *Discrete-time Markov Chains: Two-time-scale Methods and Applications*, Springer, New York, NY.

Chapter 11
Trading a Mean-Reverting Asset with Regime Switching: An Asymptotic Approach

Eunju Sohn and Qing Zhang

11.1 Introduction

This chapter is concerned with mean-reversion trading with regime switching. It is a continuation of the study developed in Zhang and Zhang [15]. In [15], a mean-reversion trading rule was considered. The objective was to buy and sell the asset so as to maximize an overall return. They followed the dynamic programming approach and used the associated HJB equations (quasi-variational inequalities) to characterize the value functions. They showed that the solution to the original optimal stopping problem can be obtained by solving two quasi-algebraic equations. In addition, they obtained sufficient conditions in the form of a verification theorem. Nevertheless, only the basic mean-reversion model with constant equilibrium was considered in [15]. It is important to extend the results to account for more realistic settings. It is the purpose of this chapter to consider the mean-reversion model in which the equilibrium is subject to random jumps governed by a two-state Markov chain and to study the corresponding trading rules.

A mean-reversion model is often used in financial and energy markets to capture price movements that have the tendency to move towards an "equilibrium" level. Studies that support the mean-reversion stock returns can be traced back to the 1930s (see Cowles and Jones [3]) in empirical literature. The research was furthered by many researchers including Fama and French [6], and Gallagher and Taylor [7] among others. In addition to stock markets, mean-reversion models are also used to characterize stochastic volatility (see Hafner and Herwartz [8]) and asset prices in energy markets (see Blanco and Soronow [1] and de Jong and Huisman [4]). See

E. Sohn
Department of Science and Mathematics, Columbia College of Chicago, Chicago, USA
e-mail: esohn@math.uga.edu

Q. Zhang (✉)
Department of Mathematics, University of Georgia, Athens, GA 30602, USA
e-mail: qingz@math.uga.edu

Y. Zeng and S. Wu (eds.), *State-Space Models: Applications in Economics and Finance*, Statistics and Econometrics for Finance 1, DOI 10.1007/978-1-4614-7789-1_11,

also related results in option pricing with a mean-reversion asset by Bos, Ware and Pavlov [2].

Trading rules in financial markets have been studied for many years. For example, an investment capacity expansion/reduction problem was considered in Merhi and Zervos [11]. Under a geometric Brownian motion market model, the authors used the dynamic programming approach and obtained an explicit solution to the singular control problem. A more general diffusion market model was treated by Løkka and Zervos [10] in connection with an optimal investment capacity adjustment problem. More recently, Johnson and Zervos [9] studied an optimal timing of investment problem under a general diffusion market model. The objective was to maximize the expected cash flow by choosing when to enter an investment and when to exit the investment. An explicit analytic solution was obtained in [9]. Recently, Dai et al. [5] provided a theoretical justification of trend following trading. In particular, the underlying stock price was formulated as a geometric Brownian motion with regime switching. Two regimes were considered: the up trend (bull market) and the down trend (bear market). The switching process was modeled as a two-state Markov chain which is not directly observable. The trading decisions were based on current information represented by both the stock price and historical information with the probability in the bull phase conditioning to all available historical price levels as a proxy. Assuming trading one share with a fixed percentage transaction cost, they showed that the strategy that optimizes the discounted expected return is a simple implementable trend following system. This strategy was characterized by two threshold curves for the conditional probability in a bull regime signaling buy and sell, respectively. The main advantage of this approach is that the conditional probability in a bull market can be obtained directly using actual historical stock price data through a differential equation.

In this chapter, we focus on a mean-reversion model in which its equilibrium is subject to random jumps. Such model can be applied to assets with a "staircase" price behavior. We consider trading involving both buying and selling actions. The objective is to buy and sell the underlying asset sequentially in order to maximize a discounted reward function. Slippage cost associated with each transaction is imposed. We assume that a fixed percentage slippage cost is incurred with each transaction. In general, this is a class of challenging problems because a closed-form solution is difficult to obtain. In this chapter, we consider the case in which the underlying Markov chain jumps frequently between its two states. This leads to a class of singular perturbation problems. The idea is to approximate the value functions of the original problem by the value functions of a limiting problem. The limiting problem is easier to solve. The solution of the limiting problem leads to admissible trading rules that are typically as good as the optimal ones for the original problem. There are substantial studies along the line of singular perturbations. We refer the readers to Sethi and Zhang [13] and Yin and Zhang [14] for related literature. In this chapter, we study the problem using the dynamic programming approach and establish the associated HJB equations (quasi-variational inequalities) for the value functions. Following a viscosity solution approach, we establish asymptotic properties of the value functions. Then using a numerical example, we show how the

solution for the limiting problem can be used to construct a set of trading rules for the original problem.

This chapter is organized as follows. In Sect. 11.2, we formulate the problem under consideration. In Sect. 11.3, we study properties of the value functions and the associated HJB equations. In Sect. 11.4, we provide asymptotic properties of the value functions and describe the corresponding limiting problem. In Sect. 11.5, we demonstrate further related approximation schemes. A numerical example is given in Sect. 11.6 in which the closed-form solution obtained in [15] is used to construct a trading rule for the original problem. The performance of the trading rule is provided in this example. Finally, some concluding remarks are provided in Sect. 11.7. Some technical definitions and assumption verification details are given in Appendix.

11.2 Problem Formulation

Let $X_t \in \mathbb{R}$ denote a mean-reverting diffusion with regime-switching governed by

$$dX_t = a(b(\alpha_t) - X_t)dt + \sigma(\alpha_t)dW_t,\ X_0 = x, \tag{11.1}$$

where $a > 0$ is the rate of reversion, $b(j)$, $j = 1,2$, is the equilibrium level for each state, $\sigma(j) > 0$, $j = 1,2$, is the volatility, $\alpha_t \in \{1,2\}$ is a two-state Markov chain, and W_t is a standard Brownian motion. In this chapter, we assume that α_t and W_t are independent.

Let $h(x)$ be a smooth function. We consider the model in which the asset price is given by $S_t = h(X_t)$. For example, the function $h(x) = e^x$ is used in Zhang and Zhang [15]. In this chapter, we consider $h(x)$ that equals e^x except when x is large. The main reason for specifying $h(x)$ is to facilitate subsequent analysis without affecting much of the applicability.

Let

$$0 \le \phi_1 \le \psi_1 \le \phi_2 \le \psi_2 \le \cdots \tag{11.2}$$

denote a sequence of stopping times. A buying decision is made at ϕ_k and a selling decision at ψ_k, $k = 1,2,\ldots$.

We consider the case that the net position at any time can be either flat (no stock holding) or long (with one share of stock holding). Let $i = 0,1$ denote the initial net position. If initially the net position is long ($i = 1$), then one should sell the stock before acquiring a share. The corresponding sequence of stopping times is denoted by $\Lambda_1 = (\psi_1, \phi_2, \psi_2, \phi_3, \ldots)$. Likewise, if initially the net position is flat ($i = 0$), then one should first buy a stock before selling a share. The corresponding sequence of stopping times is denoted by $\Lambda_0 = (\phi_1, \psi_1, \phi_2, \psi_2, \ldots)$.

In addition, we consider the problem with at most N round trips of trading. We use the notation $\Lambda_1^n = (\psi_1, \phi_2, \psi_2, \phi_3, \ldots, \phi_n, \psi_n)$ and $\Lambda_0^n = (\phi_1, \psi_1, \phi_2, \psi_2, \ldots, \phi_n, \psi_n)$ to label the corresponding stopping times limited to n round trips for $n = 0,1,\ldots,N$.

Let $0 < K < 1$ denote the percentage of slippage (or commission) per transaction. Given the initial states $X_0 = x$, $\alpha_0 = \alpha$, and initial net position $i = 0, 1$, the reward functions of the decision sequences $\{\Lambda_i^n, n = 0, 1, \ldots, N\}$ are given as follows:

$$J_i^n(x, \alpha, \Lambda_i^n) = \begin{cases} E\left\{ \sum_{k=1}^{n} \left[e^{-\rho \psi_k} S_{\psi_k}(1-K) - e^{-\rho \phi_k} S_{\phi_k}(1+K) \right] \right\}, & \text{if } i = 0, \\ E\left\{ e^{-\rho \psi_1} S_{\psi_1}(1-K) \right. \\ \quad \left. + \sum_{k=2}^{n} \left[e^{-\rho \psi_k} S_{\psi_k}(1-K) - e^{-\rho \phi_k} S_{\phi_k}(1+K) \right] \right\}, & \text{if } i = 1, \end{cases} \tag{11.3}$$

where $\rho > 0$ is the discount factor.

For $i = 0, 1$ and $n = 0, 1, \ldots, N$, let $V_i^n(x, \alpha)$ denote the value functions with the initial state $(X_0, \alpha_0) = (x, \alpha)$ and initial net positions $i = 0, 1$. That is,

$$V_i^n(x, \alpha) = \sup_{\Lambda_i^n} J_i^n(x, \alpha, \Lambda_i^n). \tag{11.4}$$

Remark 11.1. In (29), we allow the equalities, i.e., one is allowed to buy and sell at the same time. Nevertheless, owing to the existence of the positive slippage cost K, simultaneous buying and selling only cause negative returns and therefore are automatically ruled out by optimality conditions.

Let $Q = (q_{ij})$ denote the generator of α_t and let $\mathscr{A}$ denote the generator of (X_t, α_t), i.e.,

$$\mathscr{A} f(x, \alpha) = a(b(\alpha) - x) \frac{\partial f(x, \alpha)}{\partial x} + \frac{\sigma^2(\alpha)}{2} \frac{\partial^2 f(x, \alpha)}{\partial x^2} + Q f(x, \cdot)(\alpha),$$

where $Q f(x, \cdot)(\alpha) = q_{\alpha 1} f(x, 1) + q_{\alpha 2} f(x, 2)$, $\alpha = 1, 2$.

In Fig. 11.1, a sample path of (X_t, α_t) is provided. The picture was generated using the Monte Carlo method with

$$a = 0.8, b(1) = 3, b(2) = 1, \sigma(1) = 0.7, \sigma(2) = 0.3, Q = \begin{pmatrix} -0.91 & 0.91 \\ 0.62 & -0.62 \end{pmatrix}, X_0 = 1.$$

It is clear from Fig. 11.1, when $\alpha_t = 1$, the equilibrium $b(1) = 3$ serves as an attractor for X_t pulling it upwards; when α_t switched to 2, the new equilibrium $b(2) = 1$ pulls X_t downwards and so on.

As mentioned in the introduction, a closed-form solution to the problem is difficult to obtain. In this chapter, we consider the case in which the Markov chain jumps frequently between its two states. We aim at the corresponding asymptotic properties. In particular, we consider case where the generator has the following form:

$$Q^\varepsilon = \frac{1}{\varepsilon} \widetilde{Q} + \widehat{Q} = \frac{1}{\varepsilon} \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix} + \begin{pmatrix} -\lambda_1 & \lambda_1 \\ \mu_1 & -\mu_1 \end{pmatrix},$$

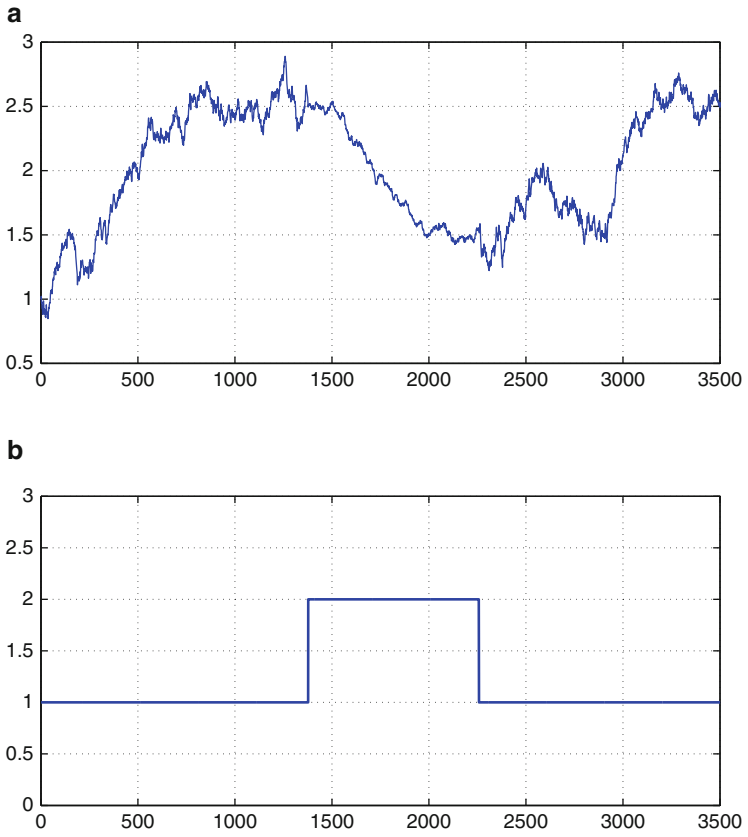

Fig. 11.1 Mean-reversion with regime switching **(a)** A sample path of X_t, **(b)** a sample path of α_t

where $\varepsilon > 0$ is a small parameter, and λ, μ, λ_1, and μ_1 are positive constants. We study the convergence of the problem as $\varepsilon \to 0$. For related Markov models in connection with manufacturing systems, see Sethi and Zhang [13].

Remark 11.2. The Markov chain α_t generated by Q^ε represents the regime of the underlying market. We focus on the market with frequent regime changes in α_t. Such a scenario often arises in a prolonged sideways market such as Dow Jones Industrial Average during the 1960s and 1980s. Its behavior can be captured by our regime-switching model with a relatively small ε. In this chapter, we aim at models with a not-so-small ε and construct near optimal trading rules from the optimal solution of the corresponding limiting problem as $\varepsilon \to 0$. A major advantage of our approach is that one does not have to identify the state of α_t, which is difficult during the period when it is changing rapidly.

The corresponding Markov chain will be labeled as α_t^ε. Similarly, we use X_t^ε for X_t, S_t^ε for S_t, $J_i^{n,\varepsilon}$ for J_i^n, and $V_i^{n,\varepsilon}$ for V_i^n from now on to emphasize the dependence

on ε. Using this notation, the optimal trading problem $\mathscr{P}^{N,\varepsilon}$ can be written as follows:

$$\mathscr{P}^{N,\varepsilon}:\begin{cases}\max & J_i^{n,\varepsilon}(x,\alpha,\Lambda_i^n)\\ & \quad=\begin{cases}E\left\{\sum_{k=1}^{n}\left[e^{-\rho\psi_k}S^{\varepsilon}_{\psi_k}(1-K)-e^{-\rho\phi_k}S^{\varepsilon}_{\phi_k}(1+K)\right]\right\}, & \text{if } i=0,\\ E\left\{e^{-\rho\psi_1}S^{\varepsilon}_{\psi_1}(1-K)\right.\\ \qquad\left.+\sum_{k=2}^{n}\left[e^{-\rho\psi_k}S^{\varepsilon}_{\psi_k}(1-K)-e^{-\rho\phi_k}S^{\varepsilon}_{\phi_k}(1+K)\right]\right\}, & \text{if } i=1,\end{cases}\\ \text{s.t.} & dX_t^{\varepsilon}=a(b(\alpha_t^{\varepsilon})-X_t^{\varepsilon})dt+\sigma(\alpha_t^{\varepsilon})dW_t,\ X_0^{\varepsilon}=x,\\ \text{value fn} & V_i^{n,\varepsilon}(x,\alpha)=\sup_{\Lambda_i^n}J_i^{n,\varepsilon}(x,\alpha,\Lambda_i^n), \qquad n=0,1,\ldots,N,\end{cases}$$

Note that the sequence $\Lambda_0^n=(\phi_1,\psi_1,\ldots,\phi_n,\psi_n)$ can be regarded as a combination of a buy at ϕ_1 and then followed by the sequence of stopping times $\Lambda_1^n=(\psi_1,\phi_2,\psi_2,\ldots,\phi_n,\psi_n)$. In view of this, we have

$$\begin{aligned}V_0^{n,\varepsilon}(x,\alpha)&\geq J_0^{n,\varepsilon}(x,\alpha,\Lambda_0^n)\\ &=E\left\{e^{-\rho\psi_1}S^{\varepsilon}_{\psi_1}(1-K)+\sum_{k=2}^{n}\left[e^{-\rho\psi_k}S^{\varepsilon}_{\psi_k}(1-K)-e^{-\rho\phi_k}S^{\varepsilon}_{\phi_k}(1+K)\right]\right\}\\ &\quad -Ee^{-\rho\phi_1}S^{\varepsilon}_{\phi_1}(1+K)\\ &=J_1^{n,\varepsilon}(X^{\varepsilon}_{\phi_1},\alpha,\Lambda_1^n)-Ee^{-\rho\phi_1}S^{\varepsilon}_{\phi_1}(1+K).\end{aligned}$$

In particular, setting $\phi_1=0$ (recall that $S_t^{\varepsilon}=h(X_t^{\varepsilon})$), we obtain the inequality

$$V_0^{n,\varepsilon}(x,\alpha)\geq V_1^{n,\varepsilon}(x,\alpha)-h(x)(1+K). \tag{11.5}$$

Similarly, we can show that

$$V_1^{n,\varepsilon}(x,\alpha)\geq V_0^{n-1,\varepsilon}(x,\alpha)+h(x)(1-K). \tag{11.6}$$

Formally, the associated HJB equations should have the form:

$$\begin{aligned}&\min\left\{\rho V_0^{n,\varepsilon}(x,\alpha)-\mathscr{A}V_0^{n,\varepsilon}(x,\alpha),\ V_0^{n,\varepsilon}(x,\alpha)-V_1^{n,\varepsilon}(x,\alpha)+h(x)(1+K)\right\}=0,\\ &\min\left\{\rho V_1^{n,\varepsilon}(x,\alpha)-\mathscr{A}V_1^{n,\varepsilon}(x,\alpha),\ V_1^{n,\varepsilon}(x,\alpha)-V_0^{n-1,\varepsilon}(x,\alpha)-h(x)(1-K)\right\}=0,\end{aligned} \tag{11.7}$$

for $n=1,2,\ldots,N$ and $\alpha=1,2$. Here, we follow the convention that $V_0^{0,\varepsilon}(x,\alpha)=0$.

Next, we impose conditions on $h(x)$.

Assumption. $h(x)$, $h'(x)$, $xh'(x)$, and $h''(x)$ are bounded and Lipschitz.

Example 11.1. An immediate example satisfying the above conditions can be given as follows. Let $h_0(x) = \begin{cases} e^x & \text{for } x \le M, \\ e^M & \text{for } x > M, \end{cases}$ for a fixed M. Take $h(x)$ to be the convolution of h_0 with the kernel $\Psi(x) = (1/\sqrt{2\pi})e^{-x^2/2}$. Validation of these conditions is provided in Appendix.

Under these assumptions, we can show, following a similar approach as in Sethi and Zhang [13,Chap. 8], that $V_i^{n,\varepsilon}(x,\alpha)$ are the viscosity solutions (see the definition given in Appendix) of the HJB equations (34).

In this chapter, C (and C_i) are generic positive constants with convention $C+C = C$ and $CC = C$, etc.

11.3 Properties of the Value Functions

In this section, we consider the basic properties of the value functions. In particular, we establish the boundedness and Lipschitz continuity of these functions.

Lemma 11.1. *There exists a constant C_0 such that*

$$0 \le V_i^{n,\varepsilon}(x,\alpha) \le C_0,$$

for $\varepsilon > 0$, $x \in \mathbb{R}$, $\alpha = 1,2$, $i = 0,1$, and $n = 0,1,\ldots,N$.

Proof. In view of the definition of $V_i^{n,\varepsilon}(x,\alpha)$, it is clear that they are nonnegative. It remains to establish their upper bounds. Let

$$F(x,\alpha) = a(b(\alpha)-x)h'(x) + \frac{\sigma^2(\alpha)}{2}h''(x) - \rho h(x).$$

Then, using Dynkin's formula, we have

$$Ee^{-\rho\psi_k}S^{\varepsilon}_{\psi_k} - Ee^{-\rho\phi_k}S^{\varepsilon}_{\phi_k} = E\int_{\phi_k}^{\psi_k} e^{-\rho s}F(X^{\varepsilon}_s,\alpha_s)ds. \tag{11.8}$$

It is easy to see that the function $F(x,\alpha)$ is bounded above on $\mathbb{R}$ by the boundedness assumptions on $h(x)$. Let C be an upper bound of F. It follows that

$$Ee^{-\rho\psi_k}S^{\varepsilon}_{\psi_k} - Ee^{-\rho\phi_k}S^{\varepsilon}_{\phi_k} \le CE\int_{\phi_k}^{\psi_k} e^{-\rho t}dt. \tag{11.9}$$

Using the definition of $J_0^{n,\varepsilon}(x,\alpha,\Lambda_0^n)$, we have

$$\begin{aligned} J_0^{n,\varepsilon}(x,\alpha,\Lambda_0^n) &\le \sum_{k=1}^{n}\left(Ee^{-\rho\psi_k}S_{\psi_k}^{\varepsilon} - Ee^{-\rho\phi_k}S_{\phi_k}^{\varepsilon}\right) \\ &\le \sum_{k=1}^{n} CE\int_{\phi_k}^{\psi_k} e^{-\rho t}dt \\ &\le C\int_0^{\infty} e^{-\rho t}dt := C_0. \end{aligned}$$

This implies that $0 \le V_0^{n,\varepsilon}(x,\alpha) \le C_0$.

Similarly, letting $C_h = \sup|h(x)|$, we have the inequalities

$$J_1^{n,\varepsilon}(x,\alpha,\Lambda_1^n) \le C_0 + Ee^{-\rho\psi_1}h(X_{\psi_1}^{\varepsilon})(1-K) \le C_0 + C_h(1-K) := C_0.$$

Therefore, $0 \le V_1^{n,\varepsilon}(x,\alpha) \le C_0$. This completes the proof. □

Lemma 11.2. *$V_i^{n,\varepsilon}(x,\alpha)$ are Lipschitz, i.e., there exists C_0 such that*

$$|V_i^{n,\varepsilon}(x_1,\alpha) - V_i^{n,\varepsilon}(x_2,\alpha)| \le C_0|x_1 - x_2|.$$

for $\varepsilon > 0$, $x_1, x_2 \in \mathbb{R}$, $\alpha = 1,2$, $i = 0,1$, and $n = 0,1,\ldots,N$.

Proof. Given x_1 and x_2, let X_t^1 and X_t^2 be solutions of (28) with $X_0^1 = x_1$ and $X_0^2 = x_2$, respectively. We claim that: There exists an constant C_0 such that for any stopping time τ,

$$\left|E\left[e^{-\rho\tau}(h(X_\tau^1) - h(X_\tau^2)\right]\right| \le C_0|x_1 - x_2|. \tag{11.10}$$

Let

$$G(x,y,\alpha) = ab(\alpha)[h'(x) - h'(y)] - a[xh'(x) - yh'(y)] + \frac{\sigma^2(\alpha)}{2}[h''(x) - h''(y)] - \rho[h(x) - h(y)].$$

Then, using the Lipschitz assumptions on $h(x)$, we can see that

$$|G(x,y,\alpha)| \le C_0|x - y|,$$

for some constant C_0. Then, applying Dynkin's formula, we have

$$E\left[e^{-\rho\tau}(h(X_t^1) - h(X_t^2))\right] = h(x_1) - h(x_2) + E\int_0^{\tau} e^{-\rho t}G(X_t^1, X_t^2, \alpha_t)dt.$$

It follows that

$$\begin{aligned} \left|E\left[e^{-\rho\tau}(h(X_t^1) - h(X_t^2))\right]\right| &\le |h(x_1) - h(x_2)| + E\int_0^{\infty} e^{-\rho t}|G(X_t^1, X_t^2, \alpha_t)|dt \\ &\le C_0|x_1 - x_2| + C_0E\int_0^{\infty} e^{-\rho t}|X_t^1 - X_t^2|dt. \end{aligned}$$

Note that

$$X_t^1 - X_t^2 = x_1 - x_2 - a\int_0^t (X_s^1 - X_s^2)ds.$$

Therefore, $X_t^1 - X_t^2 = (x_1 - x_2)e^{-at}$. In view of this, we have

$$\left|E\left[e^{-\rho\tau}(h(X_t^1) - h(X_t^2))\right]\right| \le C_0|x_1 - x_2| + C_0 E\int_0^\infty e^{-\rho t}|x_1 - x_2|e^{-at}dt = C_0|x_1 - x_2|,$$

which proves the claim. Using this inequality, for any given Λ_i^n, it is easy to see that

$$|J_i^{n,\varepsilon}(x_1, \alpha, \Lambda_i^n) - J_i^{n,\varepsilon}(x_2, \alpha, \Lambda_i^n)| \le C_0|x_1 - x_2|. \qquad \square$$

11.4 Asymptotic Properties

In this section, we study the asymptotic properties of the value functions as $\varepsilon \to 0$. We first characterize the limiting problem and then establish the desired convergence.

Lemma 11.3. *For each (x, α), if for some subsequence of ε, $V_i^{n,\varepsilon}(x, \alpha) \to V_i^{n,0}(x, \alpha)$, then $V_i^{n,0}(x, \alpha) = V_i^{n,0}(x)$.*

Proof. Let τ^ε denote the first jump time of α_t^ε. Then $\tau^\varepsilon \to 0$ a.s. as $\varepsilon \to 0$. Following the dynamic programming principle, we have

$$V_i^{n,\varepsilon}(x, \alpha) \ge Ee^{-\rho\tau^\varepsilon} V_i^{n,\varepsilon}(X_{\tau^\varepsilon}, \alpha_{\tau^\varepsilon}).$$

If $\alpha = 1$, then sending $\varepsilon \to 0$, we have

$$V^{n,0}(x, 1) \ge V^{n,0}(x, 2).$$

Similarly,

$$V^{n,0}(x, 2) \ge V^{n,0}(x, 1).$$

Therefore, $V^{n,0}(x, 1) = V^{n,0}(x, 2)$. $\square$

Let (ν_1, ν_2) denote the equilibrium distribution corresponding to $\widetilde{Q}$, i.e.,

$$(\nu_1, \nu_2) = \left(\frac{\mu}{\lambda + \mu}, \frac{\lambda}{\lambda + \mu}\right).$$

so that $(\nu_1, \nu_2)\widetilde{Q} = (0, 0)$. Let $\overline{X}_t$ denote the corresponding mean-reversion process with mean $\overline{b} = \nu_1 b(1) + \nu_2 b(2)$ and volatility $\overline{\sigma} = \sqrt{\nu_1\sigma^2(1) + \nu_2\sigma^2(2)}$. The stock price driven by $\overline{X}_t$ is denoted by $\overline{S}_t = h(\overline{X}_t)$.

Given a sequence of $\sigma\{W_r : r \le t\}$ measurable stopping times

$$0 \le \phi_1 \le \psi_1 \le \phi_2 \le \psi_2 \le \cdots,$$

one can define the set of stopping times Λ_i^n as before for $n = 0, 1, \ldots, N$ and $i = 0, 1$. The limiting problem $\mathscr{P}^{N,0}$ can be defined as follows:

$$\mathscr{P}^{N,0}: \begin{cases} \max & \overline{J}_i^n(x, \Lambda_i^n) \\ & = \begin{cases} E\left\{ \sum_{k=1}^{n} \left[e^{-\rho\psi_k} \overline{S}_{\psi_k}(1-K) - e^{-\rho\phi_k} \overline{S}_{\phi_k}(1+K) \right] \right\}, & \text{if } i = 0, \\ E\left\{ e^{-\rho\psi_1} \overline{S}_{\psi_1}(1-K) \right. \\ \qquad \left. + \sum_{k=2}^{n} \left[e^{-\rho\psi_k} \overline{S}_{\psi_k}(1-K) - e^{-\rho\phi_k} \overline{S}_{\phi_k}(1+K) \right] \right\}, & \text{if } i = 1, \end{cases} \\ \text{s.t.} & d\overline{X}_t = a(\overline{b} - \overline{X}_t)dt + \overline{\sigma} dW_t,\ \overline{X}_0 = x, \\ \text{value fn} & \overline{V}_i^n(x) = \sup_{\Lambda_i^n} \overline{J}_i(x, \Lambda_i^n). \end{cases}$$

Let $\overline{\mathscr{A}}$ denote the generator of $\overline{X}_t$, i.e.,

$$\overline{\mathscr{A}} f(x) = a(\overline{b} - x) \frac{df(x)}{dx} + \frac{\overline{\sigma}^2}{2} \frac{d^2 f(x)}{dx^2}.$$

The associated HJB equations for the limiting problem should have the form:

$$\begin{aligned} &\min\left\{ \rho \overline{V}_0^n(x) - \overline{\mathscr{A}} \overline{V}_0^n(x),\ \overline{V}_0^n(x) - \overline{V}_1^n(x) + h(x)(1+K) \right\} = 0, \\ &\min\left\{ \rho \overline{V}_1^n(x) - \overline{\mathscr{A}} \overline{V}_1^n(x),\ \overline{V}_1^n(x) - \overline{V}_0^{n-1}(x) - h(x)(1-K) \right\} = 0, \end{aligned} \tag{11.11}$$

for $n = 1, 2, \ldots, N$.

The definition of viscosity solution of the above HJB equations is also given in Appendix. We can show the following lemma, where the uniqueness can be obtained along the line of Pham [12].

Lemma 11.4. *$\overline{V}_i^n(x)$ are the unique viscosity solutions of the HJB equations (38).*

Next, we give the main result of this chapter. We show that the value functions of the original problem converge to those of the limiting problem. This suggests that the optimal solution of the limiting problem can be used to construct a trading rule for the original problem. We refer the readers to Sethi and Zhang [13] for similar approach in connection with manufacturing systems.

Theorem 11.1. *As $\varepsilon \to 0$, we have*

$$V_i^{n,\varepsilon}(x, \alpha) \to \overline{V}_i^n(x),$$

for $n = 0, 1, \ldots, N$, $i = 0, 1$, $x \in \mathbb{R}$, and $\alpha = 1, 2$.

Proof. Recall the Lipschitz properties of $V_i^{n,\varepsilon}$ in Lemma 11.2. In view of the Arzela–Ascoli Theorem, for each sequence of $\{\varepsilon \to 0\}$, there exists a further subsequence (still indexed by ε) such that $V_i^{n,\varepsilon}(x,\alpha)$ converges. Denote the limit by $V_i^{n,0}(x,\alpha)$. Then by Lemma 11.3, $V_i^{n,0}(x,\alpha) = V_i^{n,0}(x)$. It suffices to show that $V_i^{n,0}(x)$ is a viscosity solution of (38) because Lemma 11.4 implies that $V_i^{n,0}(x) = \overline{V}_i^n(x)$. Following Lemma A.25 in Yin and Zhang [14], for each $i = 0, 1$, take a function $\phi_i(x) \in C^2$ such that $V_i^{n,0}(x) - \phi_i(x)$ has a strictly local maximum at any given x_0 in a neighborhood $N(x_0)$. Choose $x_{i,\alpha}^{n,\varepsilon} \in N(x_0)$ such that

$$V_i^{n,\varepsilon}(x_{i,\alpha}^{n,\varepsilon},\alpha) - \phi_i(x_{i,\alpha}^{n,\varepsilon}) = \max_{x\in N(x_0)} \{V_i^{n,\varepsilon}(x,\alpha) - \phi_i(x)\}.$$

Then, $x_{i,\alpha}^{n,\varepsilon} \to x_0$, as $\varepsilon \to 0$. First, fix $i = 0$. We are to show the following inequality:

$$\min\left\{\rho V_0^{n,0}(x_0) - \overline{\mathscr{A}}\phi_0(x_0), V_0^{n,0}(x_0) - V_1^{n,0}(x_0) + h(x_0)(1+K)\right\} \le 0. \quad (11.12)$$

If

$$V_0^{n,0}(x_0) - V_1^{n,0}(x_0) + h(x_0)(1+K) \le 0,$$

then (39) holds. Otherwise,

$$V_0^{n,0}(x_0) - V_1^{n,0}(x_0) + h(x_0)(1+K) > 0.$$

Then there exists $N_0(x_0) \subset N(x_0)$ such that

$$V_0^{n,\varepsilon}(x) - V_1^{n,\varepsilon}(x) + h(x)(1+K) > 0$$

on $N_0(x_0)$ for ε small enough. Recall that $V_i^{n,\varepsilon}$ is a viscosity solution to (34). $V_0^{n,\varepsilon}(x,\alpha)$ must satisfy (45). Necessarily,

$$\rho V_0^{n,\varepsilon}(x_{0,\alpha}^{n,\varepsilon},\alpha) - \mathscr{A}^{\phi_0} V_0^{n,\varepsilon}(x_{0,\alpha}^{n,\varepsilon},\alpha) \le 0,$$

for $\alpha = 1, 2$.

It follows that

$$\nu_1(\rho V_0^{n,\varepsilon}(x_{0,1}^{n,\varepsilon},1) - \mathscr{A}^{\phi_0} V_0^{n,\varepsilon}(x_{0,1}^{n,\varepsilon},1)) + \nu_2(\rho V_0^{n,\varepsilon}(x_{0,2}^{n,\varepsilon},2) - \mathscr{A}^{\phi_0} V_0^{n,\varepsilon}(x_{0,2}^{n,\varepsilon},2)) \le 0. \quad (11.13)$$

Note that

$$\begin{aligned}
&\nu_1\left(\frac{\lambda}{\varepsilon}\right)(V_0^{n,\varepsilon}(x_1^\varepsilon,2) - V_0^{n,\varepsilon}(x_1^\varepsilon,1)) + \nu_2\left(\frac{\mu}{\varepsilon}\right)(V_0^{n,\varepsilon}(x_2^\varepsilon,1) - V_0^{n,\varepsilon}(x_2^\varepsilon,2)) \\
&\quad \le \nu_1\left(\frac{\lambda}{\varepsilon}\right)[V_0^{n,\varepsilon}(x_2^\varepsilon,2) - \phi(x_2^\varepsilon) - (V_0^{n,\varepsilon}(x_1^\varepsilon,1) - \phi(x_1^\varepsilon))] \\
&\qquad + \nu_2\left(\frac{\mu}{\varepsilon}\right)[V_0^{n,\varepsilon}(x_1^\varepsilon,1) - \phi(x_1^\varepsilon) - (V_0^{n,\varepsilon}(x_2^\varepsilon,2) - \phi(x_2^\varepsilon))] = 0.
\end{aligned} \quad (11.14)$$

Using this inequality and sending $\varepsilon \to 0$ in (40) to obtain $\rho V_0^{n,0}(x_0) - \overline{\mathscr{A}}\phi_0(x_0) \le 0$, which yields (39). Similarly, we can show

$$\min\left\{\rho V_1^{n,0}(x_0) - \overline{\mathscr{A}}\phi_1(x_0), V_0^{n,0}(x_0) - V_0^{n-1,0}(x_0) - h(x_0)(1-K)\right\} \le 0.$$

Thus, $V_i^{n,0}(x)$ is a viscosity subsolution to (38).

To show that $V_i^{n,0}(x)$ is a viscosity supersolution to (38), note that

$$\min\left\{\rho V_0^{n,\varepsilon}(x_0,\alpha_0) - \mathscr{A}^{\Psi} V_0^{n,\varepsilon}(x_0,\alpha_0),\ V_0^{n,\varepsilon}(x_0,\alpha_0) - V_1^{n,\varepsilon}(x_0,\alpha_0) + h(x)(1+K)\right\} \ge 0$$

implies

$$V_0^{n,0}(x_0) - V_1^{n,0}(x_0) + h(x)(1+K) \ge 0.$$

Moreover, following similar argument as in (41), we can show that

$$\rho V_0^{n,0}(x_0) - \overline{\mathscr{A}}\psi_0(x_0) \ge 0.$$

Hence,

$$\min\left\{\rho V_0^{n,0}(x_0) - \overline{\mathscr{A}}\psi_0(x_0),\ V_0^{n,0}(x_0) - V_1^{n,0}(x_0) + h(x)(1+K)\right\} \ge 0.$$

Similarly, we can show the inequality with $i = 1$. Therefore $V_i^{n,0}(x)$ is a viscosity supersolution. This completes the proof. □

11.5 Further Approximations

In this section, we show that the value function $\overline{V}_i^n(x)$ can be further approximated by taking N to be very large and $h(x)$ to be very close to e^x. In this case, we can use the closed-form solution obtained in Zhang and Zhang [15] to come up with an approximate solution for the original problem.

Recall the definition of Λ_i and its N-th round trip truncation Λ_i^N. Let

$$\overline{J}_i(x,\Lambda_i) = \limsup_{N\to\infty} \overline{J}_i^N(x,\Lambda_i^N)$$

and $\overline{V}_i(x) = \sup_{\Lambda_i} \overline{J}_i(x,\Lambda_i)$. It is easy to see that

$$\lim_{N\to\infty} \overline{V}_i^N(x) = \overline{V}_i(x).$$

In fact, for each $\delta > 0$, let $\Lambda_{i,\delta}$ be a sequence of stopping times such that $\overline{J}_i(x,\Lambda_{i,\delta}) \ge \overline{V}_i(x) - \delta$. Then, noticing that $\overline{V}_i^N(x)$ is monotonically increasing in N, we have

$$\overline{V}_i(x)-\delta\le \overline{J}_i(x,\Lambda_{i,\delta})=\limsup_{N\to\infty}\overline{J}_i^N(x,\Lambda_{i,\delta}^N)\le \limsup_{N\to\infty}\overline{V}_i^N(x)\le \overline{V}_i(x).$$

Next, we consider approximating e^x by particular choices of $h(x)$. Recall Example 11.1 and the definition of $h_0(x)$. For each $\gamma>0$, let $\Psi_\gamma(x)=(1/\gamma)\Psi(x/\gamma)$ and $h_\gamma(x)$ be the convolution of h_0 and Ψ_γ. Then, $h_\gamma(x)\to h_0(x)$ as $\gamma\to 0$ for all x. Therefore, we can approximate e^x by $h_\gamma(x)$ by choosing a small enough γ on $[-M,M]$.

In view of these, the original problem with a large N can be approximated by the limiting problem with a large N and a large M. In the next section, we study a numerical example demonstrating how these approximations work.

11.6 A Numerical Example

The optimal trading rule in the limiting problem with $N=\infty$ and $h(x)=e^x$ was treated in Zhang and Zhang [15]. The main result can be summarized as follows.

Lemma 11.5. *Let (x_1^*,x_2^*) be a pair satisfying the following conditions:*

$$x_1^*\le \frac{1}{a}\left(\frac{\overline{\sigma}^2}{2}+a\overline{b}-\rho\right)\le x_2^*,\ x_2^*-x_1^*>\log\left(\frac{1+K}{1-K}\right),$$

and

$$\begin{aligned}&\begin{pmatrix}\displaystyle\int_0^\infty \eta(t)e^{-\kappa(\overline{b}-x_1^*)t}dt & -\displaystyle\int_0^\infty \eta(t)e^{\kappa(\overline{b}-x_1^*)t}dt\\ \displaystyle\int_0^\infty t\eta(t)e^{-\kappa(\overline{b}-x_1^*)t}dt & \displaystyle\int_0^\infty t\eta(t)e^{\kappa(\overline{b}-x_1^*)t}dt\end{pmatrix}^{-1}\begin{pmatrix}e^{x_1^*}(1+K)\\ e^{x_1^*}(1+K)/\kappa\end{pmatrix}\\ &=\begin{pmatrix}\displaystyle\int_0^\infty \eta(t)e^{-\kappa(\overline{b}-x_2^*)t}dt & -\displaystyle\int_0^\infty \eta(t)e^{\kappa(\overline{b}-x_2^*)t}dt\\ \displaystyle\int_0^\infty t\eta(t)e^{-\kappa(\overline{b}-x_2^*)t}dt & \displaystyle\int_0^\infty t\eta(t)e^{\kappa(\overline{b}-x_2^*)t}dt\end{pmatrix}^{-1}\begin{pmatrix}e^{x_2^*}(1-K)\\ e^{x_2^*}(1-K)/\kappa\end{pmatrix}\end{aligned}\tag{11.15}$$

where $\kappa=\sqrt{2a}/\overline{\sigma}$ and $\eta(t)=t^{(\rho/a)-1}\exp\left(-t^2/2\right)$.

Let

$$\begin{cases}\overline{V}_0(x)=\begin{cases}C_2^*\displaystyle\int_0^\infty \eta(t)e^{\kappa(\overline{b}-x)t}dt & \text{if } x\ge x_1^*,\\ C_1^*\displaystyle\int_0^\infty \eta(t)e^{-\kappa(\overline{b}-x)t}dt-e^x(1+K) & \text{if } x<x_1^*,\end{cases}\\ \overline{V}_1(x)=\begin{cases}C_1^*\displaystyle\int_0^\infty \eta(t)e^{-\kappa(\overline{b}-x)t}dt & \text{if } x<x_2^*,\\ C_2^*\displaystyle\int_0^\infty \eta(t)e^{\kappa(\overline{b}-x)t}dt+e^x(1-K) & \text{if } x\ge x_2^*\end{cases}\end{cases}\tag{11.16}$$

with

$$\begin{pmatrix} C_1^* \\ C_2^* \end{pmatrix} = \begin{pmatrix} \int_0^\infty \eta(t)e^{-\kappa(\overline{b}-x_1^*)t}dt & -\int_0^\infty \eta(t)e^{\kappa(\overline{b}-x_1^*)t}dt \\ \int_0^\infty t\eta(t)e^{-\kappa(\overline{b}-x_1^*)t}dt & \int_0^\infty t\eta(t)e^{\kappa(\overline{b}-x_1^*)t}dt \end{pmatrix}^{-1} \begin{pmatrix} e^{x_1^*}(1+K) \\ e^{x_1^*}(1+K)/\kappa \end{pmatrix}.$$

If, on the interval (x_1^*, x_2^*), the following inequalities hold

$$e^x(1-K) \le \overline{V}_1(x) - \overline{V}_0(x) \le e^x(1+K),$$

then buying when $x \le x_1^*$ and selling when $x \ge x_2^*$ is optimal.

Example 11.2. In this example, we take

$$a = 0.8, b(1) = 3, b(2) = 1, \sigma(1) = 0.7, \sigma(2) = 0.3,$$

$$\lambda = 0.09, \mu = 0.06, \lambda_1 = 0.01, \mu_1 = 0.02, \rho = 0.5, \text{ and } K = 0.01.$$

Then, $(\nu_1, \nu_2) = (2/5, 3/5)$, $\overline{b} = 9/5$, and $\overline{\sigma} = 0.5$. We solve (42) to obtain $(x_1^*, x_2^*) = (1.115, 1.455)$ and the value functions $\overline{V}_i(x)$. These functions are plotted in Fig. 11.2. In addition, we vary $\varepsilon = 0.1, 0.01$, and 0.001 and solve the corresponding HJB equations in (34) (using the explicit finite difference method) with $N = \infty$. The value functions V_i^ε are also presented in Fig. 11.2. It is clear in this example that V_i^ε can be approximated by $\overline{V}_i$ when ε is small enough.

Next, we use $(x_1^*, x_2^*) = (1.115, 1.455)$ to construct the following trading rules for the original problem:

$$\begin{cases} \text{Buy:} & \text{if } X_t^\varepsilon \le x_1^*, \\ \text{Sell:} & \text{if } X_t^\varepsilon \ge x_2^*. \end{cases} \tag{11.17}$$

Using these trading rules, we generate the corresponding reward functions with a varying ε and $N = \infty$. In particular, we use Monte Carlo simulations based on (28) and generate 10 K sample paths. The corresponding reward functions with $\varepsilon = 1$, 0.01, and 0.0001 are plotted in Fig. 11.3. Combining these two figures, one can see how the constructed trading rules in (44) work for the original problem.

In general, the control policy obtained via a singular perturbation approach not only work when ε is small but also work for the problem with not-so-small ε. The performance with $\varepsilon = 1$ can be seen in Fig. 11.3 in which the corresponding reward functions are fairly close to the value functions of the limiting problem, and therefore, in view of Fig. 11.2, close to those of the original problem.

11.7 Concluding Remarks

In this chapter, we studied the asymptotic properties of the mean-reverting trading problem. We established the convergence of the value functions and demonstrated

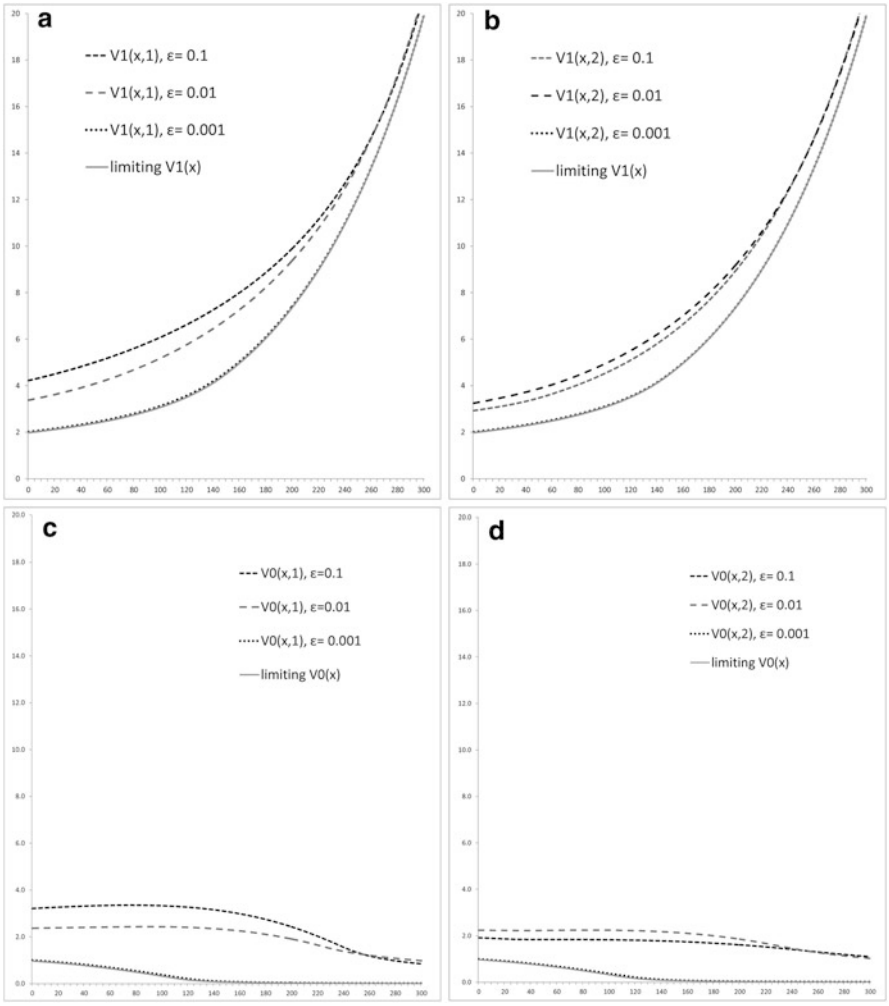

Fig. 11.2 Value function approximation. **(a)** $V_1^{\varepsilon}(x,1)$ and $\overline{V}_1(x)$, **(b)** $V_1^{\varepsilon}(x,2)$ and $\overline{V}_1(x)$, **(c)** $V_0^{\varepsilon}(x,1)$ and $\overline{V}_0(x)$, **(d)** $V_0^{\varepsilon}(x,2)$ and $\overline{V}_0(x)$

how the optimal trading rule for the limiting problem can be used to construct a trading rule for the original problem.

In general, to use an optimal trading rule for the original problem, one needs to determine the mode (or the state of α_t^{ε}). This typically involves nonlinear filtering as in Dai et al. [5]. Nevertheless, in this chapter, we showed that this is not necessary when the jump rates of α_t^{ε} is large because the constructed trading rule does not require the state information of α_t^{ε}.

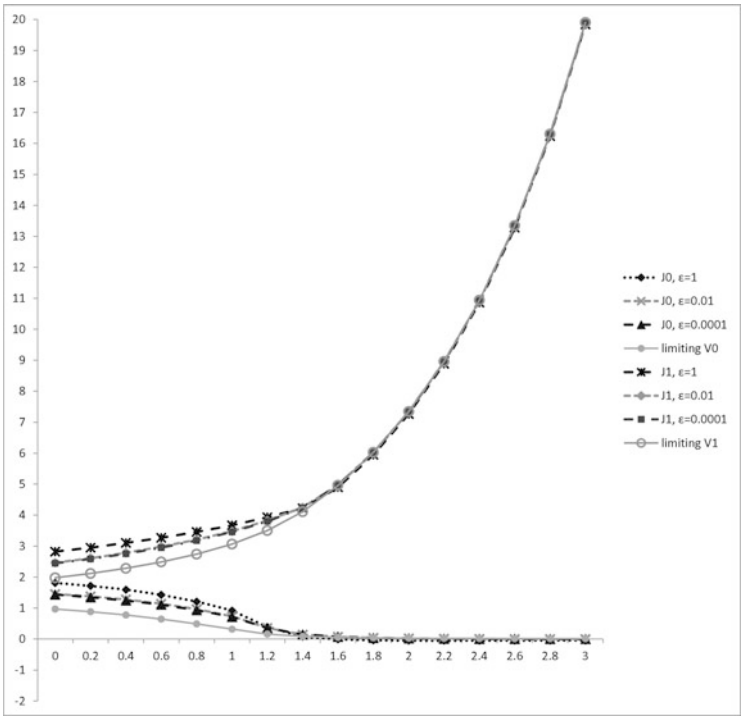

Fig. 11.3 Reward functions under trading rules constructed from that of the limiting problem.

Appendix

In this appendix, we provide the definitions of viscosity solutions of the HJB equations (34) and (38). First, we consider (34). For each $f(x,\alpha)$ and $\phi(x) \in C^2$, let

$$\mathscr{A}^{\phi} f(x,\alpha) = a(b(\alpha)-x)\frac{d\phi(x)}{dx} + \frac{\sigma^2(\alpha)}{2}\frac{d^2\phi(x)}{dx^2} + Qf(x,\cdot)(\alpha).$$

Definition 11.1. $v_i^{n,\varepsilon}(x,\alpha)$ is a *viscosity solution* of (34) if the following hold:

(a) $v_i^{n,\varepsilon}(x,\alpha)$ is uniformly continuous in x;
(b) for any $\alpha_0 \in \{1,2\}$ and x_0,

$$\begin{aligned}
&\min\Big\{\rho v_0^{n,\varepsilon}(x_0,\alpha_0) - \mathscr{A}^{\phi_0} v_0^{n,\varepsilon}(x_0,\alpha_0),\\
&\qquad\qquad v_0^{n,\varepsilon}(x_0,\alpha_0) - v_1^{n,\varepsilon}(x_0,\alpha_0) + h(x_0)(1+K)\Big\} \le 0,\\
&\min\Big\{\rho v_1^{n,\varepsilon}(x_0,\alpha_0) - \mathscr{A}^{\phi_1} v_1^{n,\varepsilon}(x_0,\alpha_0),\\
&\qquad\qquad v_1^{n,\varepsilon}(x_0,\alpha_0) - v_0^{n-1,\varepsilon}(x_0,\alpha_0) - h(x_0)(1-K)\Big\} \le 0,
\end{aligned} \tag{11.18}$$

for $n = 0, 1, \ldots, N$, whenever $\phi_i(x) \in C^2$ and $v_i^{n,\varepsilon}(x, \alpha_0) - \phi_i(x)$ has a local maximum at $x = x_0$; and

(c) for any $\alpha_0 \in \{1, 2\}$ and x_0,

$$\begin{aligned} \min\Big\{ & \rho v_0^{n,\varepsilon}(x_0, \alpha_0) - \mathscr{A}^{\psi_0} v_0^{n,\varepsilon}(x_0, \alpha_0), \\ & \qquad v_0^{n,\varepsilon}(x_0, \alpha_0) - v_1^{n,\varepsilon}(x_0, \alpha_0) + h(x_0)(1+K) \Big\} \geq 0, \\ \min\Big\{ & \rho v_1^{n,\varepsilon}(x_0, \alpha_0) - \mathscr{A}^{\psi_1} v_1^{n,\varepsilon}(x_0, \alpha_0), \\ & \qquad v_1^{n,\varepsilon}(x_0, \alpha_0) - v_0^{n-1,\varepsilon}(x_0, \alpha_0) - h(x_0)(1-K) \Big\} \geq 0, \end{aligned} \tag{11.19}$$

for $n = 0, 1, \ldots, N$, whenever $\psi_i(x) \in C^2$ and $v_i^{n,\varepsilon}(x, \alpha_0) - \psi_i(x)$ has a local minimum at $x = x_0$.

If (a) and (b) (resp. (a) and (c)) hold, we say that v is a *viscosity subsolution* (resp. *viscosity supersolution*).

Finally, we give the definition of viscosity solution of (38). Recall that

$$\overline{\mathscr{A}} f(x) = a(\overline{b} - x) \frac{df(x)}{dx} + \frac{\overline{\sigma}^2}{2} \frac{d^2 f(x)}{dx^2}.$$

Definition 11.2. $v_i^n(x)$ is a *viscosity solution* of (38) if the following hold:

(a) $v_i^n(x)$ is uniformly continuous in x;

(b) for any x_0,

$$\begin{aligned} &\min\left\{\rho v_0^n(x_0) - \overline{\mathscr{A}}\phi_0(x_0),\ v_0^n(x_0) - v_1^n(x_0) + h(x_0)(1+K)\right\} \leq 0, \\ &\min\left\{\rho v_1^n(x_0) - \overline{\mathscr{A}}\phi_1(x_0),\ v_1^n(x_0) - v_0^{n-1}(x_0) - h(x_0)(1-K)\right\} \leq 0, \end{aligned} \tag{11.20}$$

for $n = 0, 1, \ldots, N$, whenever $\phi_i(x) \in C^2$ and $v_i^n(x) - \phi_i(x)$ has a local maximum at $x = x_0$; and

(c) for any x_0,

$$\begin{aligned} &\min\left\{\rho v_0^n(x_0) - \overline{\mathscr{A}}\psi_0(x_0),\ v_0^n(x_0) - v_1^n(x_0) + h(x_0)(1+K)\right\} \geq 0, \\ &\min\left\{\rho v_1^n(x_0) - \overline{\mathscr{A}}\psi_1(x_0),\ v_1^n(x_0) - v_0^{n-1}(x_0) - h(x_0)(1-K)\right\} \geq 0, \end{aligned} \tag{11.21}$$

for $n = 0, 1, \ldots, N$, whenever $\psi_i(x) \in C^2$ and $v_i^n(x) - \psi_i(x)$ has a local minimum at $x = x_0$.

If (a) and (b) (resp. (a) and (c)) hold, we say that v is a *viscosity subsolution* (resp. *viscosity supersolution*).

Next, we give a sketch verifying the conditions in Example 11.1, i.e., we show that $h(x)$, $h'(x)$, $xh'(x)$, and $h''(x)$ are bounded and Lipschitz.

First note that $h_0(x)$ is bounded and Lipschitz. The boundedness and Lipschitz properties of h, h', and h'' follow from the equalities:

$$h(x) = \int_{-\infty}^{\infty} h_0(x-u)\Psi(u)du,$$

$$h'(x) = \int_{-\infty}^{\infty} h_0(x-u)\Psi'(u)du,$$

$$h''(x) = \int_{-\infty}^{\infty} h_0(x-u)\Psi''(u)du.$$

Next, we show that $xh'(x)$ is bounded. Note that

$$\begin{aligned} xh'(x) &= x\int_{-\infty}^{\infty} h_0'(u)\Psi(x-u)du, \\ &= x\int_{-\infty}^{M} e^u\Psi(x-u)du, \\ &= x\int_{x-M}^{\infty} e^{x-y}\Psi(y)dy, \text{ (with } y = x-u) \\ &= \frac{xe^x}{\sqrt{2\pi}}\int_{x-M}^{\infty} e^{-y}e^{-y^2/2}dy \\ &\le \frac{xe^x}{\sqrt{2\pi}}\int_{x-M}^{\infty} e^{-y^2/2}dy. \end{aligned} \tag{11.22}$$

Clearly, it is bounded on $(-\infty, M]$. To see it is also bounded on (M,∞), note also that

$$\frac{xe^x}{\sqrt{2\pi}}\int_{x-M}^{\infty} e^{-y^2/2}dy \le \frac{xe^x}{\sqrt{2\pi}}\left(\frac{\exp\left(-(x-M)^2/2\right)}{x-M}\right). \tag{11.23}$$

The boundedness follows.

Finally, to see the Lipschitz property of $xh'(x)$, in view of the Mean Value Theorem, it suffices to show that $xh''(x)$ is bounded. This can be done similarly as in (49) and (50) by noticing

$$xh''(x) = x\int_{-\infty}^{\infty} h_0'(u)\Psi'(x-u)du.$$

References

1. C. Blanco and D. Soronow, Mean reverting processes – Energy price processes used for derivatives pricing and risk management, *Commodities Now*, pp. 68-72, June 2001.
2. L.P. Bos, A.F. Ware and B.S. Pavlov, On a semi-spectral method for pricing an option on a mean-reverting asset, *Quantitative Finance*, Vol. 2, pp. 337–345, (2002).
3. A. Cowles and H. Jones, Some posteriori probabilities in stock market action, *Econometrica*, Vol. 5, pp. 280–294, (1937).
4. C. de Jong and R. Huisman, Option formulas for mean-reverting power prices with spikes, preprint.

5. M. Dai, Q. Zhang, and Q. Zhu, Trend following trading under a regime switching model, *SIAM Journal on Financial Mathematics*, Vol. 1, pp. 780–810, (2010).
6. E. Fama and K.R. French, Permanent and temporary components of stock prices, *Journal of Political Economy*, Vol. 96, pp. 246–273, (1988).
7. L.A. Gallagher and M.P. Taylor, Permanent and temporary components of stock prices: Evidence from assessing macroeconomic shocks, *Southern Economic Journal*, Vol 69, pp. 345–362, (2002).
8. C.M. Hafner and H. Herwartz, Option pricing under linear autoregressive dynamics, heteroskedasticity, and conditional leptokurtosis, *Journal of Empirical Finance*, Vol. 8, pp. 1–34, (2001).
9. T.C. Johnson and M. Zervos, The optimal timing of investment decisions, (Draft), (2006).
10. A. Løkka and M. Zervos, Long-term optimal real investment strategies in the presence of adjustment costs, preprint, (2007).
11. A. Merhi and M. Zervos, A model for reversible investment capacity expansion, *SIAM J. Contr. Optim.*, Vol. 46, pp. 839–876, (2007).
12. H. Pham, Optimal stopping of controlled jump diffusion processes: A viscosity solution approach, *J. Math. Syst. Estimation Control*, **8**, (1998), 1–27.
13. S.P. Sethi and Q. Zhang, *Hierarchical Decision Making in Stochastic Manufacturing Systems*, Birkhauser, Boston, 1994.
14. G. Yin and Q. Zhang, *Continuous-Time Markov Chains and Applications: A Singular Perturbation Approach*, Springer-Verlag, New York, 1998.
15. H. Zhang and Q. Zhang, Trading a mean-reverting asset: buy low and sell high, *Automatica*, Vol. 44, 1511-1518, (2008).

Chapter 12
CPPI in the Jump-Diffusion Model

Mingming Wang and Allanus Tsoi

12.1 Introduction

Constant Proportion Portfolio Insurance (CPPI) was introduced by [13] for equity instruments and has been further analyzed by many scholars (such as [1]). An investor invests in a portfolio and wants to protect the portfolio value from falling below a pre-assigned value. The investor shift his asset allocation over the investment period among a risk-free asset plus a collection of risky assets. The CPPI strategy is based on the dynamic portfolio allocation of two basic assets: a riskless asset (usually a treasury bill) and a risky asset (a stock index for example). This strategy relies crucially on the concept of a *cushion* C, which is defined as the difference between the *portfolio value* V and the *floor* F. This latter one corresponds to a guaranteed amount at any time t of the management period $[0,T]$. The key assumption is that the amount e invested on the risky asset, called the *exposure*, is equal to the cushion multiplied by a fixed coefficient m, called the *multiple*. The floor and the multiple can be chosen according to the investor's risk tolerance.

Consider the jump-diffusion process with $Y_n > -1$ representing the percentage of jump-size, i.e. $S_{T_n} = S_{T_n^-}(1+Y_n)$. Between two jumps, we assume that the risky asset model follows the Black–Scholes model. The number of jumps up to time t is a Poisson processes N_t with intensity λ_t. Our model becomes

$$S_t = S_0 \exp\left[\int_0^t (\mu_s - \frac{\sigma_s^2}{2})ds + \int_0^t \sigma_s dW_s + \sum_{n=1}^{N_t} \ln(1+Y_n)\right].$$

M. Wang
School of Insurance, The University of International Business & Economics,
Beijing 100029, China
e-mail: wiryming@gmail.com

A. Tsoi (✉)
Department of Mathematics, University of Missouri, Columbia, MO 65211, USA
e-mail: tsoia@missouri.edu

Y. Zeng and S. Wu (eds.), *State-Space Models: Applications in Economics and Finance*,
Statistics and Econometrics for Finance 1, DOI 10.1007/978-1-4614-7789-1_12,

We usually assume that the $\ln(1+Y_n)$ are i.i.d. with density function f_Q.

Our chapter is outlined as follows: in Sect. 12.2, we set up the jump-diffusion model, calculate the density function, and discuss the martingale measure. In Sect. 12.3, we describe the CPPI strategy and calculate the CPPI portfolio value, its expectation and variation. In Sect. 12.4, we consider the CPPI portfolio as a hedging tool (see [2]). In Sect. 12.5, we consider the mean-variance hedging for a given contingent claim H. In our jump-diffusion model, the market is not complete and so H is not attainable. Thus, we consider the mean-variance hedging as a kind of quadratic hedging [15]. We consider H as a function of the portfolio value V_T with risk measure $\mathbb{Q}$. We formulate our optimal problem (see (12.22)). We adopt the method used in Chap. 10 of [3] and give the explicit optimal solution of Z_0 and ϑ_t (see Proposition 5.2.). The explicit solution is applicable in real financial market. The main contribution of this chapter is contained in Sects. 12.4 and 12.5.

12.2 The Jump-Diffusion Model

To understand the background of our chapter we refer our readers to [3, 6, 10, 11, 14, 16], and [8].

Let $(\Omega, \mathfrak{F}, \mathfrak{F}_t, \mathbb{P})$ be a probability space satisfying the "usual assumption." Let the price S_t of a risky asset (usually stocks or their benchmark) be a right continuous with left limits stochastic process on this probability space which jumps at the random times $T_1, T_2, \ldots$ and suppose that the relative/proportional change in its value at a jump time is given by $Y_1, Y_2, \ldots$, respectively. We usually assume $\ln(1+Y_n)$s be i.i.d. and in our chapter, we usually assume the density function of $\ln(1+Y_n)$s be f_Q. We assume that, between any two near-by jump times, the price S_t follows the Black–Scholes model. Those T_ns are the jump times of a Poisson process N_t with intensity λ_t and those Y_ns are a sequence of random variables with values in $(-1, +\infty)$. We have

$$N_t = \sum_{n \geq 1} \chi_{t \geq T_n}$$

and

$$\mathbb{P}[N_t = n] = \frac{e^{-\int_0^t \lambda_s ds} (\int_0^t \lambda_s ds)^n}{n!}.$$

Then the description of the model can be formalized by letting, on the intervals $t \in [T_n, T_{n+1})$,

$$dS_t = S_t(\mu_t dt + \sigma_t dW_t),$$

and in exponential form:

$$S_t = S_{T_n} \exp\left[\int_{T_n}^t (\mu_s - \frac{\sigma_s^2}{2}) ds + \int_0^t \sigma_s dW_s\right].$$

While, at $t = T_n$, the jump size is given by $\Delta S_n = S_{T_n} - S_{T_n^-} = S_{T_n-} Y_n$, so that

$$S_{T_n} = S_{T_n^-}(1 + Y_n)$$

which, by the assumption that $Y_n > -1$, leads to positive values of the prices.

At the generic time t, S_t satisfies

$$S_t = S_0 \exp\left[\int_0^t (\mu_s - \frac{\sigma_s^2}{2})ds + \int_0^t \sigma_s dW_s\right]\left[\prod_{n=1}^{N_t}(1 + Y_n)\right] \tag{12.1}$$

$$= S_0 \exp\left[\int_0^t (\mu_s - \frac{\sigma_s^2}{2})ds + \int_0^t \sigma_s dW_s + \sum_{n=1}^{N_t} \ln(1 + Y_n)\right] \tag{12.2}$$

$$= S_0 \exp\left[\int_0^t (\mu_s - \frac{\sigma_s^2}{2})ds + \int_0^t \sigma_s dW_s + \int_0^t \ln(1 + Y_s)dN_s\right] \tag{12.3}$$

where Y_t is obtained from Y_n by a piecewise constant and left continuous time interpolation, i.e.

$$Y_t = Y_n \quad \text{if} \quad T_n < t \leq T_{n+1},$$

here we let $T_0 = 0$. The term $\sum_{n=1}^{N_t} \ln(1 + Y_n)]$ in (12.2) is a compound Poisson process. It has independent and stationary increments. Also because of (12.2), our jump-diffusion model is an exponential levy model. By the generalized Ito formula,

$$dS_t = S_{t-}[\mu_t dt + \sigma_t dW_t + Y_t dN_t], \tag{12.4}$$

with initial value S_0.

In general, if we assume $Q_n = \ln(1 + Y_n)$ are i.i.d with density function f_Q, then the density function of $\sum_{n=1}^{j} \ln(1 + Y_n)$ is $f_Q^{(j)}$. $f_Q^{(j)}(y)$ is the convolution of the density $f_Q(y)$ with itself j times. i.e.

$$f_Q^{(j)}(y) = \underbrace{f_Q(y) * f_Q(y) * \ldots * f_Q(y)}_{\text{Convoluted } j \text{ times}} \tag{12.5}$$

12.2.1 Density

We have the following proposition:

Proposition 12.1. *Let $Q_n = \ln(1 + Y_n)$ be i.i.d random variables with density function f_Q. The density function of*

$$\ln\left(\frac{S_t}{S_0}\right) = \int_0^t \left(\mu_s - \frac{\sigma_s^2}{2}\right) ds + \int_0^t \sigma_s dW_s + \Sigma_{n=1}^{N_t} \ln(1 + Y_n)$$

is:

$$p(x) = \sum_{j=0}^{\infty} \frac{e^{-\int_0^t \lambda_s ds} (\int_0^t \lambda_s ds)^j}{j!} \int_{-\infty}^{\infty} \phi\left(x-y; \int_0^t (\mu_s - \frac{\sigma_s^2}{2}) ds, \int_0^t \sigma_s^2 ds\right) f_Q^{(j)}(y) dy,$$

where $f_Q^{(j)}(y) = \underbrace{f_Q(y) * f_Q(y) * \ldots f_Q(y)}_{Convoluted\ j\ times}$ *and* $\phi(x,m,\upsilon^2) = \frac{1}{\sqrt{2\pi\upsilon^2}} e^{-\frac{(x-m)^2}{2\upsilon^2}}$.

Proof. Let $L = \int_0^t \left(\mu_s - \frac{\sigma_s^2}{2}\right) ds + \int_0^t \sigma_s dW_s$ and $M = \Sigma_{n=1}^{N_t} \ln(1+Y_n)$. Then,

$$L \sim N\left(\int_0^t (\mu_s - \frac{\sigma_s^2}{2}) ds, \int_0^t \sigma_s^2 ds\right).$$

When $N_t = j$, we have

$$\mathbb{P}(L+M \le x) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} \phi\left(y-y_2; \int_0^t \left(\mu_s - \frac{\sigma_s^2}{2}\right) ds, \int_0^t \sigma_s^2 ds\right) f_Q^{(j)}(y_2) dy_2 dy.$$

We calculate the distribution of $L+M$ in general. For all $x \in \mathbb{R}$, we have

$$\begin{aligned}
&\forall x \in \mathbb{R}, \\
&\mathbb{P}(L+M \le x) = \mathbb{P}\left[\bigcup_{j=0}^{\infty} (L+M \le x, N_t = j)\right] \\
&= \sum_{j=0}^{\infty} \mathbb{P}(L+M \le x, N_t = j) = \sum_{j=0}^{\infty} \mathbb{P}(L+M \le x | N_t = j) \mathbb{P}(N_t = j) \\
&= \sum_{j=0}^{\infty} \mathbb{P}(L+\Sigma_{n=1}^{j} \ln(1+Y_n) \le x | N_t = j) \mathbb{P}(N_t = j) \\
&= \sum_{j=0}^{\infty} \frac{\mathbb{P}(L+\Sigma_{n=1}^{j} \ln(1+Y_n) \le x, N_t = j)}{\mathbb{P}(N_t = j)} \mathbb{P}(N_t = j) \\
&= \sum_{j=0}^{\infty} \frac{\mathbb{P}(L+\Sigma_{n=1}^{j} \ln(1+Y_n) \le x) P(N_t = j)}{\mathbb{P}(N_t = j)} \mathbb{P}(N_t = j) \\
&= \sum_{j=0}^{\infty} \mathbb{P}(L+\Sigma_{n=1}^{j} \ln(1+Y_n) \le x) \mathbb{P}(N_t = j) \\
&= \sum_{j=0}^{\infty} \int_{-\infty}^{x} \phi\left(y; \int_0^t \left(\mu_s - \frac{\sigma_s^2}{2}\right) ds + j\alpha, \int_0^t \sigma_s^2 ds + j\delta^2\right) dy \frac{e^{-\int_0^t \lambda_s ds} (\int_0^t \lambda_s ds)^j}{j!}.
\end{aligned}$$

Each item in the above equations is positive, thus the series is absolute convergence. Hence, the density function is

$$
\begin{aligned}
& p(x) \\
&= \frac{d\left(\sum_{j=0}^{\infty}\int_{-\infty}^{x}\phi\left(y;\int_0^t\left(\mu_s-\frac{\sigma_s^2}{2}\right)ds+j\alpha,\int_0^t\sigma_s^2ds+j\delta^2\right)dy\frac{e^{-\int_0^t\lambda_s ds}(\int_0^t\lambda_s ds)^j}{j!}\right)}{dx} \\
&= \sum_{j=0}^{\infty}\frac{e^{-\int_0^t\lambda_s ds}(\int_0^t\lambda_s ds)^j}{j!}\phi\left(x;\int_0^t\left(\mu_s-\frac{\sigma_s^2}{2}\right)ds+j\alpha,\int_0^t\sigma_s^2ds+j\delta^2\right).
\end{aligned}
$$

□

The Merton's Model [12] and Kou's Model [9] are two common jump-diffusion models.

When we assume $\ln(1+Y_n)\sim N(\alpha,\delta^2)$, we have *Merton's* model [12].

When we assume $Q=\ln(1+Y_n)$ has an asymmetric double exponential distribution with the density function

$$f_Q(y)=p\cdot\eta_1 e^{-\eta_1 y}\chi_{y\geq 0}+q\cdot\eta_2 e^{-\eta_2 y}\chi_{y<0}$$

where $\eta_1>1$, $\eta_2>0$, $p,q\geq 0$ and $p+q=1$, then it is called *Kou's* model [9].

12.2.2 Martingale Measure

For our jump-diffusion model defined by (12.2), consider a predictable $\mathfrak{F}_t$-process ψ_t, such that $\int_0^t\psi_t\lambda_s ds<\infty$. Choose θ_t and ψ_t such that

$$\mu_t+\sigma_t\theta_t+Y_t\psi_t\lambda_t=r_t \tag{12.6}$$

and

$$\psi_t\geq 0.$$

From here we see that

$$\theta_t=\sigma_t^{-1}(r_t-\mu_t-Y_t\psi_t\lambda_t) \tag{12.7}$$

whose the choice of ψ_t is arbitrary. Define

$$L_t=\exp\left\{\int_0^t\left[(1-\psi_s)\lambda_s-\frac{1}{2}\theta_s^2\right]ds+\int_0^t\theta_s dW_s+\int_0^t\ln\psi_s dN_s\right\} \tag{12.8}$$

for $t\in[0,T]$ and the Radon–Nikodym derivative to be

$$\frac{d\mathbb{Q}}{d\mathbb{P}}=L_T. \tag{12.9}$$

Then $\mathbb{Q}$ is a risk neutral measure or martingale measure, i.e. a measure under which $\tilde{S}_t = \exp\{-\int_0^t r_s ds\} S_t$ is a martingale (see [14]).

Define

$$dW_t^{\mathbb{Q}} = dW_t - \theta_t dt; \tag{12.10}$$

$$dM_t^{\mathbb{Q}} = dN_t - \psi_t \lambda_t dt. \tag{12.11}$$

Then $W_t^{\mathbb{Q}}$ and $M_t^{\mathbb{Q}}$ are $\mathbb{Q}$-martingales. Also under the measure $\mathbb{Q}$, S_t satisfies

$$dS_t = S_{t-}[(\mu_t + \sigma_t \theta_t + Y_t \psi_t \lambda_t)dt + \sigma_t dW_t^{\mathbb{Q}} + Y_t dM_t^{\mathbb{Q}}]. \tag{12.12}$$

Under the measure $\mathbb{Q}$, N_t is a Poisson processes with intensity $\lambda_t \psi_t$.

There are many risk-neutral measures $\mathbb{Q} \sim \mathbb{P}$. A special case of a risk-neutral measure, reflecting the case of a risk-neutral world, it should satisfy

$$\mathbb{E}(S(t)) = S_0 e^{rt}.$$

(See page 312 on [6], page 248–250 on [7], page 19 on [11].)

In Merton's Model, $\ln(1+Y_n) \sim N(\alpha, \delta^2)$, by Proposition 12.1, it follows that the density satisfies

$$p(x) = \sum_{j=0}^{\infty} \frac{e^{-\int_0^t \lambda_s ds} (\int_0^t \lambda_s ds)^j}{j!} \phi\left(x; \int_0^t \left(\mu_s - \frac{\sigma_s^2}{2}\right) ds + j\alpha, \int_0^t \sigma_s^2 ds + j\delta^2\right).$$

Then

$$\begin{aligned}
&\mathbb{E}(S(t)) = S_0 \mathbb{E}\left(e^{\ln S_t / S_0}\right) = S_0 \int_{\mathbb{R}} e^x p(x) dx \\
&= S_0 \int_{\mathbb{R}} e^x \sum_{j=0}^{\infty} \frac{e^{-\int_0^t \lambda_s ds} (\int_0^t \lambda_s ds)^j}{j!} \phi\left(x; \int_0^t \left(\mu_s - \frac{\sigma_s^2}{2}\right) ds + j\alpha, \int_0^t \sigma_s^2 ds + j\delta^2\right) dx \\
&= S_0 \sum_{j=0}^{\infty} \frac{e^{-\int_0^t \lambda_s ds} (\int_0^t \lambda_s ds)^j}{j!} \int_{\mathbb{R}} e^x \phi\left(x; \int_0^t \left(\mu_s - \frac{\sigma_s^2}{2}\right) ds + j\alpha, \int_0^t \sigma_s^2 ds + j\delta^2\right) dx \\
&= S_0 \sum_{j=0}^{\infty} \frac{e^{-\int_0^t \lambda_s ds} (\int_0^t \lambda_s ds)^j}{j!} \exp\left\{\int_0^t \mu_s ds + j\alpha + j\frac{\delta^2}{2}\right\} \\
&= S_0 \exp \int_0^t \left(\mu_s - \lambda_s + e^{\alpha + \frac{\delta^2}{2}} \lambda_s\right) ds.
\end{aligned}$$

When

$$\mathbb{E}(S(t)) = S_0 e^{rt},$$

we have

$$\mu_s - \lambda_s + e^{\alpha + \frac{\delta^2}{2}} \lambda_s = r.$$

Thus under our new risk-neutral measure $\mathbb{P}^{rn}$, we can use $r+\lambda_s - e^{\alpha+\frac{\delta^2}{2}}\lambda_s$ to substitute μ_s. The model then becomes

$$S_t = S_0 \exp\left[\int_0^t \left(r+\lambda_s - e^{\alpha+\frac{\delta^2}{2}}\lambda_s - \frac{\sigma_s^2}{2}\right)ds + \int_0^t \sigma_s dW_s^{rn} + \sum_{n=1}^{N_t^{(rn)}} \ln(1+Y_n)\right].$$

$W_s^{(rn)}$ is a Brownian motion and $N_t^{(rn)}$ is Poisson process which intensity is λ_s under the probability measure $\mathbb{P}^{rn}$. For convenience, we still denote them as W_s and N_t. Then, under the probability measure $\mathbb{P}^{rn}$, the model is

$$S_t = S_0 \exp\left[\int_0^t \left(r+\lambda_s - e^{\alpha+\frac{\delta^2}{2}}\lambda_s - \frac{\sigma_s^2}{2}\right)ds + \int_0^t \sigma_s dW_s + \sum_{n=1}^{N_t} \ln(1+Y_n)\right].$$

12.3 The CPPI Strategies

12.3.1 The constant multiple case

The *CPPI* strategy is based on a dynamic portfolio allocation on two basic assets: a riskless asset (usually a treasury bill) and a risky asset (a stock index for example).

At time t, the *exposure* e_t is equal to the *cushion* C_t multiplied by the *multiple* m. The *cushion* C_t is defined as the difference between the *portfolio* value V_t and the *floor* F_t. $F_t = G\exp\{-r(T-t)\}$, G is the floor at time T. Because of the existence of jumps, it is possible to have the case that the portfolio value is less than the floor. Then, the cushion will be negative and so will be the exposure. That means short-selling should be allowed. The following proposition describes the portfolio value under this strategy.

Denote portfolio value as V_t. It consists with riskless asset $V_t - mC_t$ and risky asset mC_t. i.e. $V_t = mC_t + (V_t - mC_t)$. Let the interest rate be r and floor at time t be $F_t = F_0 e^{rt} = F_T e^{-r(T-t)}$. For convenience we summarize our notations:

Name	Notation
Interest rate	r
Time	t
Time period	$[0,T]$
Floor at time t	F_t
Portfolio value at time t	V_t
Cushion at time t	C_t
Multiple	m
Exposure at time t	e_t
Riskless asset at time t	B_t

Their relations are as follows

$$\begin{aligned} C_t &= V_t - F_t; \\ e_t &= mC_t; \\ B_t &= V_t - e_t. \end{aligned}$$

Proposition 12.2. *The CPPI portfolio value under the jump-diffusion model defined by* (12.2) *is*

$$\begin{aligned} V_t = C_0 \exp\Bigg\{ \int_0^t \left(r + m(\mu_s - r) - \frac{m\sigma_s^2}{2} \right) ds \\ + \int_0^t m\sigma_s dW_s \Bigg\} \left[\prod_{n=1}^{N_t} (1 + mY_n) \right] + F_t, \end{aligned}$$

where

$$\begin{aligned} C_0 &= \left(V_0 - Ge^{-rT} \right), \\ F_t &= G \times \exp\{-r(T-t)\}. \end{aligned}$$

Proof. We have

$$\begin{aligned} V_t &= mC_t + (V_t - mC_t) \\ &= V_t \left[\frac{mC_t}{V_t} + \left(1 - \frac{mC_t}{V_t} \right) \right] \end{aligned}$$

and

$$dV_t = V_t \left[\frac{mC_t}{V_{t-}} \frac{dS_t}{S_{t-}} + \left(1 - \frac{mC_t}{V_{t-}} \right) \frac{dB_t}{B_t} \right].$$

Since B_s is continuous, then $B_{s-} = B_s$, we have

$$\begin{aligned} dC_t &= d(V_t - F_t) \\ &= V_t \left[\frac{mC_{t-}}{V_t} \frac{dS_t}{S_{t-}} + \left(1 - \frac{mC_{t-}}{V_t} \right) \frac{dB_t}{B_t} \right] - F_t \frac{dB_t}{B_t} \\ &= C_{t-} \left(\frac{m dS_t}{S_{t-}} - (m-1) r dt \right) \\ &= C_{t-} [m(\mu_t dt + \sigma_t dW_t + Y_t dN_t) - (m-1) r dt] \\ &= C_{t-} [(r + m(\mu_t - r)) dt + m\sigma_t dW_t + mY_t dN_t]. \end{aligned} \tag{12.13}$$

Then

$$C_t = C_0 \exp\left\{ \int_0^t \left(r + m(\mu_s - r) - \frac{m^2\sigma_s^2}{2} \right) ds + \int_0^t m\sigma_s dW_s \right\} \left[\prod_{n=1}^{N_t} (1 + mY_n) \right].$$

Hence

$$\begin{aligned} V_t &= C_t + F_t \\ &= C_0 \exp\left\{ \int_0^t \left(r + m(\mu_s - r) - \frac{m\sigma_s^2}{2} \right) ds \right. \\ &\quad \left. + \int_0^t m\sigma_s dW_s \right\} \left[\prod_{n=1}^{N_t} (1 + mY_n) \right] + F_t. \end{aligned}$$

□

If we substitute μ_s by $r + \lambda_s - e^{\alpha + \frac{\delta^2}{2}} \lambda_s$, under the probability measure $\mathbb{P}^{rn}$, we get the following corollary.

Corollary 12.1. *In Merton's model, under the probability measure $\mathbb{P}^{rn}$, the CPPI portfolio value under jump-diffusion model is*

$$\begin{aligned} V_t = C_0 \exp\left\{ \int_0^t \left[r + m\left(\lambda_s - e^{\alpha + \frac{\delta^2}{2}} \lambda_s \right) - \frac{m\sigma_s^2}{2} \right] ds \right. \\ \left. + \int_0^t m\sigma_s dW_s \right\} \left[\prod_{n=1}^{N_t} (1 + mY_n) \right] + F_t, \end{aligned}$$

where

$$\begin{aligned} C_0 &= (V_0 - Ge^{-rT}) \\ F_t &= G \times \exp\{-r(T-t)\}. \end{aligned}$$

The expectation and variance of the CPPI portfolio value are deduced in the following two propositions. They are obviously two important values to describe the CPPI strategy in our jump-diffusion model.

Proposition 12.3. *The expected CPPI portfolio value at time t under the jump-diffusion model is*

$$\mathbb{E}[V_t] = C_0 \exp\left\{ \int_0^t (r + m(\mu_s - r)) ds \right\} \sum_{k=1}^{\infty} \frac{e^{-\int_0^t \lambda_s ds} (\int_0^t \lambda_s ds)^k}{k!} \mathbb{E}\left[\prod_{n=1}^{k} (1 + mY_n) \right] + F_t.$$

Proof. Because

$$\begin{aligned}
&\mathbb{P}\left[\prod_{n=1}^{N_t}(1+mY_n)\le x\right]=\mathbb{P}\left[\bigcup_{k=1}^{\infty}\left(\prod_{n=1}^{N_t}(1+mY_n)\le x,N_t=k\right)\right]\\
&=\sum_{k=1}^{\infty}\mathbb{P}\left[\prod_{n=1}^{N_t}(1+mY_n)\le x|N_t=k\right]\mathbb{P}[N_t=k]\\
&=\sum_{k=1}^{\infty}\mathbb{P}\left[\prod_{n=1}^{k}(1+mY_n)\le x|N_t=k\right]\mathbb{P}[N_t=k]\\
&=\sum_{k=1}^{\infty}\mathbb{P}\frac{\left[\prod_{n=1}^{k}(1+mY_n)\le x,N_t=k\right]}{\mathbb{P}[N_t=k]}\mathbb{P}[N_t=k]\\
&=\sum_{k=1}^{\infty}\mathbb{P}\frac{\left[\prod_{n=1}^{k}(1+mY_n)\le x\right]\mathbb{P}[N_t=k]}{\mathbb{P}[N_t=k]}P[N_t=k]\\
&=\sum_{k=1}^{\infty}\mathbb{P}\left[\prod_{n=1}^{k}(1+mY_n)\le x\right]\mathbb{P}[N_t=k]\\
&=\sum_{k=1}^{\infty}\frac{e^{-\int_0^t\lambda_s ds}(\int_0^t\lambda_s ds)^k}{k!}\mathbb{P}\left[\prod_{n=1}^{k}(1+mY_n)\le x\right],
\end{aligned}$$

we get

$$\mathbb{E}\left[\prod_{n=1}^{N_t}(1+mY_n)\right]=\sum_{k=1}^{\infty}\frac{e^{-\int_0^t\lambda_s ds}(\int_0^t\lambda_s ds)^k}{k!}\mathbb{E}\left[\prod_{n=1}^{k}(1+mY_n)\right]$$

and then

$$\begin{aligned}
&\mathbb{E}[V_t]\\
&=C_0\mathbb{E}\left[\exp\left\{\int_0^t\left(r+m(\mu_s-r)-\frac{m^2\sigma_s^2}{2}\right)ds+\int_0^t m\sigma_s dW_s\right\}\right]\mathbb{E}\left[\prod_{n=1}^{N_t}(1+mY_n)\right]+F_t\\
&=C_0\exp\left\{\int_0^t[r+m(\mu_s-r)]ds\right\}\mathbb{E}\left[\prod_{n=1}^{N_t}(1+mY_n)\right]+F_t\\
&=C_0\exp\left\{\int_0^t[r+m(\mu_s-r)]ds\right\}\sum_{k=1}^{\infty}\frac{e^{-\int_0^t\lambda_s ds}(\int_0^t\lambda_s ds)^k}{k!}\mathbb{E}\left[\prod_{n=1}^{k}(1+mY_n)\right]+F_t.
\end{aligned}$$

□

Proposition 12.4. *The variance of the CPPI portfolio value at time t under jump-diffusion model is*

$$C_0^2 \exp\left\{\int_0^t 2\left[r+m(\mu_s-r)+m^2\sigma_s^2\right]ds\right\}\sum_{k=1}^{\infty}\mathbb{E}\left[\prod_{n=1}^{k}(1+mY_n)\right]^2\frac{e^{-\int_0^t\lambda_s ds}(\int_0^t\lambda_s ds)^k}{k!}$$
$$-\left(\exp\left\{\int_0^t[r+m(\mu_s-r)]\right\}ds\sum_{k=1}^{\infty}\frac{e^{-\int_0^t\lambda_s ds}(\int_0^t\lambda_s ds)^k}{k!}\mathbb{E}\left[\prod_{n=1}^{k}(1+mY_n)\right]\right)^2.$$

Proof. Similar to the proof of Prop. 12.3, we have

$$\mathbb{E}\left(\left[\prod_{n=1}^{N_t}(1+mY_n)\right]^2\right)=\frac{e^{-\int_0^t\lambda_s ds}(\int_0^t\lambda_s ds)^k}{k!}\mathbb{E}\left(\left[\prod_{n=1}^{k}(1+mY_n)\right]^2\right).$$

Thus,

$$\begin{aligned}
&\mathrm{Var}[V_t]=\mathrm{Var}[C_t]\\
&=C_0^2\mathrm{Var}\left(\exp\left\{\int_0^t\left(r+m(\mu_s-r)-\frac{m^2\sigma_s^2}{2}\right)ds+\int_0^t m\sigma_s dW_s\right\}\left[\prod_{n=1}^{N_t}(1+mY_n)\right]\right)\\
&=C_0^2\mathbb{E}\left(\exp\left\{\int_0^t\left(r+m(\mu_s-r)-\frac{m^2\sigma_s^2}{2}\right)ds+\int_0^t m\sigma_s dW_s\right\}\left[\prod_{n=1}^{N_t}(1+mY_n)\right]\right)^2\\
&\quad-C_0^2\left(\mathbb{E}\left[\exp\left\{\int_0^t\left(r+m(\mu_s-r)-\frac{m^2\sigma_s^2}{2}\right)ds+\int_0^t m\sigma_s dW_s\right\}\left[\prod_{n=1}^{N_t}(1+mY_n)\right]\right]\right)^2\\
&=C_0^2\mathbb{E}\left[\exp\left\{\int_0^t\left(r+m(\mu_s-r)-\frac{m^2\sigma_s^2}{2}\right)ds+\int_0^t m\sigma_s dW_s\right\}\left[\prod_{n=1}^{N_t}(1+mY_n)\right]\right]^2\\
&\quad-C_0^2\left[\exp\left\{\int_0^t[r+m(\mu_s-r)]ds\right\}\sum_{k=1}^{\infty}\frac{e^{-\int_0^t\lambda_s ds}(\int_0^t\lambda_s ds)^k}{k!}\mathbb{E}\left[\prod_{n=1}^{k}(1+mY_n)\right]\right]^2\\
&=C_0^2\mathbb{E}\left[\exp\left\{\int_0^t 2[r+m(\mu_s-r)-m^2\sigma_s^2]ds+2\int_0^t m\sigma_s dW_s\right\}\right]\mathbb{E}\left[\prod_{n=1}^{N_t}(1+mY_n)\right]^2\\
&\quad-C_0^2\left[\exp\left\{\int_0^t(r+m(\mu_s-r))ds\right\}\sum_{k=1}^{\infty}\frac{e^{-\int_0^t\lambda_s ds}(\int_0^t\lambda_s ds)^k}{k!}\mathbb{E}\left[\prod_{n=1}^{k}(1+mY_n)\right]\right]^2\\
&=C_0^2\exp\left\{\int_0^t 2[r+m(\mu_s-r)+m^2\sigma_s^2]ds\right\}\sum_{k=1}^{\infty}\mathbb{E}\left[\prod_{n=1}^{k}(1+mY_n)\right]^2\frac{e^{-\int_0^t\lambda_s ds}(\int_0^t\lambda_s ds)^k}{k!}\\
&\quad-\left[\exp\left\{\int_0^t[r+m(\mu_s-r)]\right\}ds\sum_{k=1}^{\infty}\frac{e^{-\int_0^t\lambda_s ds}(\int_0^t\lambda_s ds)^k}{k!}\mathbb{E}\left[\prod_{n=1}^{k}(1+mY_n)\right]\right]^2.
\end{aligned}$$

□

Remark 12.1. Another method to calculate the expectation of the portfolio value is calculating the characteristic function of

$$\int_0^t \left([r + m(\mu_s - r)] - \frac{m\sigma_s^2}{2} \right) ds + \int_0^t m\sigma_s dW_s + \prod_{n=1}^{N_t} (1 + mY_n)$$

Remark 12.2. For the Merton's and Kou's model, $\mathbb{E}[\prod_{n=1}^{k}(1+mY_n)]$ and $\mathbb{E}[\prod_{n=1}^{k}(1+mY_n)]^2$ can be calculated and thus the expected portfolio can be calculated explicitly. In general, if we assume $Q_n = \ln(1+Y_n)$ are i.i.d with density f_Q, $\mathbb{E}[\prod_{n=1}^{k}(1+mY_n)]$ and $\mathbb{E}[\prod_{n=1}^{k}(1+mY_n)]^2$ still can be calculated in terms of the function of f_Q.

The following lemma gives the density function of $1+mY_i$.

Lemma 12.1. *Let the density function of* $\ln(1+Y_n)$ *be* $f_Q(y)$. *Then the density function* f'_Q *of the random variable* $1+mY_i$ *is*

$$f'_Q(z) = f_Q\left[\ln\left(1 + \frac{z-1}{m}\right)\right] \frac{1}{m+z-1}.$$

Proof. Since

$$\begin{aligned} \mathbb{P}(1+mY_i \le z) &= \mathbb{P}\left[\ln(1+Y_i) \le \ln\left(1+\frac{z-1}{m}\right)\right] \\ &= \int_{-\infty}^{\ln(1+\frac{z-1}{m})} f_Q(y)dy, \end{aligned}$$

the density f'_Q of the random variable $1+mY_i$ is

$$f'_Q(z) = \frac{d\left(\mathbb{P}(1+mY_i \le z)\right)}{dz} = f_Q\left[\ln\left(1+\frac{z-1}{m}\right)\right]\frac{1}{m+z-1}.$$

□

Now we can calculate

$$\begin{aligned} \mathbb{E}\left[\prod_{n=1}^{k}(1+mY_n)\right] &= \mathbb{E}\left[\exp\left\{\sum_{n=1}^{k}\ln(1+mY_n)\right\}\right] \\ &= \int_{\mathbb{R}} \exp\left\{\underbrace{f'_Q * f'_Q * \ldots * f'_Q(x)}_{k \text{ items}}\right\} dx \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}\left[\prod_{n=1}^{k}(1+mY_n)^2\right] &= \mathbb{E}\left[\exp\left\{\sum_{n=1}^{k}2\ln(1+mY_n)\right\}\right] \\ &= \int_{\mathbb{R}} \exp\left\{2\underbrace{f'_Q * f'_Q * \ldots * f'_Q(x)}_{k \text{ items}}\right\} dx. \end{aligned}$$

12.3.2 The Time-Varying Multiple Case

We study the case when the multiple is a function of time. Let m_t be the multiple at time t. The conclusion does not change much in comparison with the constant case. We still have similar propositions:

Proposition 12.5. *When the multiple is a function of time at time t, the CPPI portfolio value under the jump-diffusion model is*

$$V_t = C_0 \exp\left\{\int_0^t \left(r + m_s[\mu_s - r] - \frac{m_s^2\sigma_s^2}{2}\right) ds + \int_0^t m_s\sigma_s dW_s\right\}\left[\prod_{n=1}^{N_t}(1+m_nY_n)\right] + F_t,$$

where m_n is obtained from m_t by the formula

$$m_n = m_{T_n},$$

where $T_0 = 0$.

Proposition 12.6. *When the multiple is a function of time at time t, the expected CPPI portfolio value under jump-diffusion model is*

$$C_0 \exp\left\{\int_0^t [r + m_s(\mu_s - r)]ds \sum_{k=1}^{\infty} \frac{e^{-\int_0^t \lambda_s ds}(\int_0^t \lambda_s ds)^k}{k!}\mathbb{E}\left[\prod_{n=1}^{k}(1+m_nY_n)\right]\right\} + F_t.$$

Proposition 12.7. *When the multiple is a function of time at time t, the variance of the CPPI portfolio value under jump-diffusion model is*

$$C_0^2 \exp\left\{\int_0^t 2\left(r + m_s(\mu_s - r) + m_s^2\sigma_s^2\right) ds\right\} \sum_{k=1}^{\infty} \mathbb{E}\left[\prod_{n=1}^{k}(1+m_nY_n)\right]^2 \frac{e^{-\int_0^t \lambda_s ds}(\int_0^t \lambda_s ds)^k}{k!}$$
$$- \left\{\exp\left\{\int_0^t [r + m_s(\mu_s - r)]ds\right\} \sum_{k=1}^{\infty} \frac{e^{-\int_0^t \lambda_s ds}(\int_0^t \lambda_s ds)^k}{k!}\mathbb{E}\left[\prod_{n=1}^{k}(1+m_nY_n)\right]\right\}^2.$$

12.4 The CPPI Portfolio as a Hedging Tool

We have proved that the portfolio value is

$$V_t = C_0 \exp\left\{\int_0^t \left(r + m_s(\mu_s - r) - \frac{m_s^2\sigma_s^2}{2}\right) ds + \int_0^t m_s\sigma_s dW_s\right\}\left[\prod_{n=1}^{N_t}(1+mY_n)\right] + F_t.$$

In Sect. 12.4 of [2], the CPPI portfolio as an hedging tool under the Black–Scholes model is discussed. [5] also discusses the option on CPPI under the Black–Scholes model. In this section, we generalize the result to our jump-diffusion model.

12.4.1 PIDE Approach

Suppose that $\eta = g(S_T)$ is a contingent claim that the portfolio's manager is aiming to have at maturity. Can the CPPI portfolio be converted into a synthetic derivative with pay-off specified by $\eta = g(S_T)$?

Theorem 12.1. *If $g : \mathbb{R} \to \mathbb{R}$ is sufficiently smooth, there exists a unique self-financed $g(S_T)$ hedging CPPI portfolio V; defined by*

$$V_t = v(t, S_t) \quad t \in [0,T] \tag{12.14}$$

where $v \in C^{1,2}([0,T] \times \mathbb{R})$ is the unique solution of the following partial integro-differential equations (PIDE).

$$\frac{\partial u}{\partial t}(t,s) + (\mu_t s)\frac{\partial u}{\partial x}(t,s) + \frac{1}{2}(s\sigma_t)^2 \frac{\partial^2 u}{\partial x^2}(t,s) - ru(t,s) = 0, \tag{12.15}$$

$$sz\frac{\partial u}{\partial x}(t,s) = u(t,s+sz) - u(t,s), \tag{12.16}$$

$$u(T,s) = g(s), \quad (t,s) \in [0,T] \times \mathbb{R}), \quad u \in C^{1,2}([0,T] \times \mathbb{R}). \tag{12.17}$$

Here, $\frac{\partial u}{\partial x}$ is the partial derivative to the second variable. In particular the CPPI portfolio's gearing factor is given by:

$$m_t = \frac{\frac{\partial u}{\partial x}(t, S_t)S_{t-}}{V_{t-} - F_t}, \quad t \in [0,T]. \tag{12.18}$$

Proof. For V to be a self-financed $g(S_T)$-hedging portfolio, it is enough to ensure that at maturity time we have

$$V_T = g(S_T), \quad a.s..$$

Choose a map $v \in C^{1,2}([0,T] \times \mathbb{R})$ and set $V_t = v(t, S_t)(t \in [0,T])$. Then $v(T, S_T) = g(S_T)$ $\mathbb{P}$-a.s., therefore

$$v(T,s) = g(s), \qquad \forall s \in \mathbb{R}.$$

Second by Ito's chain rule,

$$\begin{aligned} dv(t,S_t) = &\left(\frac{\partial v}{\partial t} + \mu_t S_{t-}\frac{\partial v}{\partial x} + \frac{1}{2}(\sigma_t S_{t-})^2 \frac{\partial^2 v}{\partial x^2}\right)(t,S_t)dt \\ &+ S_{t-}\sigma_t \frac{\partial v}{\partial x}(t,S_t)dW_t + [v(t, S_{t-} + S_{t-}Y_t) - v(t,S_{t-})]dN_t. \end{aligned}$$

Now V_t satisfies

$$\begin{aligned} dV_t &= dC_t + dF_t \\ &= (V_{t-} - F_t)[r + m_t(\mu_t - r)]dt + rF_t dt + (V_{t-} - F_t)m_t\sigma_t dW_t \\ &\quad + (V_{t-} - F_t)m_t Y_t dN_t \\ &= [rV_{t-} + (V_{t-} - F_t)m_t(\mu_t - r)]dt + (V_{t-} - F_t)m_t\sigma_t dW_t \\ &\quad + (V_{t-} - F_t)m_t Y_t dN_t. \end{aligned}$$

A comparison of the above two equations implies that

$$m_t = \frac{\frac{\partial u}{\partial x}(t, S_t) S_{t-}}{V_{t-} - F_t}, \quad t \in [0, T]$$

and

$$\frac{\partial u}{\partial t}(t,s) + (\mu_t s)\frac{\partial u}{\partial x}(t,s) + \frac{1}{2}(s\sigma_t)^2 \frac{\partial^2 u}{\partial x^2}(t,s) - ru(t,s) = 0$$
$$sz\frac{\partial u}{\partial x}(t,s) = u(t, s+sz) - u(t,s).$$

□

In a financial turmoil, the portfolio's manager acting on the leverage regime may convert the CPPI portfolio in a suitable synthetic derivative whose price is specified by (12.14)–(12.17). Moreover the required dynamic gearing factor (multiple) can be easily determined, using (12.18). This is the PIDE/PDE approach hedging.

Another observation that reveals to be central in the analysis of possible portfolio's hedges is that at any time of the financial horizon the CPPI portfolio value may be regarded as a standard risky asset and therefore as an underlying for any convenient contingent claim:

Theorem 12.2. *Under the risk neutral measure $\mathbb{Q}$, the discounted CPPI portfolio's value $\{V_t\}_{t\in[0,T]}$*

$$\tilde{V}_t = e^{-rt} V_t, \quad t \in [0, T] \tag{12.19}$$

is a Martingale.

Proof. In the proof of last theorem, we have deduced

$$dV_t = [rV_{t-} + (V_{t-} - F_t)m_t(\mu_t - r)]dt + (V_{t-} - F_t)m_t\sigma_t dW_t + (V_{t-} - F_t)m_t Y_t dN_t.$$

Thus we have

$$\begin{aligned}
dV_t &= (rV_{t-} + (V_{t-} - F_t)m_t(\mu_t - r))dt + (V_{t-} - F_t)m_t\sigma_t(dW_t^{\mathbb{Q}} + \theta_t dt) \\
&\quad + (V_{t-} - F_t)m_t Y_t dN_t \\
&= (rV_{t-} + (V_{t-} - F_t)m_t(\mu_t - r + \theta_t))dt + (V_{t-} - F_t)m_t\sigma_t dW_t^{\mathbb{Q}} \\
&\quad + (V_{t-} - F_t)m_t Y_t dN_t \\
&= [rV_{t-} + (V_{t-} - F_t)m_t\sigma_t]dW_t^{\mathbb{Q}} + [(V_{t-} - F_t)m_t(-Y_t\psi_t\lambda_t)]dt \\
&\quad + (V_{t-} - F_t)m_t Y_t dN_t \\
&= [rV_{t-} + (V_{t-} - F_t)m_t\sigma_t]dW_t^{\mathbb{Q}} + (V_{t-} - F_t)m_t Y_t dM_t^{\mathbb{Q}}.
\end{aligned}$$

Integration by parts implies that

$$\begin{aligned} d\tilde{V}_t &= de^{-rt}V_t = -re^{-rt}V_t dt + e^{-rt}dV_t \\ &= e^{-rt}[(V_{t-} - F_t)m_t\sigma_t dW_t^{\mathbb{Q}} + (V_{t-} - F_t)m_t Y_t dM_t^{\mathbb{Q}}]. \end{aligned}$$

Thus, $\tilde{V}_t$ is a $\mathbb{Q}$-Martingale. □

If we substitute μ_s by $r + \lambda_s - e^{\alpha + \frac{\delta^2}{2}}\lambda_s$, under the probability measure $\mathbb{P}^{rn}$, we get the following corollary.

Corollary 12.2. *In Merton's model, under the probability measure $\mathbb{P}^{rn}$, the discounted CPPI portfolio's value $\{V_t\}_{t\in[0,T]}$*

$$\tilde{V}_t = e^{-rt}V_t, \quad t \in [0,T]$$

is a Martingale.

Proof. We have

$$\begin{aligned} dV_t &= [rV_{t-} + (V_{t-} - F_t)m_t(\mu_t - r)]dt + (V_{t-} - F_t)m_t\sigma_t dW_t \\ &\quad + (V_{t-} - F_t)m_t Y_t dN_t \\ &= \left[rV_{t-} + (V_{t-} - F_t)m_t\left(\lambda_t - e^{\alpha + \frac{\delta^2}{2}}\lambda_t\right)\right]dt \\ &\quad + (V_{t-} - F_t)m_t\sigma_t dW_t + (V_{t-} - F_t)m_t Y_t dN_t. \end{aligned}$$

Thus

$$\begin{aligned} d\tilde{V}_t &= de^{-rt}V_t = -re^{-rt}V_t dt + e^{-rt}dV_t \\ &= -re^{-rt}V_t dt + e^{-rt}(rV_{t-} + (V_{t-} - F_t)m_t\left(\lambda_t - e^{\alpha + \frac{\delta^2}{2}}\lambda_t\right)dt \\ &\quad + (V_{t-} - F_t)m_t\sigma_t dW_t + (V_{t-} - F_t)m_t Y_t dN_t) \\ &= e^{-rt}\left[(V_{t-} - F_t)m_t\left(\lambda_t - e^{\alpha + \frac{\delta^2}{2}}\lambda_t\right)dt\right. \\ &\quad \left. + (V_{t-} - F_t)m_t\sigma_t dW_t + (V_{t-} - F_t)m_t Y_t dN_t\right] \\ &= e^{-rt}\left[(V_{t-} - F_t)m_t\left(\lambda_t - e^{\alpha + \frac{\delta^2}{2}}\lambda_t + \lambda_t Y_t\right)dt\right. \\ &\quad \left. + (V_{t-} - F_t)m_t\sigma_t dW_t + (V_{t-} - F_t)m_t Y_t(dN_t - \lambda_t dt)\right]. \end{aligned}$$

Since $dN_t - \lambda_t dt$ is a martingale and

$$\mathbb{E}[Y_t] = \mathbb{E}\left[e^{\ln(1+Y_n)} - 1\right] = e^{\alpha + \frac{\delta^2}{2}} - 1,$$

we get $\mathbb{E}\left[\lambda_t - e^{\alpha + \frac{\delta^2}{2}}\lambda_t\right] = 0$, so we prove $\tilde{V}_t$ is a $\mathbb{P}^{nr}$-Martingale. □

Given any claim $\eta = g(V_T)$ which is a function of the terminal portfolio's price, there exists a unique self-financed $\eta = g(V_T)$-hedging strategy:

Theorem 12.3. *Let $g : \mathbb{R} \to \mathbb{R}$ sufficiently smooth. Then there exists a unique $\eta = g(V_T)$-hedging self-financed trading strategy (U, η) defined as*

$$U_t = u(t, V_t), \quad \eta_{t-} = \frac{\partial u}{\partial x}(t, V_t), \quad t \in [0, T],$$

where $u \in C^{1,2}([0,T] \times \mathbb{R})$ is the unique solution of the PIDE.

$$\frac{\partial u}{\partial t}(t,v) + rv\frac{\partial u}{\partial x}(t,v) + \frac{1}{2}m^2\sigma_t^2(v-f)^2\frac{\partial^2 u}{\partial x^2}(t,v) - ru(t,v) = 0 \tag{12.20}$$

$$mz(v-f)\frac{\partial u}{\partial x}(t,v) = u(t, v + m[v-f]z) - u(t,v) \tag{12.21}$$

with the final condition $u(T,v) = g(v)$.

Proof. Consider an asset $\{V_t\}_{t\in[0,T]}$, and pick a self-financed $g(V_T)$ hedging strategy space $(U_t, \eta_t)_{t\in[0,T]}$ by setting:

$$dU_t = \eta_{t-}dV_t + (U_{t-} - \eta_{t-}V_{t-})rdt$$

and

$$U_T = g(V_T) \quad a.s.$$

Since

$$\begin{aligned} dV_t = {} & [rV_{t-} + (V_{t-} - F_t)m_t(\mu_t - r)]dt + (V_{t-} - F_t)m_t\sigma_t dW_t \\ & + (V_{t-} - F_t)m_t Y_t dN_t, \end{aligned}$$

the hedging portfolio's equation may be rewritten as:

$$\begin{aligned} dU_t &= \eta_{t-}[rV_{t-} + (V_{t-} - F_t)m_t(\mu_t - r)dt + (V_{t-} - F_t)m_t\sigma_t dW_t \\ &\quad + (V_{t-} - F_t)m_t Y_t dN_t] + (U_{t-} - \eta_{t-}V_{t-})rdt \\ &= [rU_{t-} + \eta_{t-}(V_{t-} - F_t)m(\mu_t - r)]dt + \eta_{t-}(V_{t-} - F_t)m_t\sigma_t dW_t \\ &\quad + [\eta_{t-}(V_{t-} - F_t)m_t Y_t dN_t]. \end{aligned}$$

Pick $u \in C^{1,2}([0,T] \times \mathbb{R})$ and set $U_t = u(t, V_t)$, for $t \in [0,T]$.

For any $t \in [0,T]$, the Ito's formula implies that:

$$\begin{aligned} du(t,V_t) = {} & \frac{\partial u}{\partial t}(t,V_t) + [rV_{t-} + m(\mu_t - r)(V_{t-} - F_t)]\frac{\partial u}{\partial x}(t,V_t) \\ & + \frac{1}{2}(m\sigma_t)^2(V_{t-} - F_t)^2\frac{\partial^2 u}{\partial x^2}(t,V_t)dt + m\sigma_t(V_{t-} - F_t)\frac{\partial u}{\partial x}(t,V_t)dW_t \\ & + [u(t, V_{t-} + m(V_{t-} - F_t)Y_t) - u(t,V_{t-})]dN_t. \end{aligned}$$

A comparison between the above two equations implies in particular

$$\eta_{t-} = \frac{\partial u}{\partial x}(t, V_t)$$

and

$$\begin{aligned}&\frac{\partial u}{\partial t}(t,v) + (rv + m(\mu_t - r)(v-f))\frac{\partial u}{\partial x}(t,v) + \frac{1}{2}m^2\sigma_x^2(v-f)^2\frac{\partial^2 u}{\partial x^2}(t,v)\\ &= ru(t,v) + m(\mu_t - r)(v-f)\frac{\partial u}{\partial x}(t,v).\end{aligned}$$

Thus

$$\frac{\partial u}{\partial t}(t,v) + rv\frac{\partial u}{\partial x}(t,v) + \frac{1}{2}m^2\sigma_t^2(v-f)^2\frac{\partial^2 u}{\partial x^2}(t,v) - ru(t,v) = 0$$

and

$$mz(v-f)\frac{\partial u}{\partial x}(t,v) = u(t, v + m[v-f]z) - u(t,v)$$

with the final condition $u(T,v) = g(v)$. □

The rationale in constructing self-financed trading strategies that hedge the CPPI portfolio's terminal price is that there are contingent claims particularly useful to control both the closing-out-effect and the gap risk. As an example consider the case of a Vanilla options based on the CPPI portfolio's value. For instance being long in an at-the-money Put option on the portfolio with a strike at least equal to the protection required is a natural way to hedge gap risk. Similarly being long in an at-the-money Call option on the portfolio is a natural way to invest in a CPPI's portfolio preserving the capability to not purse forward the investment in the case of closed out.

12.4.2 Martingale Approach

It is possible to obtain a Black–Scholes like formula for the pricing of Vanilla options based on the CPPI portfolio:

We first consider the general case of jump-diffusion model. In this case, we assume $\ln(1+Y_i)$ are i.i.d. with the common density function f_Q.

Proposition 12.8. *Let the density of* $\ln(1+Y_i)$ *be* $f_Q(x)$ *and the density function of*

$$\ln(L_t) = \int_0^t \left[(1-\psi_s)\lambda_s - \frac{1}{2}\theta_s^2\right] ds + \int_0^t \theta_s dW_s + \int_0^t \ln\psi_s dN_s$$

be f^{L_t}, where L_t is defined by (12.8). Then the vanilla Call/Put option on the whole CPPI portfolio's value at maturity is completely determined by:

$$\begin{aligned} &Call(0,v,T,K) \\ &= \sum_{k=0}^{\infty}\left[\frac{e^{-\int_0^T \psi_s\lambda_s ds}(\int_0^T \psi_s\lambda_s ds)^k}{k!}\right]\int_{\varsigma}^{\infty}\left(C_0e^x+F_0-e^{-rT}K\right)p^{(k)}dx \end{aligned}$$

and

$$\begin{aligned} &Put(0,v,T,K) \\ &= \sum_{k=0}^{\infty}\left(\frac{e^{-\int_0^T \psi_s\lambda_s ds}(\int_0^T \psi_s\lambda_s ds)^k}{k!}\right)\int_{\varsigma}^{\infty}(-C_0e^x-F_0+e^{-rT}K)p^{(k)}dx, \end{aligned}$$

where $K > F_T$ and

$$p^{(k)} = f_1 * \underbrace{f_{Q'}^{\mathbb{Q}} * \ldots * f_{Q'}^{\mathbb{Q}}}_{k \text{ terms}},$$

where $f_{Q'}$ and $f_{Q'}^{\mathbb{Q}}$ have the following relation:

$$\int_{\mathbb{R}}\exp\left\{iuf_{Q'}^{\mathbb{Q}}(z)\right\}dz = \int_{\mathbb{R}}\exp\left\{\left[f_{Q'}(\frac{z}{iu})\frac{z}{iu}\right] * f^{L_T}(z)\right\}dz$$

and f_1 is the density function of the normal distribution

$$\mathcal{N}\left(\cdot, \int_0^T\left((m(-Y\psi_s\lambda_s)-\frac{m^2\sigma_s^2}{2}\right)ds, \int_0^T m\sigma_s dW_s^{\mathbb{Q}}\right)$$

and $\varsigma = \ln\left(\frac{e^{-rT}K-F_0}{C_0}\right)$.

Proof. Consider the process:

$$\begin{aligned} V_t &= C_0\exp\left\{\int_0^t\left(r+m(\mu_s-r)-\frac{m_s^2\sigma_s^2}{2}\right)ds+\int_0^t m\sigma_s dW_s\right\}\left[\prod_{n=1}^{N_t}(1+mY_n)\right]+F_t \\ &= C_0\exp\left\{\int_0^t\left(r+m(\mu_s-r)+m\sigma_s\theta_s-\frac{m^2\sigma_s^2}{2}\right)ds+\int_0^t m\sigma_s dW_s^{\mathbb{Q}}\right\} \\ &\quad\times\left[\prod_{n=1}^{N_t}(1+mY_n)\right]+F_t \\ &= C_0\exp\left\{\int_0^t\left(r-mY_s\psi_s\lambda_s-\frac{m^2\sigma_s^2}{2}\right)ds+\int_0^t m\sigma_s dW_s^{\mathbb{Q}}\right\}\left[\prod_{n=1}^{N_t}(1+mY_n)\right]+F_t. \end{aligned}$$

In the case of $N_T = k$, we denote

$$L^{(k)} = e^{-rT}\frac{V_T^k - F_T}{C_0} = \exp\left\{\int_0^T \left(m(-Y\psi_s\lambda_s) - \frac{m^2\sigma_s^2}{2}\right)ds + \int_0^T m\sigma_s dW_s^Q + \left[\sum_{n=1}^{k}\ln(1+mY_n)\right]\right\}.$$

(see the Remark 12.5 below the proof.) Because

$$\mathbb{P}[\ln(1+mY_i) \le z] = \mathbb{P}\left[\ln(1+Y_i) \le \ln\left(1+\frac{e^z-1}{m}\right)\right] = \int_{-\infty}^{\ln\left(1+\frac{e^z-1}{m}\right)} f_Q(y)dy,$$

the density function $f_{Q'}$ of the random variable $\ln(1+mY_i)$ under the Probability measure $\mathbb{P}$ is

$$f_{Q'}(z) = \frac{d[\mathbb{P}(\ln(1+mY_i) \le z)]}{dz} = f_Q\left[\ln\left(1+\frac{e^z-1}{m}\right)\right]\frac{e^z}{m+e^z-1}.$$

Let the density function of the random variable $\ln(1+mY_i)$ under Probability measure $\mathbb{Q}$ be $f_{Q'}^{\mathbb{Q}}$. By the properties of the Radon–Nikodym derivative and the characteristic function, we have

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}[\exp\{iu\ln(1+mY_i)\}] &= \mathbb{E}[\exp\{iu\ln(1+mY_i)\}L_T] \\ &= \mathbb{E}[\exp\{iu\ln(1+mY_i) + \ln L_T\}]\end{aligned}$$

Under $\mathbb{P}$, the density function of $iu\ln(1+mY_i)$ is $f_{Q'}\left(\frac{z}{iu}\right)\frac{z}{iu}$, thus the density function of $iu\ln(1+mY_i)+\ln L_T$ under $\mathbb{P}$ is

$$\left[f_{Q'}\left(\frac{z}{iu}\right)\frac{z}{iu}\right] * f^{L_T}(z)$$

and thus $f_{Q'}$ and $f_{Q'}^{\mathbb{Q}}$ have the following relation:

$$\int_{\mathbb{R}} \exp\left\{iuf_{Q'}^{\mathbb{Q}}(z)\right\}dz = \int_{\mathbb{R}} \exp\left\{\left[f_{Q'}\left(\frac{z}{iu}\right)\frac{z}{iu}\right] * f^{L_T}(z)\right\}dz.$$

Since

$$\begin{aligned}&\int_0^T \left[m(-Y\psi_s\lambda_s) - \frac{m^2\sigma_s^2}{2}\right]ds + \int_0^T m\sigma_s dW_s^{\mathbb{Q}} \\ &\sim \mathcal{N}\left(\cdot, \int_0^T \left[m(-Y\psi_s\lambda_s) - \frac{m^2\sigma_s^2}{2}\right]ds, \int_0^T m\sigma_s dW_s^{\mathbb{Q}}\right),\end{aligned}$$

we denote its density function by

$$f_1(x) = \phi\left(x, \int_0^T \left[m(-Y\psi_s\lambda_s) - \frac{m^2\sigma_s^2}{2}\right]ds, \int_0^T m\sigma_s dW_s^{\mathbb{Q}}\right)$$

under the Probability measure $\mathbb{Q}$ where $\phi(x,m,\upsilon)=\frac{1}{\sqrt{2\pi\upsilon^2}}e^{-\frac{(x-m)^2}{2\upsilon^2}}$. Then the density function $p^{(k)}(x)$ of $L^{(k)}$ is

$$p^{(k)}=f_1*\underbrace{f_{Q'}^{\mathbb{Q}}*\ldots*f_{Q'}^{\mathbb{Q}}}_{k\text{ terms}}.$$

We have

$$\mathbb{E}^{\mathbb{Q}}\left[e^{-rT}(V_T^{(k)}-K)^+\right]=\int_{\varsigma}^{\infty}\left(C_0e^x+F_0-e^{-rT}K\right)p^{(k)}dx,$$

where

$$\varsigma=\ln\left(\frac{e^{-rT}K-F_0}{C_0}\right),$$

thus

$$\begin{aligned}
&\text{Call}(0,v,T,K)\\
&=\mathbb{E}^{\mathbb{Q}}\left[e^{-rT}(V_T-K)^+\right]=\sum_{k=0}^{\infty}\mathbb{E}^{\mathbb{Q}}\left[e^{-rT}(V_T^{(k)}-K)^+\right]\left[\frac{e^{-\int_0^T\psi_s\lambda_sds}(\int_0^T\psi_s\lambda_sds)^k}{k!}\right]\\
&=\sum_{k=0}^{\infty}\left[\frac{e^{-\int_0^T\psi_s\lambda_sds}(\int_0^T\psi_s\lambda_sds)^k}{k!}\right]\int_{\varsigma}^{\infty}(C_0e^x+F_0-e^{-rT}K)p^{(k)}dx.
\end{aligned}$$

Similarly,

$$\mathbb{E}^{\mathbb{Q}}\left[e^{-rT}(K-V_T^{(k)})^+\right]=\int_{\varsigma}^{\infty}(-C_0e^x-F_0+e^{-rT}K)p^{(k)}dx$$

and

$$\begin{aligned}
&\text{Put}(0,v,T,K)\\
&=\mathbb{E}^{\mathbb{Q}}\left[e^{-rT}(K-V_T)^+\right]=\sum_{k=0}^{\infty}\mathbb{E}^{\mathbb{Q}}\left[e^{-rT}(K-V_T^{(k)})^+\right]\left[\frac{e^{-\int_0^T\psi_s\lambda_sds}(\int_0^T\psi_s\lambda_sds)^k}{k!}\right]\\
&=\sum_{k=0}^{\infty}\left[\frac{e^{-\int_0^T\psi_s\lambda_sds}(\int_0^T\psi_s\lambda_sds)^k}{k!}\right]\int_{\varsigma}^{\infty}(-C_0e^x-F_0+e^{-rT}K)p^{(k)}dx.
\end{aligned}$$

□

Remark 12.3. The expression is not very explicit since they contain measure transformations and convolutions.

Remark 12.4. When $\mathbb{Q}$ is the risk neutral measure, the price of a Vanilla Call option is given by,

$$\text{Call}(t,v,T,K) = E^{\mathbb{Q}}[e^{-r(T-t)}(V_T^{t,v}-K)^+] = E^{\mathbb{Q}}[e^{-r(T-t)}(V_T-K)^+|V_t=v],$$

for any $t \in [0,T]$. The CPPI portfolio's value $\{V_t\}$ is a Markov process so that

$$\text{Call}(t,v,T,K) = \text{Call}(0,v,T-t,K), \text{ for } t \in [0,T]$$

and it is sufficient to cover the case of the Vanilla Call option's price at zero.

Remark 12.5. The value of $1+mY_n$ might be negative, in this case $\ln(1+mY_n)$ is an imaginary number.

Corollary 12.3. *In Merton's Model and under the probability measure $\mathbb{P}^{rn}$, let the density of $\ln(1+Y_i)$ be $\phi(x,\alpha,\delta^2)$. Then the vanilla Call/Put option on the whole CPPI portfolio's value at maturity is completely determined by,*

$$Call(0,v,T,K) = \sum_{k=0}^{\infty}\left[\frac{e^{-\int_0^T \lambda_s ds}(\int_0^T \lambda_s ds)^k}{k!}\right]\int_{\varsigma}^{\infty}\left(C_0 e^x + F_0 - e^{-rT}K\right)p^{(k)}dx$$

and

$$Put(0,v,T,K) = \sum_{k=0}^{\infty}\left[\frac{e^{-\int_0^T \lambda_s ds}(\int_0^T \lambda_s ds)^k}{k!}\right]\int_{\varsigma}^{\infty}\left(-C_0 e^x - F_0 + e^{-rT}K\right)p^{(k)}dx,$$

where $K > F_T$ and

$$p^{(k)} = f_1 * \underbrace{f_{Q'} * \ldots * f_{Q'}}_{k \text{ terms}},$$

$$f_{Q'}(z) = \phi\left[\ln\left(1+\frac{e^z-1}{m}\right),\alpha,\delta^2\right]\frac{e^z}{m+e^z-1},$$

and f_1 is the density function of the normal distribution

$$\mathcal{N}\left[\cdot, \int_0^T\left(m\left(\lambda_s - e^{\alpha+\frac{\delta^2}{2}}\lambda_s\right) - \frac{m\sigma_s^2}{2}\right)ds, \int_0^T m\sigma_s dW_s\right]$$

and $\varsigma = \ln\left(\frac{e^{-rT}K-F_0}{C_0}\right)$.

In the following proposition we consider the special case that $Y_n = Y$ is a constant. In this case, the expression is more explicit.

Proposition 12.9. *In the case that $Y_n = Y$ is a constant, the vanilla Call/Put option on the whole CPPI portfolio's value at maturity has the explicit expression:*

$$Call(0,v,T,K) = \sum_{k=0}^{\infty} \left[\frac{e^{-\int_0^T \psi_s \lambda_s ds} \left(\int_0^T \psi_s \lambda_s ds \right)^k}{k!} \right]$$
$$\left[C_0 e^{M^{(k)} + \frac{1}{2}\sigma_{(k)}^2} \Psi \left(\frac{M^{(k)} + \sigma_{(k)}^2 - \varsigma}{\sigma_{(k)}} \right) - (F_0 - e^{-rT}K) \Psi \left(\frac{M^{(k)} - \varsigma}{\sigma_{(k)}} \right) \right]$$

and

$$Put(0,v,T,K) = \sum_{k=0}^{\infty} \left[\frac{e^{-\int_0^T \psi_s \lambda_s ds} \left(\int_0^T \psi_s \lambda_s ds \right)^k}{k!} \right]$$
$$\left[-C_0 e^{M^{(k)} + \frac{1}{2}\sigma_{(k)}^2} \Psi \left(-\frac{M^{(k)} - \sigma_{(k)}^2 + \varsigma}{\sigma_{(k)}} \right) + \left(-F_0 + e^{-rT}K \right) \Psi \left(\frac{-M^{(k)} + \varsigma}{\sigma_{(k)}} \right) \right],$$

where $K > F_T$ *and*

$$M^{(k)} = \int_0^T \left[(m - Y\psi_s\lambda_s) - \frac{m\sigma_s^2}{2} \right] ds + k \ln(1 + mY),$$
$$\sigma_{(k)}^2 = \int_0^T m\sigma_s dW_s^{\mathbb{Q}},$$

$$\varsigma = \ln \left(\frac{e^{-rT}K - F_0}{C_0} \right)$$

and

$$\Psi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt.$$

Proof. We have

$$V_t = C_0 \exp \left\{ \int_0^t \left[r + m(\mu_s - r) - \frac{m\sigma_s^2}{2} \right] ds \right.$$
$$\left. + \int_0^t m\sigma_s dW_s + \left[\sum_{n=1}^{N_t} \ln(1 + mY_n) \right] \right\} + F_t$$
$$= C_0 \exp \left\{ \int_0^t \left[r + m(-Y\psi_s\lambda_s) - \frac{m\sigma_s^2}{2} \right] ds \right.$$
$$\left. + \int_0^t m\sigma_s dW_s^{\mathcal{Q}} + \left[\sum_{n=1}^{N_t} \ln(1 + mY_n) \right] \right\} + F_t.$$

In case that $N_T = k$, we have

$$e^{-rT}\frac{V_T^k - F_T}{C_0} = \exp\left\{ \int_0^T \left[m(-Y\psi_s\lambda_s) - \frac{m\sigma_s^2}{2} \right] ds \right.$$
$$\left. + \int_0^T m\sigma_s dW_s^Q + \left[\sum_{n=1}^{N_T} \ln(1+mY_n) \right] \right\}.$$

Then we have

$$\ln\left(e^{-rT}\frac{V_T^k - F_T}{C_0} \right) \sim \mathcal{N}\left(.; M^{(k)}, \sigma_{(k)}^2 \right),$$

where

$$M^{(k)} = \int_0^T \left[m(-Y\psi_s\lambda_s) - \frac{m\sigma_s^2}{2} \right] ds + k\ln(1+mY)$$
$$\sigma_{(k)}^2 = \int_0^T m\sigma_s dW_s^{\mathbb{Q}}.$$

Thus

$$\mathbb{E}^{\mathbb{Q}}\left(e^{-rT}(V_T^{(k)} - K)^+ \right) = \int_\varsigma^\infty \left(C_0 e^x + F_0 - e^{-rT}K \right) d\left[\mathcal{N}\left(x; M^{(k)}, \sigma_{(k)}^2 \right) \right]$$
$$= C_0 e^{M^{(k)} + \frac{1}{2}\sigma_{(k)}^2} \Psi\left(\frac{M^{(k)} + \sigma_{(k)}^2 - \varsigma}{\sigma_{(k)}} \right) - \left(F_0 - e^{-rT}K \right) \Psi\left(\frac{M^{(k)} - \varsigma}{\sigma_{(k)}} \right),$$

where

$$\varsigma = \ln\left(\frac{e^{-rT}K - F_0}{C_0} \right).$$

and

$$\Psi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt.$$

Then

$$\text{Call}(0, v, T, K)$$
$$= \mathbb{E}^{\mathbb{Q}}[e^{-rT}(V_T - K)^+] = \sum_{k=0}^\infty \mathbb{E}^{\mathbb{Q}}\left[e^{-rT}(V_T^{(k)} - K)^+ \right] \left[\frac{e^{-\int_0^T \psi_s\lambda_s ds} (\int_0^T \psi_s\lambda_s ds)^k}{k!} \right]$$
$$= \sum_{k=0}^\infty \left[\frac{e^{-\int_0^T \psi_s\lambda_s ds} \left(\int_0^T \psi_s\lambda_s ds \right)^k}{k!} \right]$$
$$\left[C_0 e^{M^{(k)} + \frac{1}{2}\sigma_{(k)}^2} \Psi\left(\frac{M^{(k)} + \sigma_{(k)}^2 - \varsigma}{\sigma_{(k)}} \right) - (F_0 - e^{-rT}K)\Psi\left(\frac{M^{(k)} - \varsigma}{\sigma_{(k)}} \right) \right].$$

Similarly,

$$\mathbb{E}^{\mathbb{Q}}\left[e^{-rT}(K-V_T^{(k)})^+\right] = \int_{-\infty}^{\varsigma}\left(-C_0e^x - F_0 + e^{-rT}K\right)d\left[\mathcal{N}\left(x;M^{(k)},\sigma_{(k)}^2\right)\right]$$
$$= -C_0e^{M^{(k)}+\frac{1}{2}\sigma_{(k)}^2}\Psi\left(-\frac{M^{(k)}-\sigma_{(k)}^2+\varsigma}{\sigma_{(k)}}\right) + \left(-F_0+e^{-rT}K\right)\Psi\left(\frac{-M^{(k)}+\varsigma}{\sigma_{(k)}}\right)$$

and

$$\begin{aligned}
&\mathrm{Put}(0,v,T,K)\\
&= \mathbb{E}^{\mathbb{Q}}(e^{-rT}(K-V_T)^+) = \sum_{k=0}^{\infty}\mathbb{E}^{\mathbb{Q}}\left(e^{-rT}(K-V_T^{(k)})^+\right)\left(\frac{e^{-\int_0^T\psi_s\lambda_s ds}(\int_0^T\psi_s\lambda_s ds)^k}{k!}\right)\\
&= \sum_{k=0}^{\infty}\left[\frac{e^{-\int_0^T\psi_s\lambda_s ds}\left(\int_0^T\psi_s\lambda_s ds\right)^k}{k!}\right]\\
&\quad\left[-C_0e^{M^{(k)}+\frac{1}{2}\sigma_{(k)}^2}\Psi\left(-\frac{M^{(k)}-\sigma_{(k)}^2+\varsigma}{\sigma_{(k)}}\right) + \left(-F_0+e^{-rT}K\right)\Psi\left(\frac{-M^{(k)}+\varsigma}{\sigma_{(k)}}\right)\right].
\end{aligned}$$

□

Remark 12.6. The assumption of the jump Y_n be constant is not reasonable; however, it is a weighted sum of the option formula in Black–Scholes model with constant coefficients.

12.5 Mean-Variance Hedging

12.5.1 The Idea

Given a contingent claim H and if the financial market models do not allow arbitrage opportunities, in a complete market, H is attainable, i.e. there exists a self-financing strategy with final portfolio value $Z_T = H$, $\mathbb{P}$-a.s. However, when in our jump-diffusion model, the market is not complete and then H is not attainable. Quadratic hedging has been studied in more than 100 papers. It is used to hedge the incomplete market using the quadratic criterion. There are two approaches of quadratic hedging. One approach is risk-minimization; the other approach is mean-variance hedging. [15] is a review paper about quadratic hedging. Many symbols and definitions in this section are borrowed from that paper.

In our section, we consider the mean-variance hedging. For any contingent claim, let the payoff at T be H. Our jump-diffusion model of the risky asset price S is a semimartingale under $\mathbb{P}$. The following definition is taken from Sect. 4 in [15].

Definition 12.1. We denote by Θ_2 the set of all $\vartheta \in L(S)$ such that the stochastic integral process $G(\vartheta) := \int \vartheta dS$ satisfies $G_T \in L^2(\mathbb{P})$. For a fixed linear subspace Θ of Θ_2, a Θ-strategy is a pair $(Z_0, \vartheta) \in \mathbb{R} \times \Theta$ and its value process is $Z_0 + G(\vartheta)$. A Θ-strategy $(\tilde{Z}_0, \tilde{\vartheta})$ is called *Θ-mean-variance optimal* for a given contingent claim $H \in L^2$ if it minimizes $||H - Z_0 - G_T(\vartheta)||_{L^2}$ over all Θ-strategies (Z_0, ϑ) and $\tilde{Z}_0$ is then called the *Θ-approximation price* for H.

The linear subspace

$$\mathcal{G} := G_T(\Theta) = \left\{ \int_0^T \vartheta_u dS_u | \vartheta \in \Theta \right\}$$

of L^2 describes all outcomes of self-financing Θ-strategies with initial wealth $Z_0 = 0$ and

$$\mathcal{A} = \mathbb{R} + \mathcal{G} = \left\{ Z_0 + \int_0^T \vartheta_u dS_u | (Z_0, \vartheta) \in (\mathbb{R} \times \Theta) \right\}$$

is the space of contingent claims replicable by self-financing Θ-strategies. Our goal in mean-variance hedging is to find the projection in L^2 of H on $\mathcal{A}$ and this can be studied for a general linear subspace $\mathcal{G}$ of L^2 space. In analogy with the above definition, we introduce a *$\mathcal{G}$-mean-variance optimal* pair $(\tilde{Z}_0, \tilde{g}) \in \mathbb{R} \times \mathcal{G}$ for $H \in L^2$ and call $\tilde{Z}_0$ the *$\mathcal{G}$-approximation price* for H. In mathematics, our goal is to find

$$\min_{(Z_0, \vartheta) \in \mathbb{R} \times \Theta} ||H - Z_0 - G_T(\vartheta)||_{L^2}$$

Since

$$dS_t = S_{t-}[\mu_t dt + \sigma_t dW_t + Y_t dN_t],$$

our goal becomes to find

$$\begin{aligned}
&\min_{(Z_0, \vartheta) \in \mathbb{R} \times \Theta} \left|\left| H - Z_0 - \int_0^T \vartheta_u dS_u \right|\right|_{L^2} \\
&\min_{(Z_0, \vartheta) \in \mathbb{R} \times \Theta} \left|\left| H - Z_0 - \int_0^T \vartheta_u S_{u-} [\mu_u du + \sigma_u dW_u + Y_t dN_u] \right|\right|_{L^2} \\
&= \min_{(Z_0, \vartheta) \in \mathbb{R} \times \Theta} \left(\mathbb{E} \left\{ H - Z_0 - \int_0^T \vartheta_u S_{u-} [\mu_u du + \sigma_u dW_u + Y_t dN_u] \right\}^2 \right)^{\frac{1}{2}}.
\end{aligned}$$

[15] has pointed out that finding the optimal $\tilde{\vartheta}$ is general is an open problem. In the other hand, in real contingent claim pricing, we should always use the risk-neutral measure. [4] gives the $\mathcal{G}$-mean-variance optimal pair $(\tilde{Z}_0, \tilde{g})$ when the stocks' model is an exponential levy form martingale. For similar consideration, also see Chap. 10 in [3].

12.5.2 The Problem

Now we consider H as a function of V_T and denote $H = g(V_T)$. For any martingale measure $\mathbb{Q}$ defined in (12.9), we have proved that $\tilde{V}_t = e^{-rt}V_t$ is a $\mathbb{Q}$-martingale. Denote $\tilde{H} = e^{-rT}H$. We want to consider the following optimization problem.

$$\min_{(Z_0,\vartheta)\in\mathbb{R}\times\Theta} \mathbb{E}^{\mathbb{Q}}\left(\tilde{H} - Z_0 - \int_0^T \vartheta_u d\tilde{V}_u\right)^2. \tag{12.22}$$

Proposition 12.10. *The solution of the optimization problem* (12.22) *is*

$$Z_0 = \mathbb{E}^{\mathbb{Q}}\left[\tilde{H}\right];$$

$$\vartheta_t = \frac{\sigma_t \mathscr{C}_x(t,V_t) + \mathscr{C}(t,V_t + [V_{t-} - F_t]m_t Y_t) - \mathscr{C}(t,V_t)Y_t\lambda_t\psi_t}{\sigma_t + (V_{t-} - F_t)m_t Y_t^2\lambda_t\psi_t}.$$

Proof. We have

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}\left(\tilde{H} - Z_0 - \int_0^T \vartheta_u d\tilde{V}_u\right)^2 &= \mathbb{E}^{\mathbb{Q}}\left(\mathbb{E}^{\mathbb{Q}}\left[\tilde{H}\right] - Z_0 + \tilde{H} - \mathbb{E}^{\mathbb{Q}}\left[\tilde{H}\right] - \int_0^T \vartheta_u d\tilde{V}_u\right)^2 \\ &= \mathbb{E}^{\mathbb{Q}}\left[\left(\mathbb{E}^{\mathbb{Q}}\left[\tilde{H}\right] - Z_0\right)^2\right] \\ &\quad + \mathbb{E}^{\mathbb{Q}}\left(\tilde{H} - \mathbb{E}^{\mathbb{Q}}\left[\tilde{H}\right] - \int_0^T \vartheta_u d\tilde{V}_u\right)^2.\end{aligned}$$

We see that the optimal value for the initial capital is $Z_0 = \mathbb{E}^{\mathbb{Q}}\left[\tilde{H}\right]$.

Define $\mathscr{C}(t,x) = e^{rt}\mathbb{E}^{\mathbb{Q}}\left[\tilde{H}|V_t = x\right]$ and $\tilde{\mathscr{C}}(t,x) = e^{-rt}\mathscr{C}(t,x)$. By construction, $\tilde{\mathscr{C}}(t,x)$ is a $\mathbb{Q}$-martingale. We have deduced that

$$\begin{aligned}dV_t = {} & [rV_{t-} + (V_{t-} - F_t)m_t(\mu_t - r)]dt + (V_{t-} - F_t)m_t\sigma_t dW_t \\ & + (V_{t-} - F_t)m_t Y_t dN_t,\end{aligned}$$

and

$$d\tilde{V}_t = e^{-rt}\left[(V_{t-} - F_t)m_t\sigma_t dW_t^{\mathbb{Q}} + (V_{t-} - F_t)m_t Y_t dM_t^{\mathbb{Q}}\right].$$

Then by Ito's formula we have

$$\begin{aligned}& d\tilde{\mathscr{C}}(t,V_t) \\ = {} & \Big[-re^{-rt}\mathscr{C}(t,V_t) + e^{-rt}\mathscr{C}_t(t,V_t) + (rV_{t-} + (V_{t-} - F_t)m_t(\mu_t - r))e^{-rt}\mathscr{C}_x(t,V_t) \\ & + \frac{1}{2}(V_{t-} - F_t)^2 m_t^2\sigma_t^2 e^{-rt}\mathscr{C}_{xx}(t,V_t)\Big]dt + (V_{t-} - F_t)m_t\sigma_t e^{-rt}\mathscr{C}_x(t,V_t)dW_t \\ & + \left[e^{-rt}\mathscr{C}(t,V_t + [V_{t-} - F_t]m_t Y_t) - e^{-rt}\mathscr{C}(t,V_t)\right]dN_t \\ = {} & (V_{t-} - F_t)m_t\sigma_t e^{-rt}\mathscr{C}_x(t,V_t)dW_t^{\mathbb{Q}} \\ & + \left[e^{-rt}\mathscr{C}(t,V_t + [V_{t-} - F_t]m_t Y_t) - e^{-rt}\mathscr{C}(t,V_t)\right]dM_t^{\mathbb{Q}}.\end{aligned}$$

Thus we have

$$\begin{aligned}
&\tilde{H} - \mathbb{E}^{\mathbb{Q}}\left[\tilde{H}\right] - \int_0^T \vartheta_u d\tilde{V}_u \\
&= \tilde{\mathscr{C}}(T,V_T) - \tilde{\mathscr{C}}(0,V_0) - \int_0^T \vartheta_t e^{-rt}\left[(V_{t-}-F_t)m_t\sigma_t dW_t^{\mathbb{Q}} + (V_{t-}-F_t)m_t Y_t dM_t^{\mathbb{Q}}\right] \\
&= e^{-rt}\left[\int_0^T (V_{t-}-F_t)m_t\sigma_t(\mathscr{C}_x(t,V_t)-\vartheta_t)dW_t^{\mathbb{Q}}\right. \\
&\quad \left. + \int_0^T (\mathscr{C}(t,V_t+(V_{t-}-F_t)m_tY_t) - \mathscr{C}(t,V_t) - \vartheta_t(V_{t-}-F_t)m_tY_t)dM_t^{\mathbb{Q}}\right].
\end{aligned}$$

By the Isometry formula, we have

$$\begin{aligned}
&\mathbb{E}^{\mathbb{Q}}\left(\tilde{H} - \mathbb{E}^{\mathbb{Q}}\left[\tilde{H}\right] - \int_0^T \vartheta_u d\tilde{V}_u\right)^2 \\
&= e^{-2rt}\left(\mathbb{E}^{\mathbb{Q}}\left[\int_0^T [(V_{t-}-F_t)m_t\sigma_t(\mathscr{C}_x(t,V_t)-\vartheta_t)]^2 dt\right]\right. \\
&\quad \left. + \mathbb{E}^{\mathbb{Q}}\left[\int_0^T [\mathscr{C}(t,V_t+[V_{t-}-F_t]m_tY_t) - \mathscr{C}(t,V_t) - \vartheta_t(V_{t-}-F_t)m_tY_t]^2\lambda_t\psi_t dt\right]\right).
\end{aligned}$$

This is the minimizing problem with respect to ϑ_t. Differentiating the above expression with respect to ϑ_t and letting the first order derivative equal to 0, we have

$$\begin{aligned}
&(V_{t-}-F_t)m_t\sigma_t[\mathscr{C}_x(t,V_t)-\vartheta_t] + [\mathscr{C}(t,V_t+(V_{t-}-F_t)m_tY_t) \\
&-\mathscr{C}(t,V_t) - \vartheta_t(V_{t-}-F_t)m_tY_t](V_{t-}-F_t)m_tY_t\lambda_t\psi_t = 0,
\end{aligned}$$

thus

$$\vartheta_t = \frac{\sigma_t\mathscr{C}_x(t,V_t) + \mathscr{C}(t,V_t+[V_{t-}-F_t]m_tY_t) - \mathscr{C}(t,V_t)Y_t\lambda_t\psi_t}{\sigma_t + (V_{t-}-F_t)m_tY_t^2\lambda_t\psi_t}$$

□

Remarks: When the contingent claim is the call option with the strike price K, i.e. $H = (V_T - K)^+$, then

$$Z_0 = \mathbb{E}^{\mathbb{Q}}\left[\tilde{H}\right] = \text{Call}(0,V_0,T,K)$$

and

$$\mathscr{C}(t,x) = e^{rt}\mathbb{E}^{\mathbb{Q}}\left[\tilde{H}|V_t = x\right] = \text{Call}(t,x,T,K);$$

when the contingent claim is the put option with the strike price K, i.e. $H = (K - V_T)^+$, then

$$Z_0 = \mathbb{E}^{\mathbb{Q}}\left[\tilde{H}\right] = \text{Put}(0,V_0,T,K)$$

and

$$\mathscr{C}(t,x) = e^{rt}\mathbb{E}^{\mathbb{Q}}\left[\tilde{H}|V_t = x\right] = \text{Put}(t,x,T,K).$$

This is consistent with the calculation of call and put options.

12.6 Conclusion

In this chapter we considered CPPI by employing the jump-diffusion model. Consider an insurance firm which would like to reinvest certain amount of their income from their clients. The "floor" F can possibly be implicitly related to the calculation of the ruin probability of the company, where ruin would occur as an implicit function of F, probably with other additional ruin factors. The introduction of jump in the wealth process is suitable in situations when the investment portfolio includes, e.g., functions of the Moody Indices of some corporations. In addition to the classical martingale approach, which gives the closed-form solutions of the extended Black–Scholes type formula, the PIDE approach to the hedging problem is particularly useful to the situations when numerical calculations/programming are to be employed.

References

1. F. Black & A. Perold, Theory of Constant Proportion Portfolio Insurance. The Journal of Economic Dynamics and Control 16, 403–426(1992)
2. A. Cipollini, Capital Protection modeling the CPPI portfolio, working paper(2008)
3. R. Cont & P. Tankov, Financial Modelling With Jump Processes, Chapman & Hall/CRC financial mathematics series (2004)
4. R. Cont, P. Tankov, & E. Voltchkova, Hedging with options in models with jumps, Stochastic Analysis and Applications: Proc. Abel Sympos., Springer, pp. 197–218, 2007
5. M. Escobar, A. Kiechle, L. Seco & R. Zagst, Option on a CPPI, International Mathematical Forum, Vol. 6, no. 5, 229–262 (2011)
6. F. B. Hanson, Applied Stochastic Processes and Control for Jump-Diffusions: Modeling, Analysis, and Computation, Advances in Design and Control, 2007
7. J. Hull, Options, Futures, & Other Derivatives, 4th ed., Prentice-Hall, Englewood Cliffs, 2000
8. S. G. Kou, A Jump Diffusion Model for Option Pricing with Three Properties: Leptokurtic Feature, Volatility Smile, and Analytical Tractability, first draft, working paper, 1999
9. S. G. Kou, A Jump-Diffusion Model for option pricing, Management Science Vol.48, No. 8, 1086–1101 (2002)
10. S. G. Kou, Jump-Diffusion Models for Asset Pricing in Financial Engineering, in: J.R. Birge and V. Linetsky (eds), Handbooks in Operations Research and Management Science, Vol. 15, The Handbook of Financial Engineering, 73–116, 2007
11. K. Matsuda, Introduction to Merton Jump Diffusion Model, working paper, 2004
12. R. C. Merton, Option Pricing When Underlying Stock Returns are Discontinuous, Journal of Financial Economics 3 125–144 (1976)
13. A. Perold, A constant proportion portfolio insurance. Unpublished manuscript, Harvard Business School

14. W. Runggaldier, Jump Diffusion Models. In : Handbook of Heavy Tailed Distributions in Finance (S.T. Rachev, ed.), Handbooks in Finance, Book 1 (W. Ziemba Series Ed.), Elesevier/North-Holland, 169–209 (2003)
15. M. Schweizer, A Guided Tour through Quadratic Hedging Approaches, in: E. Jouini, J. Cvitanic, M. Musiela (eds.), Option Pricing, Interest Rates and Risk Management, Cambridge University Press 538–574 (2001)
16. P. Tankov and E. Voltchkova, Jump-diffusion models: a practitioner's guide, Banque et March'es, working paper, 2009

Part IV
Nonlinear State-Space Models for High Frequency Financial Data

Chapter 13
An Asymmetric Information Modeling Framework for Ultra-High Frequency Transaction Data: A Nonlinear Filtering Approach

Yoonjung Lee

13.1 Introduction

Understanding the joint dynamics of the price impact of a trade and the pattern of trading volume in financial markets is an important issue. The increasing availability of transaction level financial data allows empirical researchers to analyze the complex interactions among various market participants. For example, the duration between two consecutive trades might play a crucial role in conveying information to the market. Empirical researchers, however, face many challenges in analyzing a set of irregularly spaced data. Most statistical tools employed in time-series analysis are not well suited for this type of data. Besides the scarcity of standard statistical treatments, a more serious challenge lies in the lack of a theoretical framework applicable to transaction level data.

The main contribution of this chapter is to propose a continuous-time asymmetric information modeling framework that is applicable to analyzing irregularly spaced transaction data. The prosed framework specifies a full three-way interaction among the information structure, order arrivals, and price changes. Furthermore, with its dynamic nature, it can be used as a basis for empirical research. Not only does the model produce key implications consistent with the observed interactions among various market participants, but it also provides a theoretical explanation for the observed market dynamics under a traditional asymmetric information framework.

One of the first empirical models that takes into consideration the duration between trades is the ordered probit model introduced by Hausman et al. (1992) in [19]. They investigate the conditional distribution of price changes given a set of explanatory variables which includes the sequence of past prices and irregularly spaced order arrivals. Their cross-sectional analysis illustrates that the sequence of trades, not just the total volume of trades, affects the dynamics of price changes.

Y. Lee (✉)
Boston, MA, USA
e-mail: yoonjung.lee.lin@gmail.com

Y. Zeng and S. Wu (eds.), *State-Space Models: Applications in Economics and Finance*, Statistics and Econometrics for Finance 1, DOI 10.1007/978-1-4614-7789-1_13,

In their model, however, the sequence of order arrivals is used as an explanatory variable only. As a result, the feedback effects of price changes on the order arrival processes are overlooked.

In recent years, more sophisticated statistical approaches to analyzing ultra-high-frequency transactions data have been introduced by Engle (2000) in [14] and Dufour and Engle (2000) in [8]. In [14], the transaction arrival times and accompanying measures are formulated as marked point processes. In particular, an autoregressive conditional duration (ACD) model is applied to explicitly specify the dynamics of the time duration between order arrivals. Transactions data are summarized by two types of random variables. The first is the time of the transaction, and the second is a vector observed at the time of the transaction. The first variable is modeled as a process that accompanies the information on the second variables as marks. The marks in [14] consist of the volume of the contract, the price of the contract, and the posted bid and ask prices at the time. In similar vein to [14], Dufour and Engle (2000) in [8] explore a statistical model suitable for analyzing transactions data, generalizing the VAR model in [18] to incorporate the role of time between trades in stock price and trade processes. The results in [8] highlight the crucial role of duration in assessing the price impact of a trade. In particular, the price impact of a trade tends to increase as the time duration between two trades decreases, suggesting that increased trading activity would be associated with a higher level of information asymmetry. They find a positive autocorrelation of signed trades. Interestingly, a stronger positive autocorrelation is linked to higher trading intensity. This may suggest that informed traders gradually exploit their informational advantage by trading on one-side until their signal is revealed to the market. While their empirical investigation provides valuable insights into the crucial role of the inter-arrival time, it does not address how the price and volume dynamics would affect the inter-arrival time of trades. This limitation is mainly due to their exogeneity assumption on inter-trade durations. Note that they maintain the exogeneity assumption for the time process, treating inter-trade time durations as strongly exogenous to both the price and trade processes. Therefore, a theoretical link that ties the joint dynamics of the trade and volume processes would be helpful to present a complete picture of the market dynamics.

On the theoretical front, Kyle (1985) in [25] has laid a foundation for modeling the dynamics of the market with information asymmetry. An informed trader who privately observes the value of the asset optimizes his trading strategy in order to maximize his expected trading profits. Kyle's paper examines: the optimal trading strategy of the informed trader, the pricing rule that the market maker follows to set the price, and the speed at which this private information is incorporated into the market. The Kyle model has been extended in a number of directions. For instance, Back (1992) in [3] allows a more general distribution of the private signal and formally derives an equilibrium pricing rule, Aase et al. (2012) in [1] derive the optimal trading strategy of an insider in the presence of a time-varying but deterministic intensity of the liquidity traders, and finally Biagini et al. (2012) in [6] incorporate the persistent liquidity traders whose trading intensity is modeled by a fractional Brownian motion.

The stylized description of the economy in the Kyle model concisely summarizes the core of the equilibrium market dynamics. However, the continuous trading assumption may be unrealistic to some degree. In a market with frictions, investors would wait until the benefit of a trade exceeds the cost of a trade and then submit a lump-sum order. In this spirit, a strand of sequential trading models has been initiated by Glosten–Milgrom (1985) in [17]. In their model, orders are assumed to arrive sequentially at the specialist post. The market maker who observes the combined order flows from both informed and noise traders sets the price competitively. Despite its similarity to the order process of most financial markets, the Glosten–Milgrom-type model has not been explored as extensively as the Kyle-type model, perhaps due to its analytical intractability. A notable exception is the model by Back and Baruch (2004) in [4], which examines the informed traders' optimal submission of discrete orders. In their chapter, the optimal strategy of a single informed trader is numerically evaluated, under the assumption that the liquidation value of the firm follows a Bernoulli distribution. The optimal trading intensity is approximately linear in the magnitude of mispricing. I use their results to motivate the assumption for the informed traders' order arrival process in my model.

I consider a security market for a single risky asset with discrete order arrivals. In the market, there are three types of investors: informed traders, uninformed traders, and a market maker. Informed traders observe a noisy signal with regard to the liquidation value of the asset. The public information gradually arrives at the market, revealing a part of the private signal to the public. The gradual arrival of the public information also increases the quality of the informed traders' signal over time. Rationally revising the signal, each informed trader sets his reservation price to be the conditional expectation of the terminal value of the asset. An informed trader submits market orders when there is a discrepancy between his valuation and the current asset price. The intensity of his order flow depends on the magnitude of this discrepancy. A risk-neutral market maker who is assumed to know the probability structure of the order process sets the price under the zero profit condition. The current price is set at the conditional expectation of the terminal value of the asset, given the market maker's information. The market maker's information set includes not only the past history of order arrivals but also the public information. His inference problem involves rationally processing these two different sources of information. Formulating the dynamics of the market as a partially observed system, I apply a nonlinear filtering technique to provide a solution to this problem. Standard filtering techniques, which have been applied to other financial applications,[1] are not directly applicable to my model, due to the complexity of the information structure. Using enlarged filtration theory, as described in [15], I extend the standard filtering methodology. From a technical perspective, this extension is a novel contribution to the literature.

[1] For instance, Elliott (1997) in [12] applies a filtering approach to asset allocation problems. Frey and Runggaldier (2001) in [16] and Chib et al. (2002) in [26] use filtering in estimating stochastic volatility models. Zeng (2003) in [29] formulates the micro-movement of asset prices as a filtering problem, focusing on the discreteness of quoted prices.

There are some unique characteristics of the model that are particularly well suited for empirical studies. First, the generality of the structure produces various patterns of trade and volume. Second, the continual arrival of information ensures that the evolution of the market is truly dynamic and incorporates the learning processes of informed traders. Last, as a by-product of the filtering algorithm, maximum likelihood estimators can be obtained from the likelihood function. Simulation studies demonstrate that key predictions of the model are consistent with some recent empirical observations. Moreover, a theoretically consistent framework provides reasonable explanations for the observed phenomena from a traditional asymmetric information modeling point of view. With gradual revelation of the information, signed trades submitted by informed trades tend to exhibit positive autocorrelation. Given the constant trading rates of uninformed traders, a higher level of trading activity is associated with an increased positive autocorrelation of signed trades. In addition, trades arriving in a short time interval indicate an increased level of information asymmetry. Therefore, the price impact of a trade increases when the duration between trades decreases.

The information content carried in order arrival process is an important factor in setting prices. For example, Easley and O'Hara (1992) in [11] provide a model where a longer gap between trades reduces the price impact, because the market maker takes a low level of trading as a sign of a fewer information-based trades. The EKOP model, introduced in [10], has been a popular choice for many empiricists who examined the informational role in trading, cross-sectionally. At the beginning of each period, a piece of either good, bad, or no news arrives at the market independently. Informed traders participate in trading only when they observe a good or bad news and submit either buy orders or sell orders depending on the direction of news. The assumption on independent arrivals of information events retains the tractability of the model, because, under this assumption, a set of sufficient statistics consists of the total number of orders and the total number of order imbalances between buy and sell orders in each period. However the simplifying assumption leaves out the duration between trades from its analysis. Relaxing the static nature of the information structure in the EKOP model, Easley et al. (2008) in [9] consider a time-varying arrival rate model for informed and uninformed trades. A generalized autoregressive bivariate process is used to model the joint dynamics of these arrival rates. With the forecasted arrival rates, interesting interactions are found. For instance, both arrival rates tend to be highly consistent and uninformed traders tend to decrease their trading rates when informed traders trade more frequently. The relative strength of my model over the EKOP model lies in modeling the co-movement of price changes and trade arrivals under a dynamically evolving information structure. Time-varying trading rates of informed traders in the model are parsimoniously captured by stochastic changes in the size of the mispricing.

Last, I would like to point out that even with its fair generality, the proposed model is not a full-scale equilibrium model in that trading strategies of informed

traders are not optimized.[2] Rather the functional form is given exogenously. In addition, the aggregated trading rate of informed traders is obtained by simply assuming that each investor follows an identical trading strategy, regardless of how many informed traders are present in the market.

The rest of this chapter is constructed as follows: Sect. 13.2 outlines the model, describing the information structure and the inference problem that each market participant faces. Section 13.3 formalizes the market maker's inference problem as a filtering problem. Section 13.4 describes a simulation procedure and summarizes key implications of the model. In Sect. 13.5, I provide a procedure to estimate the parameters and discuss the sampling distribution of the parameter estimates. Finally, Sect. 13.6 concludes with a brief summary and directions for future research.

13.2 The Model

Following the setup described in [25], I consider a single risky asset traded over a fixed time period $[0,T]$. The terminal time T can be interpreted as the liquidation time of the firm or the time when all the uncertainty in asset valuation is resolved (e.g., the time of a public earnings announcement). The log-liquidation value of the risky asset is denoted by $V(T)$.

In the market, three kinds of investors participate in trading: informed traders, uninformed traders, and a market maker. Informed traders have superior knowledge of the distribution of the terminal value of the asset. They submit market orders to the market maker in order to exploit the mispricing with respect to their superior information set. The risk-neutral market maker absorbs combined market orders and sets the price competitively.

13.2.1 The Information Structure Dynamics

In order to fully motivate the structure of the model, I will build the information structure, adding one layer at a time.

[2] In a continuous trading model, Kyle (1985) in [25] endogenizes the trading rate of a single informed trader. Holden and Subrahmanyam (1992) in [20] demonstrate that multiple informed investors who observe a common signal tend to trade more aggressively, revealing their information at a higher speed, than a single informed trader would have otherwise. A recent paper of [4] endogenizes the trading intensity of a single informed trader who submits his order sequentially. Along another line of research, implications of strategic behavior of noise traders have been under great scrutiny. Admati and Pfleiderer (1988) [2] examine the interaction between strategic informed traders and strategic liquidity traders who have some discretion over when to place their trades. Bhushan (1991) in [5] models cross-sectional variation in trading costs that arise from a strategic behavior of discretionary liquidity traders.

13.2.1.1 The Perfect Signal Case Without Public Information Revealed

First I consider a case where informed traders observe a perfect signal and their trading activity is the only source of information in the market. In this case, informed traders observe $V(T)$ at the beginning of the period, and this is the only source of information in the market. As a result, the information structure is static. Some early market microstructure models, such as [25] and [3], are built on a similar information structure. As is shown in [25], the optimal trading strategy for a monopolistic informed trader, when continuous trades are allowed, is proportional to the magnitude of mispricing, namely the difference between $V(T)$ and the current price.

With the private signal being the only source of information and the trading activity being the only channel through which this information is transmitted, the price is completely determined by the combined order flows from informed and uninformed trades. However, in most financial markets, there are many other factors that influence the price formation process other than information carried by trades. Among those factors are the macro-economic conditions of the market, updated earnings estimates announced by investment banks, or buy and sell recommendations of analysts. If these multiple sources of information are available to the public, the price would be adjusted to reflect this information, in addition to the information conveyed through order arrivals. In the next section, I describe how this public information is explicitly modeled.

13.2.1.2 The Perfect Signal Case With Public Information Revealed

Next, consider a more general setting in which the private information is slowly revealed to the public. I have two objectives in mind in modeling the public information process. The first objective is to build a dependency structure between the public and private information. The second objective is to incorporate a learning process of the public, by introducing gradual improvements of the information over time. It would be desirable to have a structure in which the public estimates $V(T)$ with more accuracy, as the terminal time gets closer.

In order to achieve these objectives while maintaining tractability, I model $V(T)$ as the terminal value of some stochastic process, say $\{V(t) : 0 \leq t \leq T\}$, with $V(t)$ revealed to the public at time t. Informed traders are capable of acquiring and processing the information more quickly than the public. The public information would lead the log-price of the asset to slowly converge to $V(T)$, even in the absence of informed traders. This gives us a natural interpretation of the terminal time as the time when the information asymmetry is completely resolved. For modeling purposes, I assume $V(\cdot)$ to be a Brownian motion with variance σ^2. Based on the public information only, the variance of the prediction error of estimating $V(T)$ at time t is $\sigma^2(T-t)$. Therefore, the public information is effectively modeled to be revealed over time at a constant rate.

Addressing the evolution of the public information explicitly is a unique feature of the model. The role of public information has not been fully explored in

Glosten–Milgrom-type models, even though the original Glosten–Milgrom paper (1985) incorporates it in the market maker's information set. I believe that this is mainly because designing a theoretically consistent mechanism to process multiple sources of information is a challenging problem. As will be discussed later, I provide a solution to this problem with the aid of the filtering approach.

13.2.1.3 The Imperfect Signal Case With Public Information Revealed

So far, unlike the public information that is gradually revealed, the private signal has been assumed to be perfectly observable to informed traders at the beginning of the period, leaving no room for incorporating a learning process for informed traders. The last layer I add to the information structure is a dynamic learning process of the informed traders.

Informed traders are now assumed to observe a noisy path of the $V(\cdot)$ process rather than the perfect signal, $V(T)$. To be explicit, I model their imperfect signal at the beginning of the period, to be the whole path of a noisy signal, $\{V(t)+W(t) : 0 \le t \le T\}$, where the noise process $W(\cdot)$ is again modeled as a Brownian motion. I assume that the processes $W(\cdot)$ and $V(\cdot)$ are independent and denote the variance parameter of $W(\cdot)$ by γ^2. Upon the arrival of the public information at time t, informed traders are able to separate $W(t)$ from their noisy signal $V(t)+W(t)$, hence having a better estimate of $V(T)+W(T)$. Their learning process will be considered in detail in the next section.

In sum, the merits of explicitly modeling the public information are multi-fold. First, it provides a natural interpretation of the end of the trading period, T, as the time when the information asymmetry completely dissipates. Second, it allows us to decompose the price impact of a trade into two parts, one part due to the informational content of a trade and the other due to the arrival of new public information that is unrelated to order flow. Most asymmetric information models assume that trades are the only source of information. Their usage in empirical studies of transaction data is somewhat limited, since the observed price dynamics in financial markets are much more complex than can be explained by the trading activity alone. Last, from a modeling point of view, $V(t)$ concisely models natural learning processes of both informed and uninformed investors.

13.2.2 Informed Traders' Signal Extraction

In this section, I formalize the signal extraction procedure of informed traders, when informed traders observe an imperfect signal and the public information gradually reveals the private information. As briefly discussed above, informed traders are able to sharpen their signal as they gradually learn the public information.

I summarize the information available to informed traders as a sigma field, $\mathscr{F}_t^I$, defined by

$$\mathscr{F}_t^I = \sigma(\{V(u) : 0 \le u \le t\}) \vee \sigma(\{W(u) : 0 \le u \le t\}) \vee \sigma(\{V(u) + W(u) : t \le u \le T\}).$$

Then the conditional expectation of the terminal value of the asset given $\mathscr{F}_t^I$, denoted by $I(t)$, is shown below:

$$\begin{aligned} I(t) &\equiv E\left[\exp(V(T)) \,|\, \mathscr{F}_t^I\right] \\ &= E\left[\exp(V(T)) \,|\, \{(V(u), W(u)) : 0 \le u \le t\}, \{V(u) + W(u) : t \le u \le T\}\right] \\ &= E\left[\exp(V(t) + V(T) - V(t)) \,|\, \{V(t), W(t), V(T) + W(T)\}\right] \\ &= \exp\left(V(t) + \frac{\sigma^2}{\sigma^2 + \gamma^2}(V(T) - V(t) + W(T) - W(t)) + \frac{1}{2}\sigma^2(T-t)\frac{\gamma^2}{\sigma^2 + \gamma^2}\right). \end{aligned}$$

Note that I have used the independent increment property of Brownian motion and that the conditional distribution of $V(T) - V(t)$ given $V(T) - V(t) + W(T) - W(t)$ is also normal with the mean and variance given by,

$$\begin{aligned} &V(T) - V(t) | V(T) - V(t) + W(T) - W(t) \\ &\sim \mathscr{N}\left(\frac{\sigma^2}{\sigma^2 + \gamma^2}(V(T) - V(t) + W(T) - W(t)),\ \sigma^2(T-t)\frac{\gamma^2}{\sigma^2 + \gamma^2}\right). \end{aligned}$$

The mean-square prediction error at time t on estimating $V(T)$ based on the information available to informed traders is $\sigma^2(T-t)\frac{\gamma^2}{\sigma^2+\gamma^2}$, whereas the variance of the prediction error based on the public signal is $\sigma^2(T-t)$. Consequently, the ratio $\frac{\sigma^2}{\sigma^2+\gamma^2} = 1 - \frac{\gamma^2}{\sigma^2+\gamma^2}$ measures the variance reduction due to the private signal. The larger this ratio, the more important the role informed traders play in the market. In this sense, this ratio measures the degree of information asymmetry in the market.

13.2.3 Order Arrivals

In this section, I describe the probabilistic structure of the order arrival processes. With an application to transaction level data in mind, I follow [17] and consider a market where orders arrive sequentially. Uninformed traders submit both buy and sell orders with intensity η. Let $P(t)$ denote the current price of the asset set by the market maker at the expected value of $\exp(V(T))$ based on the market maker's information. Informed traders submit market buy (sell) orders, if $P(t)$ is lower (higher) than his reservation price $I(t)$, according to counting processes whose intensities are $\alpha(\log(I(t)) - \log(P(t)))^+$ for buy orders and $\alpha(\log(I(t)) - \log(P(t)))^-$ for sell orders, where x^+ (x^-) is the positive (negative) part of x.

Summarizing, I model the total number of buy orders up to time t, denoted by $B(t)$, and the total number of sell orders up to time t, denoted by $S(t)$, as counting processes whose intensities are given by:

$$\begin{cases} \lambda_B(t) = \alpha(\log(I(t)) - \log(P(t)))^+ + \eta \\ \lambda_S(t) = \alpha(\log(I(t)) - \log(P(t)))^- + \eta \end{cases}$$

The parameter α governs how aggressively informed traders, as a whole, exploit their informational advantage. Therefore, it governs the speed at which the private signal is revealed to the market. Too high a value of α may not be desirable for informed traders, in terms of their total expected profits, since profitable future trading opportunities may be lost by revealing their private information too quickly. In fact, finding an optimal strategy for a single informed trader, even without considering the strategic behavior of other informed traders, is a challenging yet interesting problem. In an equilibrium market setting, [4] numerically evaluates an optimized strategy, though in a simpler setting, where $V(T)$ takes only two different values. The key insight from their research is that the trading intensity of a single informed trader does depend on the perceived mispricing and is approximately linear in the magnitude of mispricing.

So far I have described how the information structure evolves over time and how investors in the market submit their orders. In the next section, the market maker's inference problem will be discussed. Specifically, the Bayesian updating procedure of the market maker will be developed.

13.3 Bayesian Updating of the Market Maker's Beliefs via Filtering

The market maker is assumed to have correct beliefs on the probabilistic structure of the order arrival processes. The information set available to the market maker at time t includes the history of the public information V process and the history of the order processes B and S. Therefore, the market maker's information set, $\mathscr{F}_t^M$, can be summarized as

$$\mathscr{F}_t^M = \sigma\left((V(u), B(u), S(u)) : 0 \le u \le t\right).$$

The market maker is risk-neutral and sets his price competitively under the zero profit condition. In other words, the ask/bid price, $P_a(t)/P_b(t)$ is his unbiased estimate of $V(T)$ given the past information and given the assumption that a buy/sell order arrives at the market instantaneously at time t:

$$\begin{cases} P(t) &= E\left[\exp(V(T)) \,|\, \mathscr{F}_t^M\right]. \\ P_a(t) &= E\left[\exp(V(T)) \,|\, \mathscr{F}_{t-}^M,\ \Delta B(t) = 1\right]. \\ P_b(t) &= E\left[\exp(V(T)) \,|\, \mathscr{F}_{t-}^M,\ \Delta S(t) = 1\right]. \end{cases}$$

If a buy order arrives at time t, $P(t) = P_a(t)$, and the buy order is transacted at this price. Similarly, if a sell order arrives at time t, $P(t) = P_b(t)$, and the sell order is transacted at this price. The market maker updates his conditional distribution of the signal $V(T)$, as new information arrives at the market. The information extraction problem that the market maker faces fits a general setting of a filtering problem. A filtering problem is concerned with estimation of the state at time t of a given

stochastic system in which some of the information is hidden. Based on the incomplete information available at time t, a filtering algorithm computes the conditional distribution of hidden signals, or state variables.

The filtering approach, in general, has been used in various financial applications. The two main approaches to deriving filtering equations are the innovations approach and the reference measure approach. The innovations approach hinges upon the martingale representation of the projection of the state process on the observed history. The reference measure approach involves constructing a reference measure under which the history of observations is independent of the unobserved state variables. A generalized Bayes formula provides the link between the objective measure and the reference measure. In this chapter, I follow the reference measure approach to compute the conditional distribution of the signal $V(T)$. I refer interested readers to [13] for a wide range of applications of the reference measure approach.

For a formal description of the filtering problem, I consider a probability space $(\Omega, \mathscr{F}, P)$ with a filtration $\{\mathscr{F}_t\}_{0\le t\le T}$. Under the measure P and with respect to the filtration $\mathscr{F}_t$, $V(\cdot)$ and $W(\cdot)$ are independent Brownian motions with variance parameters σ^2 and γ^2, respectively. Consider two independent unit Poisson processes $\xi_B(\cdot)$ and $\xi_S(\cdot)$ so that $\xi_B(t)-t$ and $\xi_S(t)-t$ are martingales with respect to $\mathscr{F}_t$ and are independent of $V(\cdot)$ and $W(\cdot)$. Then the order arrival processes $B(\cdot)$ and $S(\cdot)$ can be expressed as follows with λ_B and λ_S defined in Sect. 13.2:

$$\begin{cases} B(t) = \xi_B\left(\int_0^t \lambda_B(u)du\right) \\ S(t) = \xi_S\left(\int_0^t \lambda_S(u)du\right) \end{cases}$$

so that $B(t)-\int_0^t \lambda_B(u)du$ and $S(t)-\int_0^t \lambda_S(u)du$ are martingales under P with filtration $\mathscr{F}_t$.

Under the measure P, the observed quantities, namely the history of the V, B, and S processes up to time t, depend on the hidden variable $V(T)$. Indeed, the observed processes interact with each other through $V(T)$. The key step in the reference measure approach is to construct a reference measure, say Q, under which the observed processes are independent of the unobserved processes. Due to the independence between the observed and unobserved processes, certain computations such as taking expectations become analytically easier to handle under the reference measure Q than the objective measure P. However, the inference needs to be based on the objective measure P. So, there needs to a link that connects the inferences made under the reference measure to the inferences made under the objective measure. A generalized version of the classic Bayes formula described below provides this link.

Bayes Formula (Rule):

If $(\Omega, \mathscr{F}, Q)$ is a probability space and P is also a probability measure on $(\Omega, \mathscr{F})$ such that

$$L = \frac{dP}{dQ},$$

then for any sub-σ-algebra $\mathscr{G} \subset \mathscr{F}$ and $L^1(P)$-random variable Z,

$$E^P[Z|\mathcal{G}] = \frac{E^Q[ZL|\mathcal{G}]}{E^Q[L|\mathcal{G}]}.$$

In the next section, I illustrate how to construct a reference measure Q, defining the Radon–Nikodym derivative of P with respect to Q.

13.3.1 Construction of a Reference Measure

I construct a reference measure under which the state variables $V(T)$ and $W(T) - W(t)$ are independent of the market maker's information set, $\mathcal{F}_t^M$, in two steps. The construction requires two steps, because the price $P(t)$ incorporates two sources of information: one source from the order flow and the other from the public information, $V(t)$. The first change of measure ensures the buy and sell processes $B(\cdot)$ and $S(\cdot)$ become independent of the firm value process V. The second change of measure ensures the terminal value of the firm $V(T)$ becomes independent of the public information, $\{V(u) : 0 \le u \le t\}$.

As a first step, I construct a reference measure Q_1 under which the processes $B(\cdot)$ and $S(\cdot)$ are independent unit Poisson processes that are independent of $V(\cdot)$. Filtering problems with counting process observations are developed in depth in [7]. Girsanov's theorem on the change of measure for Poisson processes gives

$$\begin{aligned} L_1(t) &\equiv \frac{dP}{dQ_1}(t) \\ &= \exp\Big(\int_0^t \log(\lambda_B(u-))dB(u) - \int_0^t (\lambda_B(u) - 1)du \\ &\quad + \int_0^t \log(\lambda_S(u-))dS(u) - \int_0^t (\lambda_S(u) - 1)du \Big). \end{aligned}$$

In a stochastic differential equation (SDE) form,

$$L_1(t) = 1 + \int_0^t (\lambda_B(u-) - 1)L_1(u-)(dB(u) - du) + \int_0^t (\lambda_S(u-) - 1)L_1(u-)(dS(u) - du).$$

Under the measure Q_1, the observed order arrival processes $B(\cdot)$ and $S(\cdot)$ are independent of $V(\cdot)$ and $W(\cdot)$.

The market maker's information set, $\mathcal{F}_t^M$, however, includes the history of V up to time t, as well as the history of the order arrival processes. Hence, the second step of the construction involves finding a reference measure Q_2 under which $V(T)$ is independent of $\{V(u) : 0 \le u \le t\}$. Constructing such a reference measure involves several well-known results on Brownian motion, such as the enlargement of filtration technique, a decomposition of the Brownian motion, and Girsanov's theorem. The enlargement of filtration technique has been used in other financial applications, such as in [15] and in [27]. The theory on the decomposition of a Brownian motion and Girsanov's theorem can be found in [23].

I first consider an enlarged filtration, denoted by $\mathscr{G}_t$, which is enlarged to include $V(T)$:

$$\mathscr{G}_t = \mathscr{F}_t^M \vee \sigma(V(T)).$$

Then a standard result on Brownian motion allows us to decompose $V(t)$ into the following two components,

$$V(t) = Z(t) + \int_0^t \frac{V(T) - V(u)}{T-u} du,$$

where $Z(t)$ is a Brownian motion with respect to the enlarged filtration, $\mathscr{G}_t$, and is independent of $V(T)$. Hence, $V(t)$, which is a martingale with respect to $\mathscr{F}_t^M$, becomes a semi-martingale with respect to $\mathscr{G}_t$.

Consider a process $Y(t)$, defined by

$$Y(t) \equiv \int_0^t (T-u) d\left(\frac{V(u)}{T-u}\right)$$

so that

$$Y(t) = Z(t) + \int_0^t \frac{V(T)}{T-u} du.$$

Notice that $Y(t)$ is decomposed into two pieces: $Z(t)$, a Brownian motion with respect to $\mathscr{G}_t$, and $\int_0^t \frac{V(T)}{T-u} du$, which is a function of the private signal, $V(T)$. In other words, $Y(t)$ has a standard form of the observed process, namely, the form of *"noise+signal."* Moreover, $Y(t)$ can be constructed from the observed process $V(t)$, giving

$$\mathscr{F}_t^M = \sigma\left((V(u), B(u), S(u)) : 0 \le u \le t\right) = \sigma\left((Y(u), B(u), S(u)) : 0 \le u \le t\right).$$

This allows us to model the observations by $Y(\cdot)$ rather than $V(\cdot)$. Currently, under the objective measure P, the process $Y(\cdot)$ is a semi-martingale with respect to the enlarged filtration $\mathscr{G}_t$ with its drift term $\frac{V(T)}{T-t}$. I construct a reference measure Q_2 under which the process $Y(\cdot)$ becomes a martingale, more precisely, a Brownian motion with variance parameter σ^2 with respect to the enlarged filtration $\mathscr{G}_t$. Applying Girsanov's theorem, which allows us to change the drift term of a Brownian motion, yields for $t < T$,

$$\begin{aligned} L_2(t) &\equiv \frac{dP}{dQ_2}(t) \\ &= \exp\left(\frac{1}{\sigma^2}\int_0^t \frac{V(T)}{T-u} dY(u) - \frac{1}{2\sigma^2}\int_0^t \frac{V(T)^2}{(T-u)^2} du\right) \\ &= 1 + \frac{1}{\sigma^2}\int_0^t \frac{V(T)}{T-u} L_2(u) dY(u). \end{aligned}$$

It can also be shown that under the reference Q_2, $Y(\cdot)$ is independent of the unobserved signal $V(T)$.

Finally, putting the pieces together, I define a reference measure Q such that

$$\frac{dP}{dQ}(t) = L(t),$$

where

$$\begin{aligned} L(t) &\equiv L_1(t)L_2(t) \\ &= \exp\Big(\int_0^t \log(\lambda_B(u-))dB(u) - \int_0^t (\lambda_B(u)-1)du \\ &\quad + \int_0^t \log(\lambda_S(u-))dS(u) - \int_0^t (\lambda_S(u)-1)du \\ &\quad + \frac{1}{\sigma^2}\int_0^t \frac{V(T)}{T-u}dY(u) - \frac{1}{2\sigma^2}\int_0^t \frac{V(T)^2}{(T-u)^2}du\Big). \end{aligned}$$

In other words,

$$\begin{aligned} L(t) = 1 &+ \int_0^t L(u-)\left(\lambda_B(u-)-1\right)d\left(B(u)-u\right) \\ &+ \int_0^t L(u-)\left(\lambda_S(u-)-1\right)d\left(S(u)-u\right) + \frac{1}{\sigma^2}\int_0^t L(u)\frac{V(T)}{T-u}dY(u). \end{aligned}$$

By the construction of the measures Q_1 and Q_2, under the reference measure Q, the following properties are satisfied: the processes $B(\cdot)$ and $S(\cdot)$ are independent unit Poisson processes and they are independent of $Y(\cdot)$ and $W(\cdot)$. $W(\cdot)$ is a Brownian motion with variance γ^2 and it is independent of $Y(\cdot)$. Furthermore, with respect to $\mathcal{G}_t$, $Y(\cdot)$ is a Brownian motion with variance σ^2 and $Y(\cdot)$ is independent of $V(T)$. Lastly, the Radon–Nikodym derivative $L(t)$ is a martingale with respect to $\mathcal{G}_t$.

13.3.2 Filtering Equation

With the reference measure constructed above, I apply a generalized Bayes formula, also known as the Kallianpur–Striebel formula [22], to compute the conditional distribution of $V(T)$ given the market maker's information. The informed traders' unbiased estimate of the asset value, denoted by $I(t)$, involves two quantities $V(t)$ and $V(T)-V(t)+W(T)-W(t)$. With $V(t)$ revealed to the public at time t, the observed order flows convey the information on the quantity $V(T)+W(T)-W(t)$, from which the market maker is ultimately interested in estimating $V(T)$. As a first step, the market maker computes the joint conditional distribution of $\{V(T), W(T)-W(t)\}$. Next, he/she obtains the marginal conditional distribution of $V(T)$ by integrating out the distribution of $W(T)-W(t)$. This leads us to consider the following quantity,

$$\phi(t,f) \equiv E^{Q}\left[f\left(V(T),W(T)-W(t)\right)L(t)|\mathscr{F}_t^M\right]$$

for some function f. $\phi(t,f)$ determines the conditional distribution of $(V(T), W(T)-W(t))$ for a general class of f. The Kallianpur–Striebel formula gives

$$\begin{aligned}\pi(t,f) &\equiv E^{P}\left[f\left(V(T),W(T)-W(t)\right)|\mathscr{F}_t^M\right] \\ &= \frac{E^{Q}\left[f\left(V(T),W(T)-W(t)\right)L(t)|\mathscr{F}_t^M\right]}{E^{Q}\left[L(t)|\mathscr{F}_t^M\right]} \\ &= \frac{\phi(t,f)}{\phi(t,\mathbf{1})}.\end{aligned}$$

The Zakai equation [21] that the unnormalized conditional distribution $\phi(t,f)$ satisfies is derived in Proposition 13.1.

Proposition 13.1. *$\phi(t,f)$ is the solution of the following stochastic partial differential equations (SPDEs):*

$$\begin{aligned}\phi(t,f) = \phi(0,f) &+ \int_0^t \phi\left(u-, f\left(\lambda_B - 1\right)\right) d\left(B(u)-u\right) + \int_0^t \phi\left(u-, f\left(\lambda_S - 1\right)\right) d\left(S(u)-u\right) \\ &+ \frac{1}{\sigma^2}\int_0^t \phi\left(u, g_1\right) dY(u) - \int_0^t \phi\left(u, g_2\right) du + \frac{\gamma^2}{2}\int_0^t \phi\left(u, f_{ww}\right) du,\end{aligned}$$

where

$$g_1(u,v,w) = f(v,w)\frac{v}{T-u},\ \ g_2(u,v,w) = f_w(v,w)\frac{w}{T-u},$$

and

$$P(t) = \frac{\phi(t,f_0)}{\phi(t,\mathbf{1})},$$

with

$$f_0\left(v,w\right) \equiv \exp\left(v\right).$$

(All proofs are in Appendix.)

The Zakai equation displayed in Proposition 13.1 may appear linear, but it is highly nonlinear in its nature, since the solution for $\phi(t,f)$ depends on $P(t)$, while $P(t)$ is in turn determined by $\phi(t,f)$.

13.3.3 Uniqueness of the System

As pointed out in [17], establishing the existence and uniqueness of the P_a and P_b is not trivial. Since the quoted prices affect the trading intensities and the trades are in part used in formulating the prices, there is a feedback effect of quoted prices in the system. This problem is similar to what is known as a fixed point problem in a

rational equilibrium model setting. Proposition 13.2 shows that prices are uniquely determined by filtering equations derived in the previous section. For technical details of the proof, I rely on methods developed in [24], where they examine the uniqueness of a class of nonlinear SPDEs.

Proposition 13.2. *P_a and P_b are uniquely determined by filtering equations in Proposition 13.1.*

(All proofs are in Appendix.)

13.4 Key Implications of the Model

This section investigates some of the key implications of the model, the imperfect signal case when informed traders observe a slowly improving imperfect signal.

The degree of information asymmetry in the market will be measured by the conditional variance of $V(T)$ given the market maker's information set and the quoted spread, $P_a(t) - P_b(t)$. The conditional variance of $V(T)$ defined below measures how much uncertainty remains in the market:

$$Var\left[V(T)|\mathcal{F}_t^M\right] = E^P\left[V(T)^2|\mathcal{F}_t^M\right] - \left(E^P\left[V(T)|\mathcal{F}_t^M\right]\right)^2.$$

The quoted spread, $P_a(t) - P_b(t)$, is frequently used quantity in the literature to measure liquidity. It measures how much the price would change if a market order arrived instantaneously. More precisely, $P_a(t) - P(t-)$ is the price impact of a buyer-initiated order arrival and $P(t-) - P_b(t)$ is the price impact of a seller-initiated order arrival.

13.4.1 The Quality of the Signal

The remaining uncertainty of the terminal value of the asset clearly depends on the quality of the signal for informed traders. When the signal that informed traders observe has a high noise-to-signal ratio (relatively high value of γ with respect to σ), the market maker faces higher uncertainty in estimating the value of the asset. When γ converges to zero, the conditional variance curve moves close to the conditional variance curve with the perfect signal case. The conditional variance curve with the perfect signal case then serves as a lower bound for the conditional variance plots. As in the perfect signal case, the line of $\sigma^2(T-t)$ serves as an upper bound, since if the signal is very noisy, then there is little advantage of observing the signal and the role of the public information dominates the role of the private information. The parameter γ governs how close the conditional variance is to either bound.

The signal that informed traders observe in the perfect signal case is static, in the sense that $V(T)$ is revealed to informed traders at the beginning of the time period.

When informed traders observe a noisy signal, the signal is gradually improved over time with a new piece of public information arriving continuously.

Since informed traders maintain their superiority of knowledge of the terminal value of the asset over the market maker ($\mathscr{F}_t^M \subset \mathscr{F}_t^I \vee \sigma(B(u), S(u)) : 0 \le u \le t)$),

$$\begin{aligned} P(t) &= E\left[\exp(V(T)) \,|\, \mathscr{F}_t^M\right] \\ &= E\left[E\left[\exp(V(T)) \,|\, \mathscr{F}_t^I \vee \sigma(B(u), S(u)) : 0 \le u \le t)\right] \,|\, \mathscr{F}_{t-}^M\right] \\ &= E\left[I(t) | \mathscr{F}_t^M\right]. \end{aligned}$$

Consequently, the price process can be interpreted as the conditional expectation of the signal process of the informed traders as well as the conditional expectation of the terminal asset value. You can think of $I(t)$ as a moving target that is continuously revised by informed traders and think of $P(t)$ as the market maker's best prediction of where this moving target is. In terms of estimating the terminal asset value, $P(t)$ does not perform as well as $I(t)$, since only a part of the private information is revealed through trades.

13.4.2 Informed Traders' Trading Rate

In order to investigate the impact of the trading intensity parameter α, the remaining parameters are fixed at $\eta = 50$, and $\sigma = 1$, and $\gamma = .5$. As in the perfect signal case, the trading intensity parameter α controls how rapidly the uncertainty is resolved. From Fig. 13.1, one can notice that under the imperfect signal assumption, a large value of α alone cannot decrease the conditional distribution to zero. The uncertainty about $V(T)$ that informed traders face is not resolved by increasing the trading intensity of informed traders. In fact, the lower bound of the conditional variance curve is $\frac{\gamma^2}{\sigma^2+\gamma^2}(T-t)$. As displayed in Fig. 13.1, when $\alpha = 160$, the conditional variance curve is close to $\frac{\gamma^2}{\sigma^2+\gamma^2}(T-t)$. When α approaches infinity, the private information is revealed immediately to the market.

13.4.3 The Price Impact of a Trade

The quoted spreads, the sequence of order arrivals, and the price dynamics are plotted in Fig. 13.2. The parameter values used in simulation are $(\alpha, \eta, \sigma, \gamma) = (40, 50, 1, .5)$. The plots in Fig. 13.2 are drawn on a large time scale to illustrate the relationship between the price impact of a trade and the duration between trades. In Fig. 13.2, quoted spreads, the sequence of order arrivals (a positive directional vertical bar indicates a buy order and a negative directional vertical bar a sell order), and the price dynamics are plotted in that order. A striking feature is the pattern of narrowing spreads, as the market maker waits for the next order to arrive. The market maker associates a high volume of trading, especially consecutive trades of the same sign, with a high level of information asymmetry. So, consecutive buy or

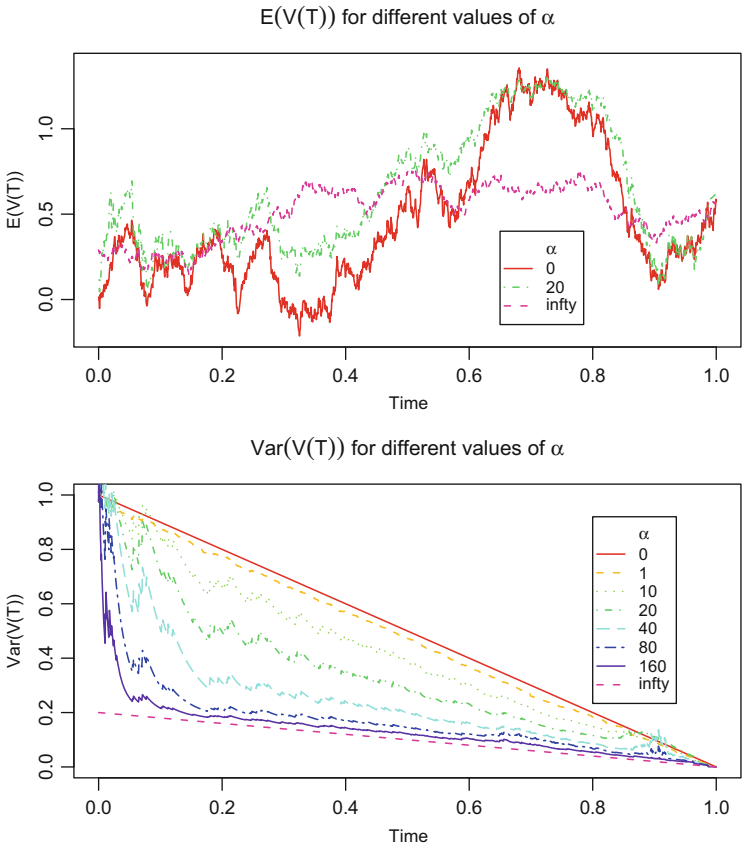

Fig. 13.1 Time-series plots of $\left(E\left[V(T)|\mathscr{F}_t^M\right], Var\left[V(T)|\mathscr{F}_t^M\right]\right)$ for different values of α in an imperfect signal case: the other parameters are fixed at $(\eta, \sigma, \gamma) = (50, 1, .5)$.

sell orders in a short time period should widen quoted spreads, resulting in a large price impact of a trade.

13.5 Parameter Estimation

There are two main approaches to estimating the parameters in the model. A Bayesian approach extends the state space to include the parameters in the model as state variables. With a given prior distribution of the state variables, the filtering algorithm produces the joint conditional distribution of the signal process and the parameters. The dimensionality quickly becomes an issue, since implementing a filtering equation of a high dimension of signal is generally computationally intractable. Several simulation algorithms have been suggested for enhancing the

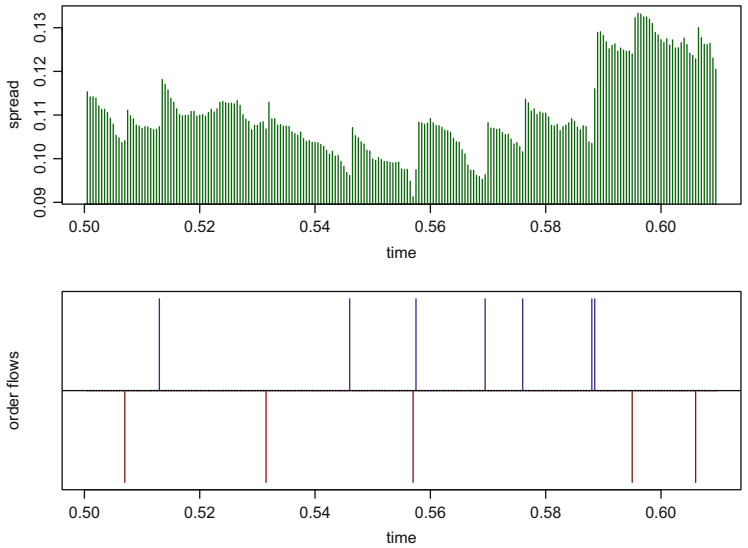

Fig. 13.2 Time-series plots of spreads, order arrivals, and price dynamics in an imperfect signal case: the parameter values used in simulation are $(\alpha, \eta, \sigma, \gamma) = (40, 50, 1, .5)$.

computational efficiency. For example, Pitt and Shephard (1999) in [28] suggest a particle filter algorithm for online Bayesian calculations about the parameters. Alternatively, one can take a classical maximum likelihood approach which involves finding a set of parameters that maximize the likelihood function.

13.5.1 Maximum Likelihood Estimation

I take a classical approach and consider a continuous-time likelihood function $L(\theta, T)$, defining $\theta = \{\gamma, \alpha, \eta\}$. Since σ is the volatility parameter of the public information process V, I assume that σ is a known parameter and focus on estimating the remaining parameters in the model, $\theta \equiv \{\alpha, \eta, \gamma\}$.

If all the state variables in the system are observable, a complete likelihood function is given by

$$\frac{dP}{dQ} = L(T, \theta),$$

where

$$L(T, \theta) = 1 + \int_0^T L(u-, \theta)\,(\lambda_B(u-, \theta) - 1)\,d\,(B(u) - u) + \int_0^T L(u-, \theta)\,(\lambda_S(u-, \theta) - 1)\,d\,(S(u) - u)$$

with

$$\begin{cases} \lambda_B(u) = \alpha\left(\log(I(u)) - \log(P(u))\right)^+ + \eta \\ \lambda_S(u) = \alpha\left(\log(I(u)) - \log(P(u))\right)^- + \eta \end{cases}$$

At the end of the period, I have full access to $\{P(\cdot), V(\cdot), B(\cdot), S(\cdot)\}$ for $t \in [0,T]$. However, the process $W(\cdot)$ is still unobserved at the end of the period. Hence, the complete likelihood function is not available for computing maximum likelihood estimators. The conditional likelihood function based on the observed quantities is

$$E^Q\left[L(T,\theta)|\mathscr{H}_T\right],$$

where

$$\mathscr{H}_t \equiv \sigma\left((B(u),S(u)) : 0 \le u \le t\right) \vee \sigma\left((V(u),P(u)) : 0 \le u \le T\right).$$

Proposition 13.3 shows that this conditional distribution function can be computed from a recursive filtering algorithm, similar to the filtering equation that is used for the market maker's inference problem.

Proposition 13.3.

$$\begin{aligned} E^Q\left[L(t,\theta)|\mathscr{H}_t\right] = \exp\Big(& \int_0^t \log\left(E^P\left[\lambda_B(u-,\theta)|\mathscr{H}_u\right]\right) dB(u) - \int_0^t \left(E^P\left[\lambda_B(u,\theta)|\mathscr{H}_u\right] - 1\right) du \\ & + \int_0^t \log\left(E^P\left[\lambda_S(u-,\theta)|\mathscr{H}_u\right]\right) dS(u) - \int_0^t \left(E^P\left[\lambda_S(u,\theta)|\mathscr{H}_u\right] - 1\right) du\Big). \end{aligned}$$

(All proofs are in Appendix.)

A maximum likelihood estimator of θ, denoted by $\hat{\theta}$, is then a quantity that maximizes $E^Q\left[L(T,\theta)|\mathscr{H}_T\right]$.

13.5.2 Parameter Estimation for Simulated Data

The simulation procedure can be summarized as follows: I first fix a set of parameter values $(\alpha,\eta,\gamma,\sigma)$ and the time interval $[0,T]$. For the given parameter values, the recursive filtering algorithm in Proposition 13.1 is applied to update the price process. For each sample path, assuming that public information is available in computing the likelihood function and henceforth σ is known, I first choose an initial value of (α,η,γ) to evaluate the likelihood function according to Proposition 13.3. In order to find a set of parameter values that maximizes the likelihood function, I use a nonlinear minimization algorithm built-in the statistical package R.

The sampling distribution of the MLEs for simulated data is summarized in Fig. 13.3. The parameter values of $(\alpha,\eta,\gamma,\sigma) = (20,20,2,1)$ and $T = 20$ are used. The parameter estimates are based on 250 simulated data sets. The histograms in Fig. 13.3 seem to suggest that the distributions of $\hat{\eta}$ and $\hat{\gamma}$ are fairly symmetric around the true values of the parameters, while the distribution of $\hat{\alpha}$ is skewed to the

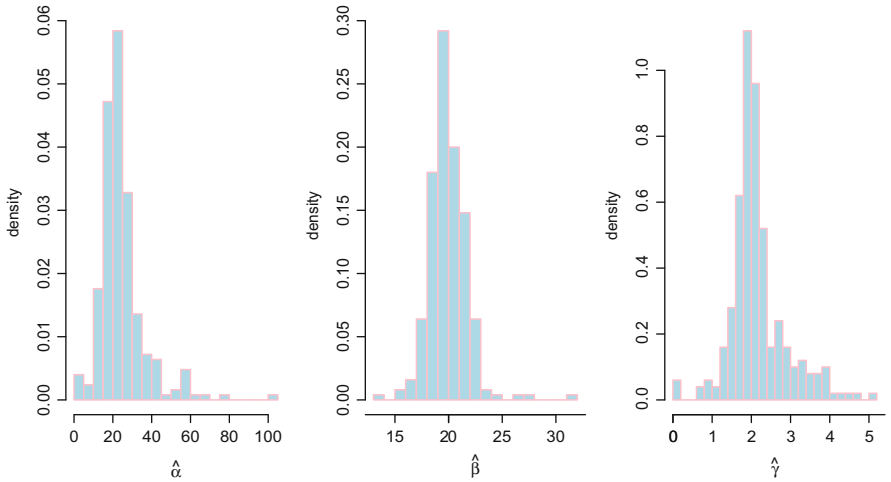

Fig. 13.3 Histograms of $(\hat{\alpha}, \hat{\eta}, \hat{\gamma})$ in an imperfect signal case: the true parameter values used in simulation are $(\alpha, \eta, \gamma) = (20, 20, 2)$.

Table 13.1 The summary statistics for $(\hat{\alpha}, \hat{\eta}, \hat{\gamma})$ in an imperfect signal case: the true parameter values used in simulation are $(\alpha, \eta, \gamma) = (20, 20, 2)$ with $T = 20$. The parameter estimates are based on 250 simulations.

	$\hat{\alpha}$	$\hat{\eta}$	$\hat{\gamma}$
Mean	24.55	19.92	2.15
Median	22.70	19.84	2.02
Std. dev.	11.86	1.81	0.72

right. The summary statistics are also presented in Table 13.1. The parameter estimates of α exhibit relatively wider sample variation than the other two parameters.

13.6 Conclusion

Formulating the economy as a partially observed dynamical system, I propose a new theoretical framework that is geared toward modeling transaction level data. A theoretical link that has been missing in the empirical market microstructure literature is the three-way interaction among the price impact of a trade, the duration between trades, and the degree of information asymmetry. The model fully specifies this interaction under a fairly general dynamically evolving information structure with a continuously distributed terminal value of the asset. The generality of the model demands a rigorous treatment of the price formulation process, for which a nonlinear filtering technique has been applied. The filter provides a computationally efficient recursive algorithm for Bayesian updating by the market maker.

The set of parameters in the model parsimoniously captures characteristics of the market and provides flexibility in modeling. The speed at which the private information is incorporated into the price depends on the trading rate of informed traders and the quality of their signal. The liquidity of the market, measured by quoted spreads, dynamically changes depending on the particular sequence of order arrivals. When the current price veers away from the informed traders' valuation, orders of the same sign are more likely to arrive. In turn, by observing the increased trading activity, the market maker widens his quoted spread in order to be compensated for bearing the risk of trading with informed traders. Meanwhile, he revises his view on the asset value and adjusts the price toward the private valuation, decreasing the magnitude of the mispricing. Simulation studies confirm that the proposed model reconciles some of the empirical findings reported in the market microstructure literature. In particular, the empirical role of time between trades and the impact of a particular pattern of order arrivals are captured in simulation studies.

Simulation studies on the maximum likelihood estimates demonstrate that the parameters that govern the uninformed trading rate and the quality of the signal process can be estimated with better accuracy than the parameter that governs the informed trading rate. The continuous trading model provides some insights into the magnitude of the bias of the maximum likelihood estimator. From a theoretical point of view, the results on the optimal trading rate of an insider is an extension of the Kyle (1985) model. When the information advantage of an insider gradually decreases as the public slowly learns his private signal, he/she trades more aggressively.

There are a number of interesting directions for future research. These can be grouped into two categories: theoretical issues and empirical issues. An important path for improving the current model is to fully endogenize the trading intensity for a single informed trader in the market. This problem can be formulated as a stochastic control problem. However, unlike most standard control problems, the control variable in the model would be of infinite dimension, since the conditional distribution of the asset value given the market maker's information becomes a quantity that the informed trader controls. Another interesting application is to develop an option pricing framework when investors are heterogeneously informed. A key quantity in pricing a contingent claim is the volatility parameter of the underlying asset. The filtering algorithm implemented produces the conditional distribution of the asset valuation, which can be used to estimate the volatility of the price.

Along the lines of empirical research, a cross-sectional analysis of transaction data would be of primary importance. In particular, I would like to compare the results of the model with those of other well-known microstructure models in [9] and [18]. Second, it would be interesting to examine the proposed mechanism for setting quoted prices against other algorithms that market makers implement in practice. The filtering algorithm developed may serve as a useful tool in designing an automated market maker system, since it successfully processes multiple sources of information that arrive at the market. Additionally, the proposed model has a potential for addressing some interesting research questions outside of market microstructure. For instance, the parameter that controls the intensity with which informed

traders trade could be correlated with the number of analysts who follow the stock. A cross-sectional comparison of the estimated parameters can be used to provide valuable insights as to what kind of stocks tend to be followed by a large number of analysts. Alternatively, after controlling for other firm characteristics, one could test if and to what extent the analyst coverage affects the firm's managerial decisions on investments or mergers and acquisitions. Furthermore, the quality of informed traders' signal is an important characteristic of the firm. It measures how transparent the firm's prospects are to informationally advanced traders. A low signal-to-noise ratio is an indication that the firm's future is rather opaque. Therefore, the quality of the signal indirectly measures the benefit of putting resources toward the fundamental analysis, from an investor's point of view.

Acknowledgments I am grateful to Thomas Kurtz for providing me with his expertise in the stochastic modeling and to Mark Ready for introducing me to the asymmetric information modeling literature and guiding me with his insightful comments. I am also grateful for helpful comments from Elizabeth Odders-White, Andrew Roper, and David Brown.

Appendix

Proof of Proposition 13.1

Proof. Note that

$$W(t) = \hat{W}(t) + \int_0^t \frac{W(T) - W(u)}{T-u} du,$$

where $\hat{W}(\cdot)$ is a Brownian motion with respect to the enlarged sigma field, $\mathscr{F}_t^W \vee \sigma(W(T))$. For a smooth function f,

$$\begin{aligned}
&f(V(T), W(T) - W(t)) \\
&\quad = f(V(T), W(T) - W(0)) + \int_0^t f_w(V(T), W(T) - W(u))\, d(W(T) - W(u)) \\
&\qquad + \frac{\gamma^2}{2} \int_0^t f_{ww}(V(T), W(T) - W(u))\, du \\
&\quad = f(V(T), W(T) - W(0)) + \int_0^t f_w(V(T), W(T) - W(u)) \left(-d\hat{W}(u) - \left(\frac{W(T) - W(u)}{T-u} \right) du \right) \\
&\qquad + \frac{\gamma^2}{2} \int_0^t f_{ww}(V(T), W(T) - W(u))\, du.
\end{aligned}$$

Itô's formula applying on $f(V(T), W(T) - W(t))L(t)$ gives:

$$\begin{aligned}
&f(V(T), W(T) - W(t))L(t) \\
&\quad = f(V(T), W(T) - W(0))L(0) + \int_0^t f(V(T), W(T) - W(u))\, dL(u)
\end{aligned}$$

$$
\begin{aligned}
&+\int_0^t L(u)df(V(T),W(T)-W(u))+[f(V(T),W(T)-W(\cdot)),L(\cdot)]_t \\
&=f(V(T),W(T))+\int_0^t f(V(T),W(T)-W(u))L(u-)(\lambda_B(u-)-1)d(B(u)-u) \\
&+\int_0^t f(V(T),W(T)-W(u))L(u-)(\lambda_S(u-)-1)d(S(u)-u) \\
&+\int_0^t f(V(T),W(T)-W(u))L(u)\left(\frac{V(T)}{T-u}\right)dY(u) \\
&+\int_0^t f_w(V(T),W(T)-W(u))L(u)\left(-d\hat{W}(u)-\left(\frac{W(T)-W(u)}{T-u}\right)du\right) \\
&+\frac{\gamma^2}{2}\int_0^t f_{ww}(V(T),W(T)-W(u))L(u)du.
\end{aligned}
$$

Taking averages given the observed processes under the reference measure,

$$
\begin{aligned}
&E^Q\left[f(V(T),W(T)-W(t))L(t)|\mathscr{F}_t^{B,S,Y}\right] \\
&=E^Q\left[f(V(T),W(T))|\mathscr{F}_0^{B,S,Y}\right] \\
&+\int_0^t E^Q\left[f(V(T),W(T)-W(u))L(u-)(\lambda_B(u-)-1)|\mathscr{F}_u^{B,S,Y}\right]d(B(u)-u) \\
&+\int_0^t E^Q\left[f(V(T),W(T)-W(u))L(u-)(\lambda_S(u-)-1)|\mathscr{F}_u^{B,S,Y}\right]d(S(u)-u) \\
&+\frac{1}{\sigma^2}\int_0^t E^Q\left[f(V(T),W(T)-W(u))L(u)\left(\frac{V(T)}{T-u}\right)|\mathscr{F}_u^{B,S,Y}\right]dY(u) \\
&-\int_0^t E^Q\left[L(u)f_w(V(T),W(T)-W(u))\left(\frac{W(T)-W(u)}{T-u}\right)|\mathscr{F}_u^{B,S,Y}\right]du \\
&+\frac{\gamma^2}{2}\int_0^t E^Q\left[L(u)f_{ww}(V(T),W(T)-W(u))|\mathscr{F}_u^{B,S,Y}\right]du.
\end{aligned}
$$

Note that I utilize the independence between the observed processes and the state variables, when taking averages under the reference measure. $\phi(t,f)$ is the solution of the following SPDEs:

$$
\begin{aligned}
\phi(t,f)&=\phi(0,f)+\int_0^t \phi(u-,f(\lambda_B-1))d(B(u)-u)+\int_0^t \phi(u-,f(\lambda_S-1))d(S(u)-u) \\
&+\frac{1}{\sigma^2}\int_0^t \phi(u,g_1)dY(u)-\int_0^t \phi(u,g_2)du+\frac{\gamma^2}{2}\int_0^t \phi(u,f_{ww})du,
\end{aligned}
$$

where

$$
g_1(u,v,w)=f(v,w)\frac{v}{T-u},\ \ g_2(u,v,w)=f_w(v,w)\frac{w}{T-u},
$$

and

$$\begin{aligned} P(t) &= E^P\left[\exp\left(V(T)\right)|\mathscr{F}_t^M\right] \\ &= \frac{E^Q\left[\exp\left(V(T)\right)L(t)|\mathscr{F}_t^M\right]}{E^Q\left[L(t)|\mathscr{F}_t^M\right]} \\ &= \frac{\phi(t,f_0)}{\phi(t,\mathbf{1})}, \end{aligned}$$

with

$$f_0\left(v,w\right) \equiv \exp\left(v\right).$$

□

Proof of Proposition 13.2

Proof. For $t \in [0, T-\varepsilon]$, consider

$$X_i(t) = \left(V_i(T), W_i(T) - W_i(t)\right),$$

such that $\{X_i(t) : i \in \mathbf{N}\}$ is independent and identically distributed. Think of $\{X_i(t) : 0 \le t \le T-\varepsilon, i \in \mathbf{N}\}$ as a system of particles with locations in $\mathbf{R}^2$ and time-varying weights $\{L_i(t) : 0 \le t \le T-\varepsilon, i \in \mathbf{N}\}$ and define $M(t)$ to be the weighted empirical measure of $\{X_i(t) : i \in \mathbf{N}\}$ such that

$$M(t) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} L_i(t)\delta_{X_i(t)},$$

where δ_x is the Dirac measure of x.

Let

$$\begin{aligned} L_i(t) = \exp\Big(&\int_0^t \log\left(\lambda_B(X_i(u-), M(u-))\right) dB(u) - \int_0^t (\lambda_B(X_i(u), M(u)) - 1)du \\ &+ \int_0^t \log\left(\lambda_S(X_i(u-), M(u-))\right) dS(u) - \int_0^t (\lambda_S(X_i(u), M(u)) - 1)du \\ &+ \frac{1}{\sigma^2}\int_0^t \frac{V_i(T)}{T-u} dY(u) - \frac{1}{2}\int_0^t \frac{V_i(T)^2}{(T-u)^2} du\Big), \end{aligned}$$

where

$$\begin{aligned} &\lambda_B(X_i(u), M(u)) \\ &= \alpha\left(V(t) + \frac{\sigma^2}{\sigma^2+\gamma^2}\left(V_i(T) - V(t) + W_i(T) - W_i(t)\right) + \frac{1}{2}\sigma^2(T-t)\frac{\gamma^2}{\sigma^2+\gamma^2} - \log\left(P(t)\right)\right)^+ + \eta, \end{aligned}$$

$$\lambda_S(X_i(u),M(u))$$
$$= \alpha\left(V(t)+\frac{\sigma^2}{\sigma^2+\gamma^2}\left(V_i(T)-V(t)+W_i(T)-W_i(t)\right)+\frac{1}{2}\sigma^2(T-t)\frac{\gamma^2}{\sigma^2+\gamma^2}-\log\left(P(t)\right)\right)^- + \eta.$$

Note that $P(t)$ is a function of $M(t)$, since

$$\begin{aligned} P(t) &= E^P\left[\exp\left(V(T)\right)|\mathscr{F}_t^{B,S,Y}\right] \\ &= \frac{E^Q\left[\exp\left(V(T)\right)L(t)|\mathscr{F}_t^{B,S,Y}\right]}{E^Q\left[L(t)|\mathscr{F}_u^{B,S,Y}\right]} \\ &= \frac{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \exp\left(V_i(T)\right)L_i(t)}{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n L_i(t)} \\ &= \frac{\langle f_0, M(t)\rangle}{\langle \mathbf{1}, M(t)\rangle}. \end{aligned}$$

where

$$f_0(v,w) \equiv \exp(v)$$

and

$$\langle f, M(t)\rangle \equiv \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n f\left(X_i(t)\right)L_i(t).$$

For given $\{X_i(t) : 0 \le t \le T-\varepsilon, i \in \mathbf{N}\}$, suppose that (L,M) and $(\tilde{L},\tilde{M})$ are solutions of the system. Let

$$P(t) = \frac{\langle f_0, M(t)\rangle}{\langle \mathbf{1}, M(t)\rangle} \text{ and } \tilde{P}(t) = \frac{\langle f_0, \tilde{M}(t)\rangle}{\langle \mathbf{1}, \tilde{M}(t)\rangle}.$$

Let $l_i(\cdot)$ be $\log(L_i(\cdot))$ and $\tilde{l}_i(\cdot)$ be $\log(\tilde{L}_i(\cdot))$. Then

$$\begin{aligned} &\left|l_i(t)-\tilde{l_i}(t)\right| \\ &= \Big|\int_0^t \left(\log\left(\lambda_B(X_i(u-),M(u-))\right)-\log\left(\lambda_B(X_i(u-),\tilde{M}(u-))\right)\right)dB(u) \\ &\quad+\int_0^t \left(\log\left(\lambda_S(X_i(u-),M(u-))\right)-\log\left(\lambda_S(X_i(u-),\tilde{M}(u-))\right)\right)dS(u) \\ &\quad-\int_0^t \left(\lambda_B(X_i(u),M(u))-\lambda_B(X_i(u),\tilde{M}(u))+\lambda_S(X_i(u),M(u))-\lambda_S(X_i(u),\tilde{M}(u))\right)du \\ &\quad+\frac{1}{\sigma^2}\int_0^t\left(\frac{V_i(T)}{T-u}-\frac{V_i(T)}{T-u}\right)dY(u)-\frac{1}{2}\int_0^t\left(\frac{V_i(T)^2}{(T-u)^2}-\frac{V_i(T)^2}{(T-u)^2}\right)du\Big| \end{aligned}$$

$$
\begin{aligned}
&\le \int_0^t \left|\log\left(\lambda_B(X_i(u-),M(u-))\right) - \log\left(\lambda_B(X_i(u-),\tilde{M}(u-))\right)\right| dB(u) \\
&\quad + \int_0^t \left|\log\left(\lambda_S(X_i(u-),M(u-))\right) - \log\left(\lambda_S(X_i(u-),\tilde{M}(u-))\right)\right| dS(u) \\
&\quad + \int_0^t \left(\left|\lambda_B(X_i(u),M(u)) - \lambda_B(X_i(u),\tilde{M}(u))\right| + \left|\lambda_S(X_i(u),M(u)) - \lambda_S(X_i(u),\tilde{M}(u))\right|\right) du
\end{aligned}
$$

$$
\begin{aligned}
&\left|\lambda_B(X_i(u),M(u)) - \lambda_B(X_i(u),\tilde{M}(u))\right| \\
&= \Big|\alpha\left(V(t) + \frac{\sigma^2}{\sigma^2+\gamma^2}\left(V_i(T) - V(t) + W_i(T) - W_i(t)\right) + \frac{1}{2}\sigma^2(T-t)\frac{\gamma^2}{\sigma^2+\gamma^2} - \log(P(t))\right)^+ \\
&\quad - \alpha\left(V(t) + \frac{\sigma^2}{\sigma^2+\gamma^2}\left(V_i(T) - V(t) + W_i(T) - W_i(t)\right) + \frac{1}{2}\sigma^2(T-t)\frac{\gamma^2}{\sigma^2+\gamma^2} - \log(P(t))\right)^+ \Big| \\
&\le \Big|\alpha\left(V(t) + \frac{\sigma^2}{\sigma^2+\gamma^2}\left(V_i(T) - V(t) + W_i(T) - W_i(t)\right) + \frac{1}{2}\sigma^2(T-t)\frac{\gamma^2}{\sigma^2+\gamma^2} - \log(P(t))\right) \\
&\quad - \alpha\left(V(t) + \frac{\sigma^2}{\sigma^2+\gamma^2}\left(V_i(T) - V(t) + W_i(T) - W_i(t)\right) + \frac{1}{2}\sigma^2(T-t)\frac{\gamma^2}{\sigma^2+\gamma^2} - \log(P(t))\right) \Big| \\
&= \alpha\left|P(u) - \tilde{P}(u)\right|.
\end{aligned}
$$

Note that I have used $\left|x^+ - y^+\right| \le \left|x - y\right|$. Similarly,

$$
\left|\lambda_S(X_i(u),M(u)) - \lambda_S(X_i(u),\tilde{M}(u))\right| \le \alpha\left|P(u) - \tilde{P}(u)\right|.
$$

Since both $\log\left(\lambda_B(X_i(u-),M(u-))\right)$ and $\log\left(\lambda_B(X_i(u-),\tilde{M}(u-))\right)$ are bounded below by η, I have for some $C_1 > 0$ such that

$$
\begin{aligned}
&\left|\log\left(\lambda_B(X_i(u-),M(u-))\right) - \log\left(\lambda_B(X_i(u-),\tilde{M}(u-))\right)\right| \\
&\quad \le C_1\left|\lambda_B(X_i(u),M(u)) - \lambda_B(X_i(u),\tilde{M}(u))\right| \\
&\quad \le \alpha C_1\left|P(u) - \tilde{P}(u)\right|.
\end{aligned}
$$

Similarly,

$$
\left|\log\left(\lambda_S(X_i(u),M(u))\right) - \log\left(\lambda_S(X_i(),\tilde{M}(u))\right)\right| \le \alpha C_1\left|P(u) - \tilde{P}(u)\right|.
$$

Hence,

$$
\begin{aligned}
&\left|l_i(t) - \tilde{l}_i(t)\right| \\
&\le \int_0^t \alpha C_1\left|P(u-) - \tilde{P}(u-)\right| dB(u) + \int_0^t \alpha C_1\left|P(u-) - \tilde{P}(u-)\right| dS(u) + \int_0^t \left(2\alpha\left|P(u) - \tilde{P}(u)\right|\right) du \\
&\le \int_0^t C\left|P(u-) - \tilde{P}(u-)\right| dB(u) + \int_0^t C\left|P(u-) - \tilde{P}(u-)\right| dS(u) + \int_0^t C\left|P(u) - \tilde{P}(u)\right| du,
\end{aligned}
$$

where $C \equiv \max(\alpha C_1, 2\alpha)$.

On the other hand,

$$\begin{aligned}\left|P(t)-\tilde{P}(t)\right| &= \left|\log\left(\frac{\langle f_0,M(t)\rangle}{\langle \mathbf{1},M(t)\rangle}\right)-\log\left(\frac{\langle f_0,\tilde{M}(t)\rangle}{\langle \mathbf{1},\tilde{M}(t)\rangle}\right)\right| \\ &= \left|\log\left(\frac{\langle f_0,M(t)\rangle}{\langle f_0,\tilde{M}(t)\rangle}\right)-\log\left(\frac{\langle \mathbf{1},M(t)\rangle}{\langle \mathbf{1},\tilde{M}(t)\rangle}\right)\right| \\ &\le \left|\log\left(\frac{\langle f_0,M(t)\rangle}{\langle f_0,\tilde{M}(t)\rangle}\right)\right|+\left|\log\left(\frac{\langle \mathbf{1},M(t)\rangle}{\langle \mathbf{1},\tilde{M}(t)\rangle}\right)\right|\end{aligned}$$

$$\begin{aligned}&\log\left(\frac{\langle f_0,M(t)\rangle}{\langle f_0,\tilde{M}(t)\rangle}\right) \\ &= \log\left(\frac{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\exp\left(V_i(T)\right)L_i(t)}{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\exp\left(V_i(T)\right)\tilde{L}_i(t)}\right) \\ &= \log\left(\frac{\lim_{n\to\infty}\sum_{i=1}^{n}\exp\left(V_i(T)+l_i(t)\right)}{\lim_{n\to\infty}\sum_{i=1}^{n}\exp\left(V_i(T)+\tilde{l}_i(t)\right)}\right) \\ &= \log\left(\frac{\lim_{n\to\infty}\sum_{i=1}^{n}\exp\left(V_i(T)+\tilde{l}_i(t)\right)\exp\left(l_i(t)-\tilde{l}_i(t)\right)}{\lim_{n\to\infty}\sum_{i=1}^{n}\exp\left(V_i(T)+\tilde{l}_i(t)\right)}\right) \\ &\le \log\left(\exp\left(\int_0^t C\left|P(u-)-\tilde{P}(u-)\right|dB(u)+\int_0^t C\left|P(u-)-\tilde{P}(u-)\right|dS(u)+\int_0^t C\left|P(u)-\tilde{P}(u)\right|du\right)\right) \\ &= \int_0^t C\left|P(u-)-\tilde{P}(u-)\right|dB(u)+\int_0^t C\left|P(u-)-\tilde{P}(u-)\right|dS(u)+\int_0^t C\left|P(u)-\tilde{P}(u)\right|du,\end{aligned}$$

since

$$\begin{aligned}\left|l_i(t)-\tilde{l}_i(t)\right| \le \int_0^t C\left|P(u-)-\tilde{P}(u-)\right|dB(u)+\int_0^t C\left|P(u-)-\tilde{P}(u-)\right|dS(u) \\ +\int_0^t C\left|P(u)-\tilde{P}(u)\right|du\end{aligned}$$

and

$$\frac{\lim_{n\to\infty}\sum_{i=1}^{n}\exp\left(V_i(T)+\tilde{l}_i(t)\right)\exp\left(l_i(t)-\tilde{l}_i(t)\right)}{\lim_{n\to\infty}\sum_{i=1}^{n}\exp\left(V_i(T)+\tilde{l}_i(t)\right)}$$

is a convex combination of $\exp\left(l_i(t)-\tilde{l}_i(t)\right)$ and $\log(\cdot)$ is an increasing function.

$$\begin{aligned}&-\log\left(\frac{\langle f_0,M(t)\rangle}{\langle f_0,\tilde{M}(t)\rangle}\right) \\ &\quad= \log\left(\frac{\langle f_0,\tilde{M}(t)\rangle}{\langle f_0,M(t)\rangle}\right) \\ &\quad\le \int_0^t C\left|P(u-)-\tilde{P}(u-)\right|dB(u)+\int_0^t C\left|P(u-)-\tilde{P}(u-)\right|dS(u)+\int_0^t C\left|P(u)-\tilde{P}(u)\right|du.\end{aligned}$$

Therefore,

$$\left|\log\left(\frac{\langle f_0, M(t)\rangle}{\langle f_0, \tilde{M}(t)\rangle}\right)\right| \le \int_0^t C\left|P(u-)-\tilde{P}(u-)\right|dB(u) + \int_0^t C\left|P(u-)-\tilde{P}(u-)\right|dS(u) + \int_0^t C\left|P(u)-\tilde{P}(u)\right|du.$$

Applying similar arguments gives:

$$\left|\log\left(\frac{\langle \mathbf{1}, M(t)\rangle}{\langle \mathbf{1}, \tilde{M}(t)\rangle}\right)\right| \le \int_0^t C\left|P(u-)-\tilde{P}(u-)\right|dB(u) + \int_0^t C\left|P(u-)-\tilde{P}(u-)\right|dS(u) + \int_0^t C\left|P(u)-\tilde{P}(u)\right|du.$$

Finally,

$$\left|P(t)-\tilde{P}(t)\right| \le \int_0^t 2C\left|P(u-)-\tilde{P}(u-)\right|dB(u) + \int_0^t 2C\left|P(u-)-\tilde{P}(u-)\right|dS(u) + \int_0^t 2C\left|P(u)-\tilde{P}(u)\right|du.$$

Then by Gronwall's inequality,

$$P(t) = \tilde{P}(t) \text{ a.s.}$$

Hence,

$$L_i(t) = \tilde{L}_i(t) \text{ a.s.}$$

This shows that the solution to the system (L, M), for given $\{X_i(t) : 0 \le t \le T-\varepsilon, i \in \mathbf{N}\}$, is unique. □

Proof of Proposition 13.3

Proof. Define a reference measure Q such that

$$\frac{dP}{dQ}(t) = L(t),$$

where

$$L(t,\theta) = 1 + \int_0^t L(u-,\theta)\left(\lambda_B(u-,\theta)-1\right)d\left(B(u)-u\right) + \int_0^t L(u-,\theta)\left(\lambda_S(u-,\theta)-1\right)d\left(S(u)-u\right)$$

with

$$\begin{cases} \lambda_B(t,\theta) = \alpha\,(I(t)-P(t))^+ + \eta. \\ \lambda_S(t,\theta) = \alpha\,(I(t)-P(t))^- + \eta. \end{cases}$$

Under the reference measure Q, $B(\cdot)$ and $S(\cdot)$ are independent unit Poisson processes that are independent of $W(\cdot)$. $W(\cdot)$ is a Brownian motion with a volatility parameter γ.

$$\begin{aligned} E^Q[L(t,\theta)|\mathscr{H}_t] &= 1 + \int_0^t E^Q[L(u-,\theta)\,(\lambda_B(u-,\theta)-1)\,|\mathscr{H}_u]\,d\,(B(u)-u) \\ &\quad + \int_0^t E^Q[L(u-,\theta)\,(\lambda_S(u-,\theta)-1)\,|\mathscr{H}_u]\,d\,(S(u)-u) \\ &= 1 + \int_0^t E^P[(\lambda_B(u-,\theta)-1)\,|\mathscr{H}_u]\,E^Q[L(u-,\theta)|\mathscr{H}_u]\,d\,(B(u)-u) \\ &\quad + \int_0^t E^P[(\lambda_S(u-,\theta)-1)\,|\mathscr{H}_u]\,E^Q[L(u-,\theta)|\mathscr{H}_u]\,d\,(S(u)-u). \end{aligned}$$

Hence,

$$\begin{aligned} E^Q[L(t,\theta)|\mathscr{H}_t] = \exp\big(&\int_0^t \log\left(E^P[\lambda_B(u-,\theta)|\mathscr{H}_u]\right)dB(u) - \int_0^t \left(E^P[\lambda_B(u,\theta)|\mathscr{H}_u]-1\right)du \\ &+ \int_0^t \log\left(E^P[\lambda_S(u-,\theta)|\mathscr{H}_u]\right)dS(u) - \int_0^t \left(E^P[\lambda_S(u,\theta)|\mathscr{H}_u]-1\right)du\big). \end{aligned}$$

In order to evaluate $E^P\left[\lambda_B(t,\theta)\middle|\mathscr{H}_t\right]$ and $E^P\left[\lambda_S(t,\theta)\middle|\mathscr{H}_t\right]$, we need to know the conditional distribution of $W(T)-W(t)$ given $\mathscr{H}_t$. The conditional distribution of $W(T)-W(t)$ given $\mathscr{H}_t$ is derived below.

For a smooth function f,

$$\begin{aligned} &f(W(T)-W(t)) \\ &= f(W(T)-W(0)) + \int_0^t f_w(W(T)-W(u))\,d\,(W(T)-W(u)) + \frac{\gamma^2}{2}\int_0^t f_{ww}(W(T)-W(u))\,du \\ &= f(W(T)-W(0)) + \int_0^t f_w(W(T)-W(u))\left(-d\hat{W}(u) - \left(\frac{W(T)-W(u)}{T-u}\right)du\right) \\ &\quad + \frac{\gamma^2}{2}\int_0^t f_{ww}(W(T)-W(u))\,du. \end{aligned}$$

$$\begin{aligned} &f(W(T)-W(t))\,L(t) \\ &= f(W(T)-W(0))\,L(0,\theta) + \int_0^t f(W(T)-W(u))\,dL(u,\theta) \\ &\quad + \int_0^t L(u,\theta)\,df(W(T)-W(u)) + [f(W(T)-W(\cdot)),L(\cdot,\theta)]_t \\ &= f(W(T)) + \int_0^t f(W(T)-W(u))\,L(u-,\theta)\,(\lambda_B(u-,\theta)-1)\,d\,(B(u)-u) \end{aligned}$$

$$
\begin{aligned}
&+\int_0^t f\left(W(T)-W(u)\right)L(u-,\theta)\left(\lambda_S(u-,\theta)-1\right)d\left(S(u)-u\right)\\
&+\int_0^t f_w\left(W(T)-W(u)\right)L(u,\theta)\left(-d\hat{W}(u)-\left(\frac{W(T)-W(u)}{T-u}\right)du\right)\\
&+\frac{\gamma^2}{2}\int_0^t f_{ww}\left(W(T)-W(u)\right)L(u,\theta)du.
\end{aligned}
$$

$$
\begin{aligned}
&E^Q\left[f\left(W(T)-W(t)\right)L(t,\theta)|\mathscr{H}_t\right]\\
&\quad=E^Q\left[f\left(W(T)\right)|\mathscr{H}_0\right]\\
&\qquad+\int_0^t E^Q\left[f\left(W(T)-W(u)\right)L(u-,\theta)\left(\lambda_B(u-,\theta)-1\right)|\mathscr{H}_u\right]d\left(B(u)-u\right)\\
&\qquad+\int_0^t E^Q\left[f\left(W(T)-W(u)\right)L(u-,\theta)\left(\lambda_S(u-,\theta)-1\right)|\mathscr{H}_u\right]d\left(S(u)-u\right)\\
&\qquad-\int_0^t E^Q\left[L(u,\theta)f_w\left(W(T)-W(u)\right)\left(\frac{W(T)-W(u)}{T-u}\right)|\mathscr{H}_u\right]du\\
&\qquad+\frac{\gamma^2}{2}\int_0^t E^Q\left[L(u,\theta)f_{ww}\left(W(T)-W(u)\right)|\mathscr{H}_u\right]du.
\end{aligned}
$$

$$
E^P\left[f\left(W(T)-W(t)\right)|\mathscr{H}_t\right]=\frac{E^Q\left[f\left(W(T)-W(t)\right)L(t,\theta)|\mathscr{H}_t\right]}{E^Q\left[L(t,\theta)|\mathscr{H}_t\right]}.
$$

□

References

1. Aase, K., Bjuland, T., Øksendal, B.: Strategic insider trading equilibrium: a filter theory approach. Afrika Matematika **23**, 145–162 (2012)
2. Admati, A., Pfleiderer, P.: A theory of intraday patterns: Volume and price variability. Review of Financial Studies **1**, 3–40 (1988)
3. Back, K.: Insider trading in continuous time. Review of Financial Studies **5**, 387–409 (1992)
4. Back, K., Baruch, S.: Information in securities markets: Kyle meets Glosten and Milgrom. Econometrica **72**(2), 433–465 (2004)
5. Bhushan, R.: Trading costs, liquidity, and asset holdings. Review of Financial Studies **4(2)**, 343–60 (1991)
6. Biagini, F., Hu, Y., Meyer-Brandis, T., Øksendal, B.: Insider trading equilibrium in a market with memory. Mathematics and Financial Economics **6**, 229–247 (2012)
7. Bremaud, P.: Point processes and queues, martingale dynamics. Springer series in statistics. Springer-Verlag, New York, N.Y. (1981)
8. Dufour, A., Engle, R.F.: Time and the price impact of a trade. The Journal of Finance **55**(6), 2467–2498 (2000)
9. Easley, D., Engle, R.F., O'Hara, M., Wu, L.: Time-varying arrival rates of informed and uninformed trades. Journal of Financial Econometrics, **6**, 171–207 (2008)
10. Easley, D., Kiefer, N.M., O'Hara, M., Paperman, J.B.: Liquidity, information, and infrequently traded stocks. Journal of Financial Economics **LI(4)**, 1405–1436 (1996)

11. Easley, D., O'Hara, M.: Time and the process of security price adjustment. The Journal of Finance **47**, 577–607 (1992)
12. Elliott, R.J.: An application of hidden Markov models to asset allocation problems. Finance and Stochastics **3**, 229–238 (1997)
13. Elliott, R.J., Aggoun, L., Moore, J.B.: Hidden Markov Models: Estimation and Control. Applications of Mathematics. Springer-Verlag, New York, N.Y. (1997)
14. Engle, R.F.: The econometrics of ultra high frequency data. Econometrica **68**, 1–22 (2000)
15. Follmer, H., Imkeller, P.: Anticipation canceled by a Girsanov transformation: a paradox on wiener space. Annales Inst. H. Poincar'e **29(4)**, 569–58 (1993)
16. Frey, R., Runggaldier, W.J.: Nonlinear filtering techniques for volatility estimation with a view towards high frequency data. International Journal of Theoretical and Applied Finance **4**, 199–210, (2001)
17. Glosten, L.R., Milgrom, P.R.: Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders. Journal of Financial Economics **14**, 71–100 (1985)
18. Hasbrouck, J.: Measuring the information content of stock trades. Journal of Financial Economics **46 (1)**, 179–207 (1991)
19. Hausman, J.A., Lo, A.W., MacKinlay, A.C.: An ordered probit analysis of transaction stock prices. Journal of Financial Economics **31**(2), 319–379 (1992)
20. Holden, C.W., Subrahmanyam, A.: Long-lived private information and imperfect competition. Journal of Financial Economics **47**, 247–270 (1992)
21. Kallianpur, G.: Stochastic Filtering Theory. Stochastic Filtering Theory. Springer-Verlag, New York, N.Y. (1980)
22. Kallianpur, G., Striebel, C.: Arbitrary system process with additive white noise observation errors. Ann. Math. Statist. **39**, 785–801 (1968)
23. Karatzas, I., Shreve, S.E.: Brownian motion and stochastic calculus. Graduate texts in mathematics. Springer-Verlag, New York, N.Y. (1991)
24. Kurtz, T.G., Xiong, J.: Particle representations for a class of nonlinear SPDEs. Stochastic Processes and their Applications **83**, 103–126 (1999)
25. Kyle, A.S.: Continuous auctions and insider trading. Econometrica **53**, 1315–1335 (1985)
26. Chib, S., Nardari, F., Shephard, N.: Markov chain Monte Carlo methods for stochastic volatility models. Journal of Econometrics **108**, 281–316 (2002)
27. Pikovsk, I., Karatzas, I.: Anticipative portfolio optimization. Advances in Applied Probability **28**, 1095–112 (1996)
28. Pitt, M., Shephard, N.: Filtering via simulation: Auxiliary particle filters. Journal of the American Statistical Association **94**, 590–599 (1999)
29. Zeng, Y.: A partially observed model for micromovement of asset prices with Bayes estimation via filtering. Mathematical Finance **13**, 411–444 (2003)

Chapter 14
Heterogenous Autoregressive Realized Volatility Model

Yazhen Wang and Xin Zhang

14.1 Introduction

Volatility plays a central role in modern finance. There is extensive literature on volatility estimation and forecast based on financial data. For financial data observed at daily or longer time horizons, which are often referred to as low-frequency financial data, many parametric models were developed in past three decades to model volatility processes, and the well-known volatility models include GARCH models, discrete stochastic volatility models, and diffusive stochastic volatility models (see Bollerslev et al. (1992); Drost and Nijman (1993) and Shephard (1996)). For intraday financial data, which are called high-frequency financial data, various realized volatility methods were developed for estimating integrated volatility in the past decade (see Ait-Sahalia et al. (2005); Andersen et al. (2003a,b); Barndorff-Nielsen et al. (2008); Tao et al. (2013); Wang and Zou (2010)).

For low-frequency data, due to the lack of enough data to directly estimate volatilities, the low-frequency modeling of volatilities assumes that volatility processes follow stationary models such as autoregressive (AR) processes. For example, for daily data there is only one observation in each day, and daily volatility cannot be estimated from the single observation. We often fit daily return data to a stationary model with an AR volatility process for volatility estimation and forecast. On the other hand, given a time interval we can directly estimate integrated volatility based on high-frequency data over the time interval. Because of the small time scale in high-frequency financial observations, the high-frequency models need to incorporate micro-structure noise, and the volatility processes in the high-frequency models no longer obey any known parametric models. Instead nonparametric models are used for volatility processes in the high-frequency models, and nonparametric methods such as realized volatility are used to estimate integrated volatility.

Y. Wang (✉) • X. Zhang
Department of Statistics, University of Wisconsin-Madison,
1300 University Avenue, Madison, WI, USA
e-mail: yzwang@stat.wisc.edu; zhangxin@stat.wisc.edu

Y. Zeng and S. Wu (eds.), *State-Space Models: Applications in Economics and Finance*, Statistics and Econometrics for Finance 1, DOI 10.1007/978-1-4614-7789-1_14,

The volatility processes in the low-frequency models have stationary parametric AR structures that are good for prediction, but the low-frequency models take volatilities as latent processes and thus make statistical inferences difficult. The high-frequency methods can directly estimate volatility nonparametrically, without relying on stationary assumption, but the nonparametric methods are hard for volatility prediction. We may take the advantages of the strengths of high-frequency and low-frequency methods by combining them together. One combining approach is to use the high-frequency data to estimate volatilities and then fit the estimated volatilities to an AR volatility model (Andersen et al. (2003a,b); Corsi (2009), and Tao et al. (2011)). Instead of latent volatility processes in the low-frequency models, the combining approach enables us to fit an AR volatility model directly for volatility estimation and prediction. The approach is largely based on good intuitions and nice empirical results.

This chapter is the first attempt to provide some theoretical justifications for the combining approach. As it is widely known that at large time scales of low-frequency data volatility processes have an AR structure, but at the small time scales of high-frequency data, the volatility processes do not obey any simple parametric models like the AR model. Since the volatility behaviors at larger time scales may be treated as the results of temporal aggregations of volatility activities at smaller time scales, we may speculate that the low-frequency parametric AR volatility structures are due to the temporal aggregations of high-frequency volatility models. In this paper we will show that for appropriate underlying price and volatility processes, temporal aggregations of volatility processes and their corresponding realized volatility estimators approximately obey a heterogenous AR (HAR) model. The obtained results provide some theoretic justifications for the described approach of combining low-frequency and high-frequency methods.

The rest of the chapter is as follows. Section 14.2 reviews the model framework for high-frequency financial data. Section 14.3 presents a continuous-time bivariate diffusion model for the underlying price process and describes the GARCH approximation and the stochastic volatility (SV) approximation to the underlying continuous-time price and volatility processes. Section 14.4 shows that the temporal aggregations of the GARCH volatility process, the SV volatility process, and the underlying continuous-time volatility process all approximately obey a HAR model. Section 14.5 illustrates realized volatility estimators of integrated volatilities based on the high-frequency financial data and establishes an approximate HAR model for the realized volatility estimators. We collect some technical results about the temporal aggregation of AR processes in Sect. 14.6.

14.2 High-Frequency Financial Data and Price Model

Let $X(t)$ be the true log price of an asset at time t over an interval $[0,T]$. High-frequency finance assumes that, because of micro-structure noise in high-frequency data, the observed data are not the underlying true log price process $X(t)$ in contin-

uous time. Instead we observe only the high-frequency noisy version, $Y(t_\ell)$, of $X(\cdot)$ at time points t_ℓ. In this chapter we assume

$$Y(t_\ell) = X(t_\ell) + e(t_\ell), \qquad t_\ell = T\ell/n, \ \ \ell = 1, \cdots, n, \tag{14.1}$$

where $e(t_\ell)$, $\ell = 1, \cdots, n$, represent micro-structure noise and are assumed to be i.i.d. random variables with mean zero and variance η, and $e(\cdot)$ and $X(\cdot)$ are assumed to be independent.

Modern finance theory assumes that $X(t)$ follows a continuous-time diffusion model,

$$dX(t) = \mu_t\, dt + \sigma_t\, dB_t, \qquad t \in [0, T], \tag{14.2}$$

where B_t is a standard Brownian motion, μ_t is the drift, and σ_t^2 is the volatility of $X(t)$. High-frequency data $Y(t_\ell)$ are used to estimate the integrated volatility.

14.3 GARCH and Stochastic Volatility Approximations to the Price Model

Consider the following bivariate diffusion model for the log price process $X_t, t \in [0, T]$,

$$dX_t = (\nu_0 + \nu_1\, \sigma_t^2)\, dt + \sigma_t\, dB_t, \tag{14.3}$$

$$d\sigma_t^2 = (\eta_0 + \eta_1\, \sigma_t^2)\, dt + \eta_2\, \sigma_t^2\, dW_t, \tag{14.4}$$

where B_t and W_t are two independent standard Brownian motions, σ_t^2 is the volatility process, and $(\nu_0, \nu_1, \eta_0, \eta_1, \eta_2)$ are parameters.

The bivariate diffusion process (X_t, σ_t^2) can be approximated by a GARCH process as follows. Divide the time interval $[0, T]$ into n subinterval of length $s_n = T/n$ and set $t_k = k\, s_n$, $k = 0, 1, \cdots, n$. For i.i.d. standard normal random variables $\{\varepsilon_k\}$, let

$$\zeta_k = 2^{-1/2}(\varepsilon_k^2 - 1). \tag{14.5}$$

We define a linear GARCH(1,1) approximating process as follows. For $k = 1, \cdots, n$, let

$$X_{n,k} - X_{n,k-1} = (\nu_0 + \nu_1\, \sigma_{n,k}^2)\, s_n + \sigma_{n,k}\, s_n^{1/2}\, \varepsilon_k, \tag{14.6}$$

$$\begin{aligned} \sigma_{n,k}^2 &= \eta_0\, s_n + \sigma_{n,k-1}^2\, (1 + \eta_1\, s_n + \eta_2\, s_n^{1/2}\, \zeta_{k-1}) \\ &= \alpha_0 + \alpha_1\, \sigma_{n,k-1}^2 + \alpha_2\, \sigma_{n,k-1}^2\, \varepsilon_{k-1}^2, \end{aligned} \tag{14.7}$$

where

$$\alpha_0 = \eta_0\, s_n, \qquad \alpha_1 = 1 + \eta_1\, s_n - \eta_2\, s_n^{1/2}/2^{1/2}, \qquad \alpha_2 = \eta_2\, s_n^{1/2}/2^{1/2}.$$

The approximating process $(X_{n,t}, \sigma^2_{n,t})$, $t \in [0,T]$, is given by

$$X_{n,t} = X_{n,k}, \qquad \sigma^2_{n,t} = \sigma^2_{n,k}, \qquad \text{for } t \in [t_k, t_{k+1}), \ \ k = 0, \cdots, n. \tag{14.8}$$

As $n \to \infty$, the normalized partial sum process of (ε_k, ζ_k) weakly converges to a planar Wiener process (B_t, W_t) and thus the GARCH process $(X_{n,t}, \sigma^2_{n,t})$ converges in distribution to bivariate diffusion process (X_t, σ^2_t) described by (14.3) and (14.4). The diffusion model (14.3) and (14.4) [or the process (X_t, σ^2_t)] is called the diffusion limit of the linear GARCH model (14.6) and (14.7).

We may discretize (14.3) and (14.4) to obtain a discrete stochastic volatility (SV) model,

$$X_{n,k} - X_{n,k-1} = (\nu_0 + \nu_1 \sigma^2_{n,k})\, s_n + \sigma_{n,k}\, s_n^{1/2}\, \varepsilon_k, \tag{14.9}$$

$$\begin{aligned} \sigma^2_{n,k} &= \eta_0\, s_n + \sigma^2_{n,k-1}\,(1 + \eta_1\, s_n + \eta_2\, s_n^{1/2}\, \delta_k) \\ &= \alpha_0 + \alpha_1\, \sigma^2_{n,k-1} + \alpha_2\, \sigma^2_{n,k-1}\, \delta_k, \end{aligned} \tag{14.10}$$

where $\varepsilon_k = (B_{t_k} - B_{t_{k-1}})/s_n^{1/2}$ and $\delta_k = (W_{t_k} - W_{t_{k-1}})/s_n^{1/2}$

$$\alpha_0 = \eta_0\, s_n, \qquad \alpha_1 = 1 + \eta_1\, s_n, \qquad \alpha_2 = \eta_2\, s_n^{1/2}.$$

The approximating process $(X_{n,t}, \sigma^2_{n,t})$, $t \in [0,T]$, is given by

$$X_{n,t} = X_{n,k}, \qquad \sigma^2_{n,t} = \sigma^2_{n,k}, \qquad \text{for } t \in [t_k, t_{k+1}), \ \ k = 0, \cdots, n. \tag{14.11}$$

For the SV process $(X_{n,t}, \sigma^2_{n,t})$, since ε_k and δ_k are discretizations of Brownian motions B and W, $(X_{n,t}, \sigma^2_{n,t})$ will converge in probability to bivariate diffusion process (X_t, σ^2_t) described by (14.3) and (14.4). See Nelson (1990) and Wang (2002).

Note that we abuse notations by using the same set of notations for the GARCH model (14.6)–(14.8) and for the SV model (14.9)–(14.11).

14.4 The HAR Model for Volatility Processes

For both the GARCH approximation model (14.6) and (14.7) and the SV approximation model (14.9) and (14.10), the volatility process $\sigma^2_{n,k}$ obeys a HAR(1) model (a heterogeneous autoregressive model of order 1). Consider a temporal aggregation of the volatility process

$$\bar{\sigma}^2_{m,n,j} = \frac{1}{m} \sum_{k=1+j}^{m+j} \sigma^2_{n,k},$$

where m is an integer. Propositions 14.1 and 14.2 in Sect. 14.6 shows that the aggregated volatility process $\bar{\sigma}^2_{m,n,j}$ still follows a HAR(1) model, that is,

$$\bar{\sigma}^2_{m,n,j} = \eta_0 + \eta_1 \bar{\sigma}^2_{m,n,j-1} + z^*_j, \tag{14.12}$$

where z_j^* is an innovation process specified by Propositions 14.1 and 14.2 in Sect. 14.6.

Sine the GARCH and SV volatility processes $\sigma_{n,k}^2$ converge in probability to diffusion volatility σ_t^2 given by (14.4), the aggregated SV volatility process provides an approximation to a temporal aggregation of the diffusion volatility σ_t^2

$$\bar{\sigma}_{m,t_j}^2 = \frac{1}{m}\sum_{k=1+j}^{m+j} \sigma_{t_k}^2. \tag{14.13}$$

Hence, from (14.12) we have that the temporal aggregated diffusion volatility process $\bar{\sigma}_{m,t_j}^2$ approximately follows a HAR(1) model, that is,

$$\bar{\sigma}_{m,t_j}^2 = \eta_0 + \eta_1 \bar{\sigma}_{m,t_{j-1}}^2 + z_j^* + o_P(1). \tag{14.14}$$

Define integrated volatility

$$\gamma_{m,t_j} = \frac{n}{mT}\int_{t_j}^{t_{m+j}} \sigma_t^2 dt.$$

which is the continuous-time average of σ_t^2 over interval $[t_j, t_{m+j}]$. Note that

$$\begin{aligned}
&\gamma_{m,t_j} = \frac{n}{mT}\sum_{k=1+j}^{m+j}\int_{t_{k-1}}^{t_k} \sigma_t^2 dt,\\
&|\bar{\sigma}_{m,t_j}^2 - \gamma_{m,t_j}| \le \frac{1}{m}\sum_{k=1+j}^{m+j}\left|\sigma_{t_k}^2 - \frac{n}{T}\int_{t_{k-1}}^{t_k}\sigma_t^2 dt\right|\\
&= \frac{1}{m}\sum_{k=1+j}^{m+j}\left|\sigma_{t_k}^2 - \sigma_{t_k^*}^2\right| \le \max_{|s-t|\le 1/n}|\sigma_t^2 - \sigma_s^2| \to 0,
\end{aligned}$$

where t_k^* is between t_k and t_{k-1}. Thus from (14.14) we obtain

$$\gamma_{m,t_j} = \eta_0 + \eta_1 \gamma_{m,t_{j-1}} + z_j^* + o_P(1). \tag{14.15}$$

Thus the temporal aggregated continuous-time volatility process γ_{m,t_j} approximately follows a HAR(1) model.

14.5 The HAR Model for Realized Volatilities

Various realized volatility estimators are proposed to estimate integrated volatility based on high-frequency data $Y(t_\ell)$ from model (14.1). Suppose we like to estimate integrated volatility $\int_{a_0}^{a_1}\sigma_t^2 dt$ over $[a_0, a_1]$ (a subinterval of $[0, T]$) based on high-frequency data $Y(t_\ell)$, $t_\ell \in [a_0, a_1]$. Denote by n_1 the number of observations $Y(t_\ell)$ with $t_\ell \in [a_0, a_1]$.

Given an integer K let

$$[Y,Y]_1^{(K)} = \frac{1}{K}\sum_{\ell=1}^{n_1-K}(Y(t_{\ell+K})-Y(t_\ell))^2.$$

We choose $K = cn_1^{2/3}$ for some constant c and define the two-scale realized volatility (TSRV) estimator

$$\tilde{\Gamma}_1 = [Y,Y]_1^{(K)} - \frac{1}{K}[Y,Y]_1^{(1)}. \tag{14.16}$$

TSRV estimator has a suboptimal convergence rate $n_1^{1/6}$, that is,

$$\tilde{\Gamma}_1 - \int_{a_0}^{a_1}\sigma_t^2 dt = O_P(n_1^{-1/6}).$$

See Zhang et al. (2005).

The multi-scale realized volatility (MSRV) estimator is given by

$$\hat{\Gamma}_1 = \sum_{j=1}^{M} a_j [Y,Y]_1^{(K_j)} + \xi([Y,Y]_1^{(K_1)} - [Y,Y]_1^{(K_M)}), \tag{14.17}$$

where $K_j = j+N$

$$a_j = \frac{12(j+N)(j-M/2-1/2)}{M(M^2-1)},$$

$$\xi = \frac{(M+N)(N+1)}{(n+1)(M-1)},$$

and we take M and N to be the integer part of $n_1^{1/2}$ (i.e., the largest integer not greater than $\sqrt{n_1}$). The MSRV estimator $\hat{\Gamma}_1$ has optimal convergence rate $n_1^{-1/4}$, that is,

$$\hat{\Gamma}_1 - \int_{a_0}^{a_1}\sigma_t^2 dt = O_P(n_1^{-1/4}).$$

See Zhang (2006) and Fan and Wang (2007).

We divide $[0,T]$ into n/m subintervals $[a_{j-1},a_j]$, $a_j = t_{jm}$, $j = 1,\cdots,n/m$. We apply the methods described above to the high-frequency data on the jth subinterval and construct TSRV estimator $\tilde{\Gamma}_j$ and MSRV estimator $\hat{\Gamma}_j$ for the integrated volatility $\int_{a_{j-1}}^{a_j}\sigma_t^2 dt$, $j = 1,\cdots,n/m$. Then

$$\tilde{\Gamma}_j - \int_{a_{j-1}}^{a_j}\sigma_t^2 dt = O_P(m^{-1/6}), \qquad \hat{\Gamma}_j - \int_{a_{j-1}}^{a_j}\sigma_t^2 dt = O_P(m^{-1/4}). \tag{14.18}$$

Since the continuous average of σ_t^2 over interval $[a_{j-1}, a_j]$ is

$$\frac{1}{a_j - a_{j-1}} \int_{a_{j-1}}^{a_j} \sigma_t^2 dt = \frac{n}{mT} \int_{a_{j-1}}^{a_j} \sigma_t^2 dt,$$

we define the TSRV estimator and MSRV estimator of the average volatility as follows,

$$\tilde{\gamma}_j = \frac{n}{mT} \tilde{\Gamma}_j, \qquad \hat{\gamma}_j = \frac{n}{mT} \hat{\Gamma}_j.$$

Hence, from (14.18) we conclude

$$\gamma_{m,t_j} = \frac{n}{mT} \int_{a_{j-1}}^{a_j} \sigma_t^2 dt = \frac{n}{mT} \hat{\Gamma}_j + O_P(nm^{-5/4}), \tag{14.19}$$

$$\gamma_{m,t_j} = \frac{n}{mT} \int_{a_{j-1}}^{a_j} \sigma_t^2 dt = \frac{n}{mT} \tilde{\Gamma}_j + O_P(nm^{-7/6}). \tag{14.20}$$

We choose m such that $nm^{-7/6} \to 0$ for the TSRV case and $nm^{-5/4} \to 0$ for the MSRV case. Then the above results and (14.15) imply that TSRV estimator $\tilde{\gamma}_j$ and MSRV estimator $\hat{\gamma}_j$ approximately follows a HAR(1) model, that is,

$$\tilde{\gamma}_j = \eta_0 + \eta_1 \tilde{\gamma}_{j-1} + z_j^* + o_P(1), \tag{14.21}$$

$$\hat{\gamma}_j = \eta_0 + \eta_1 \hat{\gamma}_{j-1} + z_j^* + o_P(1). \tag{14.22}$$

The established HAR(1) models in (14.21) and (14.21) for TSRV estimator $\tilde{\gamma}_j$ and MSRV estimator $\hat{\gamma}_j$ provide some justification for the combining approach of fitting the estimated realized volatilities to a HAR(1) model used in Andersen et al. (2003a,b) and Corsi (2009) (see also Tao et al. (2011)).

14.6 The Temporal Aggregation of AR Processes

This section provides some technical results on the aggregation of AR model.

Proposition 14.1. *Consider the following HAR(1) process* $\{u_k\}$,

$$u_k = \alpha_0 + \alpha_1 u_{k-1} + z_k, \tag{14.23}$$

where innovation process z_i *satisfies (i)* $Var(z_k) = \tau_k^2$ *may depend on* k *and (ii)* $Cov(z_k, z_j) = 0$ *for* $j \neq k$. *Define an aggregation,* $\{v_i\}$, *of process* u_k,

$$v_i = \frac{1}{m} \sum_{k=(i-1)m+1}^{im} u_k. \tag{14.24}$$

Then $\{v_i\}$ *follows a HAR(1) model,*

$$v_i = \alpha_0^* + \alpha_1^* v_{i-1} + z_i^*, \tag{14.25}$$

where α_0^*, α_1^* *and* z_i^* *are given by*

$$\alpha_0^* = \alpha_0 \sum_{k=0}^{m-1} \alpha_1^k, \qquad \alpha_1^* = \alpha_1^m, \qquad z_i^* = \frac{1}{m} \sum_{j=(i-1)m+1}^{im} \sum_{k=0}^{m-1} \alpha_1^k z_{j-k}. \tag{14.26}$$

Proof. It is sufficient to show that $v_2 = \alpha_0^* + \alpha_1^* v_1 + z_2^*$. First, we substitute $\{u_{k+m}\}$ successively into the expressions of $\{u_k\}$ and obtain

$$\begin{aligned} u_{m+1} &= \alpha_0 + \alpha_1 u_m + z_{m+1}, \\ &= (\alpha_0 + \alpha_0\alpha_1) + \alpha_1^2 u_{m-1} + (z_{m+1} + \alpha_1 z_m), \\ &= \alpha_0 \sum_{k=0}^{m-1} \alpha_1^k + \alpha_1^m u_1 + \sum_{k=0}^{m-1} \alpha_1^k z_{m+1-k}, \end{aligned}$$

$$u_{m+2} = \alpha_0 \sum_{k=0}^{m-1} \alpha_1^k + \alpha_1^m u_2 + \sum_{k=0}^{m-1} \alpha_1^k z_{m+2-k},$$

and

$$u_{2m} = \alpha_0 \sum_{k=0}^{m-1} \alpha_1^k + \alpha_1^m u_m + \sum_{k=0}^{m-1} \alpha_1^k z_{2m-k}.$$

Plugging these expression of $\{u_k\}$ into v_2 in (14.24) we have

$$\begin{aligned} v_2 &= \frac{1}{m} \sum_{k=m+1}^{2m} u_k \\ &= \alpha_0 \sum_{k=0}^{m-1} \alpha_1^k + \alpha_1^m \frac{1}{m} \sum_{k=1}^{m} u_k + \frac{1}{m} \sum_{j=m+1}^{2m} \sum_{k=0}^{m-1} \alpha_1^k z_{j-k} \\ &= \alpha_0 \sum_{k=0}^{m-1} \alpha_1^k + \alpha_1^m v_1 + \frac{1}{m} \sum_{j=m+1}^{2m} \sum_{k=0}^{m-1} \alpha_1^k z_{j-k} \\ &= \alpha_0^* + \alpha_1^* v_1 + z_2^*. \end{aligned}$$

The proof is completed. □

Proposition 14.2. *The innovation process* $\{z_i^*\}$ *in model (14.25) satisfies*

$$z_i^* = \frac{1}{m} \sum_{j=(i-1)m+1}^{im} \sum_{k=0}^{m-1} \alpha_1^k z_{j-k},$$

$$Var(z_i^*) = O\left(\frac{1}{m}\right), \qquad Cov(z_i^*, z_{i+1}^*) = O\left(\frac{1}{m^2}\right).$$

Proof. Proposition 14.1 assumes $Var(z_k) = \tau_k^2$, $Cov(z_k, z_j) = 0$ for $j \neq k$. It is sufficient to show that

$$Var(z_2^*) = O\left(\frac{1}{m}\right), \qquad Cov(z_2^*, z_3^*) = O\left(\frac{1}{m^2}\right).$$

Direct computations show

$$\begin{aligned}
Var(z_2^*) &= Var\left(\frac{1}{m}\sum_{j=m+1}^{2m}\sum_{k=0}^{m-1}\alpha_1^k z_{j-k}\right) \\
&= \frac{1}{m^2}\sum_{j=2}^{m}\left(\sum_{k=m+1-j}^{m-1}\alpha_1^k\right)^2 Var(z_j) + \frac{1}{m^2}\sum_{j=m+1}^{2m}\left(\sum_{k=0}^{2m-j}\alpha_1^k\right)^2 Var(z_j) \\
&= \frac{1}{m^2}\left(\sum_{j=2}^{m}\frac{\alpha_1^{2m-2j+2} + \alpha_1^{2m} - 2\alpha_1^{2m-j+1}}{(1-\alpha_1)^2}\tau_j^2\right) \\
&+ \frac{1}{m^2}\left(\sum_{j=m+1}^{2m}\frac{1 + \alpha_1^{4m-2j+2} - 2\alpha_1^{2m-j+1}}{(1-\alpha_1)^2}\tau_j^2\right) \\
&= \frac{1}{m}\frac{\bar{\tau}_{m+1,2m}^2}{(1-\alpha_1)^2} + o\left(\frac{1}{m}\right),
\end{aligned}$$

where $\bar{\tau}_{m+1,2m}^2$ is the average of τ_j^2 over $m+1 \leq j \leq 2m$, that is,

$$\bar{\tau}_{m+1,2m}^2 = \frac{1}{m}\sum_{j=m+1}^{2m}\tau_j^2.$$

Similarly, simple manipulations lead to

$$\begin{aligned}
Cov(z_2^*, z_3^*) &= Cov\left(\frac{1}{m}\sum_{j=m+1}^{2m}\left(\sum_{k=0}^{2m-j}\alpha_1^k\right)z_j, \frac{1}{m}\sum_{j=m+2}^{2m}\left(\sum_{k=2m+1-j}^{m-1}\alpha_1^k\right)z_j\right) \\
&= \frac{1}{m^2}\sum_{j=m+2}^{2m}\left(\sum_{k=0}^{2m-j}\alpha_1^k\sum_{k=2m+1-j}^{m-1}\alpha_1^k\right)Var(z_j) \\
&= \frac{1}{m^2}\sum_{j=m+2}^{2m}\left(\frac{1-\alpha_1^{2m-j+1}}{1-\alpha_1}\right)\left(\frac{\alpha_1^{2m-j+1} - \alpha_1^m}{1-\alpha_1}\right)\tau_j^2 \\
&= \frac{1}{m^2}\sum_{j=m+2}^{2m}\left(\frac{\alpha_1^{2m-j+1} - \alpha_1^{4m-2j+2} - \alpha_1^m + \alpha_1^{3m-j+1}}{(1-\alpha_1)^2}\right)\tau_j^2 \\
&= O\left(\frac{1}{m^2}\right).
\end{aligned}$$

We complete the proof. □

Acknowledgment Yazhen Wang's research was partially supported by the NSF grant DMS-105635.

References

1. Ait-Sahalia, Y., Mykland, P. A. and Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. Review of Financial Studies 18, 351-416.
2. Andersen, T. G., Bollerslev, T. and Diebold, F. X. (2003). Some like it smooth, and some like it rough: untangling continuous and jump components in measuring, modeling, and forecasting asset return volatility. Manuscript.
3. Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003). Modeling and forecasting realized volatility. Econometrica 71, 579-625.
4. Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. Econometrica 76, 1481-1536.
5. Bollerslev, T., Chou, R. Y. and Kroner, K. F. (1992). ARCH modeling in finance: A review of the theory and empirical evidence. Journal of Econometrics 52, 5-59.
6. Corsi, F. (2009). A simple long memory model of realized volatility. Journal of Financial Econometrics 7, 174-196.
7. Drost, F. C. and Nijman, T. E. (1993). Temporal Aggregation of GARCH Processes. Econometrica 61, 909-927.
8. Fan, J. and Wang, Y. (2007) Multi-scale jump and volatility analysis for high-frequency financial data. Journal of the American Statistical Association 102, 1349-1362.
9. Nelson, D. B. (1990). ARCH models as diffusion approximations. Journal of Econometrics 45, 7-38.
10. Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In Time Series Models in Econometrics, Finance, and Other Fields (D. R. Cox, D. V. Hinkley and O. E. Barndorff-Nielsen, eds.), London: Chapman & Hall, pp. 1-67.
11. Tao, M., Wang, Y., Yao, Q. and Zou, J. (2011). Large Volatility Matrix Inference via Combining Low-Frequency and High-Frequency Approaches. Journal of the American Statistical Association 106, 1025-1040.
12. Tao, M., Wang, Y. and Chen, X. (2013). Fast convergence rates in estimating large volatility matrices using high-frequency financial data. Econometric Theory 29 (FirstView), pp. 1–19. doi:10.1017/S0266466612000746
13. Wang, Y. (2002). Asymptotic nonequivalence of ARCH models and diffusions. The Annals of Statistics, 30, 754-783.
14. Wang, Y. and Zou, J. (2010). Vast volatility matrix estimation for high-frequency financial data. Annals of Statistics 38, 943-978.
15. Zhang, L., Mykland, P. A. and Ait-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. Journal of the American Statistical Association 100, 1394-1411.
16. Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. Bernoulli 12, 1019-1043.

Chapter 15
Parameter Estimation via Particle MCMC for Ultra-High Frequency Models

Cai Zhu and Jian Hui Huang

15.1 Introduction

The recent availability of high frequency data[1] has motivated a growing literature devoted to extracting information from intraday trading prices and returns. Compared with low frequency data, such as daily observations, high frequency data has two distinguishable features. On the one hand, the intraday trading occur at random trading times. On the other hand, trading prices are contaminated by market microstructure noise.

According to market microstructure theory, noise is generated from two main sources. Firstly, noise is introduced by noise traders, as discussed by Black (1986). Noise traders may enter into market driven by transitory liquidity needs or misunderstanding of information. Secondly, noise may be a reflection of trading cost or the effect of price discreteness and clustering (Harris, 1991), inventory control by dealers (Hasbrouck, 1988), and delayed price discovery (Cohen et al., 1980), among other sources.

Microstructure noise has significant effect on asset dynamic modeling and parameter estimation. For instance, Aït-Sahalia et al. (2005); Bandi and Russell (2006), among others, find out the existence of microstructure noise will generate

[1] There is a difference between high frequency data in the literature related to realized variance, which are equally spaced in time and ultra-high frequency data which are irregularly spaced. In this chapter, we use high frequency data to denotes ultra-high frequency data to keep in line with other literature. But readers should notice the difference.

C. Zhu (✉)
Department of Finance, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong
e-mail: czhuaa@ust.hk

J.H. Huang
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
e-mail: majhuang@inet.polyu.edu.hk

Y. Zeng and S. Wu (eds.), *State-Space Models: Applications in Economics and Finance*, Statistics and Econometrics for Finance 1, DOI 10.1007/978-1-4614-7789-1_15,

over-valued estimator for realized variance. More recently, Asparouhova et al. (2010) study the effects of microstructure noise on prices, showing that coefficients estimated by standard ordinary least squares regressions of security returns on security characteristics or factor loadings are biased and inconsistent if microstructure noise is considered. Moreover, Duan and Fulop (2009) show that ignoring trading noise can lead to a significant over-estimation for the firm asset volatility, thus will generate bias in credit risk modeling through structure method (Merton, 1974). Therefore, models explicitly considering microstructure noise are needed.

In market microstructure literature, most models for asset dynamics have two components: a permanent one and a transitory one. The permanent component is commonly assumed to be information-related and affected by the degree of information asymmetry. The transitory component is regarded as trade-related perturbations due to market imperfection. Some stochastic processes, such as geometric Brownian motion (GBM), are often applied to model the intrinsic value of assets, and the observed prices differ from this intrinsic value due to all kinds of noise.

For most papers related to parameters estimation involving high frequency data, the various market microstructure noises are commonly summarized by white noise. However, the actual structure of noise is rich. Zeng (2003) proposes a model explicitly considering the structure of three different kinds of market microstructure noises: discrete, clustering, and non-clustering noise, with 1/8 trading rule. The details of the model is introduced in later section. Along with his model, Zeng (2003) also develops a Bayesian filtering estimation method. The method is a recursive algorithm relying on the Markov chain approximation method to compute the approximate posterior and then the Bayes estimator. While efficient, the method is computational intensive, and once the underlying dynamics for stock changes, derivation of the algorithm according to the new dynamics is also difficult.

In this chapter, a general estimation method, namely, particle Markov Chain Monte Carlo (PMCMC), is applied to parameter estimation for a couple of Zeng's models under both 1/8 and 1/100 trading rules. This method combines particle filtering with Markov Chain Monte Carlo (MCMC) to achieve sequential parameter learning in a Bayesian way. In a nutshell, MCMC is used to propose new values for parameters in the model, and then particle filtering is used to calculate values of marginal likelihood functions in the state-space model based on those proposed parameters. Similar insights are provided by Liu and West (2001) and Storvik (2002), among others.

The idea of PMCMC method has been used in finance and economics by several scholars. Since the seminal paper by Gordon, Salmond and Smith (1993) with bootstrap filter (SIR), particle filtering is regarded as an important simulation-based estimation tool for nonlinear and non-Gaussian state-space model when likelihood values and sequential parameter learning are needed. Particle filtering-based estimation methods draw much attention in finance recently, especially in stochastic variance modeling and option pricing, because in most models, except for GARCH family, stochastic variance is treated as an unobservable factor. Johannes et al. (2009) use particle filtering to extract latent stochastic variance from stochastic volatility

models with and without jumps. Malik and Pitt (2011) adapt particle filtering within maximum likelihood method to study a basic stochastic volatility model with leverage effect. Carvalho and Lopes (2007) and Rios and Lopes (2012) use the Liu and West particle filtering framework and its extension to estimate Markov switching stochastic volatility models. Christoffersen et al. (2010) apply particle filtering to both estimate model parameters and filter latent variance factor for option pricing under several stochastic volatility models. In economics, some researchers also use particle filtering-based algorithm to estimate dynamic stochastic general equilibrium (DSGE) models, such as Fernández-Villaverde and Rubio-RamÍrez (2007) and An and Schorfheide (2007). These researchers use particle filtering to calculate likelihood function values for DSGE models, and then use either Bayesian method or numerical optimization to estimate parameters in the models. Recently, Andrieu et al. (2010) summarize this PMCMC idea as a general calculation framework and provide some theoretical foundations.

PMCMC method has several nice features. First of all, although it is based on simulation, the resampling schemes in particle filtering step of the method can efficiently reduce the variance for likelihood calculation. Secondly, approximation to likelihood functions by particle filtering is proven to be unbiased in a general auxiliary particle filtering case by Pitt et al. (2010). Thirdly, the Markov Chain Monte Carlo step of the method approximates the posterior distributions of parameters, containing more information to conduct statistical inference than just point estimates. Moreover, due to the nature of this algorithm, parallel programming is able to be used, making this method more efficient in practice. Some numeric methods that are able to enhance the algorithm efficiency are discussed. Numerical studies through simulation and real data show that PMCMC method is able to yield reasonable estimates for model parameters.

The rest of the chapter is as follows. In Sect. 15.2, Zeng (2003) model is presented. Due to the intrinsic value plus noise structure, the model can be cast into nonlinear and non-Gaussian state-space framework. Hence, in Sect. 15.3, estimation method based on particle Markov Chain Monte Carlo method (PMCMC, Andrieu et al., 2010) is introduced to address the parameter estimation problem. Then both simulation and empirical studies are conducted in Sect. 15.4. Section 15.5 makes conclusions.

15.2 The Model

Zeng's model is based on the intuition that trading price should arise from an intrinsic price process in combination with market noise from trading activities at trading times. The model consists of three parts: trading time series, micro-structure noise in observed price and the unobservable intrinsic process, which are described subsequently below.

15.2.1 Trading Times

Trading time series $\{t_i : i \geq 1\}$ are modeled as a doubly-stochastic Poisson process including a Poisson process with constant intensity.

15.2.2 Micro-Structure Noise

Three important kinds of micro-structure noise, discrete, clustering, and non-clustering noise are considered. Discrete noise, which is generated from trading mechanism, exists because intraday prices move discretely, that is, tick by tick, and the smallest tick size for trading is set by security exchanges.

Clustering noise means prices gather more on some ticks, instead of distributing evenly on all ticks. Harris (1991) documents this phenomenon, noticing that stock prices cluster on some fractions: integers more common than halves, halves more than odd quarters, etc. The reason for clustering noise to exist, as suggested by Harris, is that traders want to lower the negotiation cost. With a coarser tick size scheme, it is easier to get agreements. Hasbrouck (1999) also studies the clustering noise, and puts up a modeling idea used by Zeng (2003).

Clustering noise is generated from trading procedure. There should be other kinds of trading noise, such as fixed transaction cost, cost generated from inventory control of market makers, and cost from loss of market makers when trading with informed investors. All these noise is modeled as non-clustering noise. Non-clustering noise is important in asset price dynamic modeling, because its existence allows the prices of trades to occur within the same second to be different and the difference can be two or more ticks. Besides, it can generate observed outliers in price.

Zeng's model considers all these three kinds of noise. At trading time t_i, there is an unobservable intrinsic value $\{S(t_i)\}$ of an asset. $S(t)$ or S_t is commonly modeled by a certain stochastic process and is to be described in the next subsection. The price at trading time t_i, $Y(t_i)$, is constructed from the intrinsic value $S(t_i)$ by incorporating the above three kinds of noise.

Step 1. Incorporate price discreteness by rounding off S_t to its closest tick: Round$[S(t_i), \frac{1}{M}]$, where M is tick size. Tick size is determined by trading regulations of different market and varies with time. For example, the tick size in New York Stock Exchange (NYSE) was switched to $\$\frac{1}{16}$ from $\$\frac{1}{8}$ on June 24, 1997 and then further adjusted to \$0.01 beginning from January 29, 2001. Zeng (2003) deals with the case of $M = 8$, Spalding, Tsui and Zeng (2006) the case of $M = 16$, and this chapter the case of $M = 100$.

Step 2. Incorporate non-clustering noise by adding V_i: $Y'(t_i) = \text{Round}[S(t_i) + V_i, \frac{1}{M}]$, where V is a i.i.d. random variable and independent of the intrinsic value $S(t)$. The distribution of V should be unimodal, symmetric, and bell-shaped in order to conform to the desirable features that the trading price at a tick closer to the stock value is more likely to occur and trading prices with the same distance to

the stock value have equal probabilities. A good candidate is the doubly geometric distribution with parameter ρ and $M = 100$, whose probability mass function is given by

$$P(V = v) = \begin{cases} (1-\rho) & \text{if } v = 0, \\ \frac{1}{2}(1-\rho)\rho^{M|v|} & \text{if } v = \pm\frac{1}{M}, \pm\frac{2}{M}, \ldots. \end{cases} \tag{15.1}$$

Step 3. Incorporate clustering noise by biasing $Y'(t_i)$. The biasing function $b(\cdot)$ moves $Y'(t_i)$ to some close ticks according to certain probability defined by parameters α, η. The construction of biasing function is related to tick size and illustrated by Zeng (2003) in detail with 1/8 trading rules. Market data with a tick size of 1 cent exhibits elevated frequencies of prices that are increments of 5 cents and 10 cents. From the motivation of Harris (1991) and Hasbrouck (1999), when tick size is $\$\frac{1}{100}$, $b(\cdot)$ is constructed in the following way. If $Y'(t_i)$ is a multiple of 5 cents, then $Y(t_i)$ stays on $Y'(t_i)$ with probability one. Otherwise, move Y'_t to the nearest odd increment of 5 cents with probability α and to the nearest increment of 10 cents with probability η to obtain $Y(t_i)$. Note that $\alpha + \eta$ should be less than 1 in general. The two parameters α, η can be estimated via relative frequency methods. The details are illustrated in Zeng (2003).

The conditional likelihood function $P(Y(t_i) \mid S(t_i))$ with $M = 100$ is listed in Table 15.1, with two variables D and R defined as

$$D = M * |Y_t - \text{Round}[S_t, \frac{1}{M}]|, \tag{15.2}$$

and

$$R = \begin{cases} 0 & Y_t \text{ is an increment of 10 cents}, \\ 1 & Y_t \text{ is an increment of 5 cents}, \\ 2 & \text{otherwise}. \end{cases} \tag{15.3}$$

Table 15.1 Likelihood function for high frequency data model under 1/100 trading rule

R	D	Likelihood function
0	0	$(1-\rho)(1+\eta(\rho+\rho^2+\rho^3+\rho^4))$
0	1	$0.5(1-\rho)(\rho+\eta(2+\rho+2*\rho^2+2*\rho^3+\rho^4))$
0	2	$0.5(1-\rho)(\rho^2+\eta(2+2\rho+\rho^2+\rho^3+\rho^4+\rho^5+\rho^6))$
0	3	$0.5(1-\rho)(\rho^3+\eta(2+2\rho+\rho^2+\rho^4+\rho^5+\rho^6+\rho^7))$
0	4	$0.5(1-\rho)(\rho^4+\eta(2+\rho+\rho^2+\rho^3+\rho^5+\rho^6+\rho^7+\rho^8))$
0	≥ 5	$0.5(1-\rho)(\rho^5+\eta(2+\rho+\rho^2+\rho^3+\rho^4+\rho^6+\rho^7+\rho^8+\rho^9))\rho^{D-5}$
1	0	$(1-\rho)(1+\alpha(\rho+\rho^2+\rho^3+\rho^4))$
1	1	$0.5(1-\rho)(\rho+\alpha(2+\rho+2*\rho^2+2*\rho^3+\rho^4))$
1	2	$0.5(1-\rho)(\rho^2+\alpha(2+2\rho+\rho^2+\rho^3+\rho^4+\rho^5+\rho^6))$
1	3	$0.5(1-\rho)(\rho^3+\alpha(2+2\rho+\rho^2+\rho^4+\rho^5+\rho^6+\rho^7))$
1	4	$0.5(1-\rho)(\rho^4+\alpha(2+\rho+\rho^2+\rho^3+\rho^5+\rho^6+\rho^7+\rho^8))$
1	≥ 5	$0.5(1-\rho)(\rho^5+\alpha(2+\rho+\rho^2+\rho^3+\rho^4+\rho^6+\rho^7+\rho^8+\rho^9))\rho^{D-5}$
2	0	$(1-\rho)(1-\alpha-\eta)$
2	≥ 1	$0.5(1-\rho)(1-\alpha-\eta)\rho^D$

The corresponding R and likelihood function under 1/8 trading rule is given in (15.4) and Table 15.2:

$$R = \begin{cases} 0 & \text{if } Y_t \text{ is an integer,} \\ 1 & \text{if the fractional part of } Y_t \text{ is } \frac{1}{2}, \\ 2 & \text{if the fractional part of } Y_t \text{ is } \frac{1}{4}, \frac{3}{4}, \\ 3 & \text{if the fractional part of } Y_t \text{ is } \frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}. \end{cases} \tag{15.4}$$

Table 15.2 Likelihood function for high frequency data model under 1/8 trading rule

R	D	Likelihood function
0	0	$(1-\rho)*(1+\gamma*\rho*(1+\rho^2))$
0	1	$0.5*(1-\rho)*(\rho+\gamma*(2+2*\rho^2+\rho^4))$
0	2	$0.5*(1-\rho)*\rho*(\rho+\gamma*(2+\rho^2+\rho^4))$
0	3	$0.5*(1-\rho)*(\rho^3+\gamma*(2+\rho^2+\rho^4+\rho^6))$
0	≥ 4	$0.5*(1-\rho)*\rho^{D-3}*(\rho^3+\gamma*(1+\rho^2+\rho^4+\rho^6))$
1	0	$(1-\rho)*(1+\eta*\rho*(1+\rho^2))$
1	1	$0.5*(1-\rho)*(\rho+\eta*(2+2*\rho^2+\rho^4))$
1	2	$0.5*(1-\rho)*\rho*(\rho+\eta*(2+\rho^2+\rho^4))$
1	3	$0.5*(1-\rho)*(\rho^3+\eta*(2+\rho^2+\rho^4+\rho^6)$
1	≥ 4	$0.5*(1-\rho)*\rho^{D-3}*(\rho^3+\eta*(1+\rho^2+\rho^4+\rho^6))$
2	0	$(1-\rho)*(1+\alpha*\rho)$
2	1	$0.5*(1-\rho)(\rho+\alpha*(2+\rho^2))$
2	≥ 2	$0.5*(1-\rho)\rho^{D-1}*(\rho+\alpha*(1+\rho^2))$
3	0	$(1-\rho)*(1-\alpha-\eta-\gamma)$
3	≥ 1	$0.5*(1-\alpha-\eta-\gamma)*(1-\rho)*\rho^D$

15.2.3 Intrinsic Value Processes

In general, the intrinsic value S_t can be modeled by all kinds of stochastic process. It is natural to assume GBM, or a jump-diffusion process, or a stochastic volatility process in the model. In this chapter, GBM and Merton's model, a jump-diffusion process are adopted. GBM is first estimated to study the method in detail, and then the algorithms are extended to the jump-diffusion case.

The stochastic differential equation for GBM is as follows:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t, \tag{15.5}$$

where μ is a drift parameter, σ is a positive diffusion parameter, and W_t is a standard Brownian motion.

The stochastic differential equation for Merton's jump-diffusion process (Merton, 1976) is as follows:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t + J_t dN_t, \tag{15.6}$$

where μ, σ, and W_t are of the same meanings as those in (15.5), N_t is a Poisson process with intensity λ, and J_t is normally distributed jump size with mean μ_J and variance σ_J^2, which is independent of W_t and N_t.

At trading time t_i, the two processes yield two $S(t_i)$ and then evolve to two different $Y(t_i)$ according to either Table 15.1 or 15.2 depending on different tick size rules. The target is to estimate all parameters in both stochastic processes together with ρ in (15.1). The estimation method is introduced in next section.

15.3 Estimation Method

Commonly, estimation methods for parametric model are built upon likelihood function. However, the estimation for model parameters is more difficult and computationally expensive for Zeng's model. The main reason is that the intrinsic values are not observed directly due to the existence of market microstructure noise, which makes the model a state-space one. Another reason is that the high randomness of market microstructure noise makes the state-space model nonlinear and non-Gaussian. Because of these challenges in model estimation, an efficient estimation method should be developed.

In this section, particle Markov Chain Monte Carlo (PMCMC) method is introduced. As to be shown, this method is suitable for state-space model estimation, especially for ultra-high frequency data models with complicated noise structures. For different underlying stock dynamics, only the simulation part of this algorithm needs revision.

A general review of particle filtering is given by Doucet and Johansen (2011). Other reviews in a financial and economical setting are given by Creal (2009); Lopes and Tsay (2011). The following intuitive explanation of particle filtering is based on Doucet and Johansen (2011).

For adapting PMCMC algorithm in the setting of ultra-high frequency data models, we first address one notation issue. For particle filtering algorithm (Algorithm 1), in most cases, the input data time series is equally spaced in time. However, one characteristic of ultra-high frequency data is random trading time interval. The generalization of particle filtering algorithm for Zeng' model is straightforward. The only thing is to alter the time sequence $\{1:t\}$ in Algorithm 1 to $\{t_1, t_2, \ldots, t_n\}$, where n is the length of data, and $\{t_i\}_{i=1}^n$ is the sequence of trading times, which is commonly modeled as a Poisson process. Hence in the following introduction of the algorithms, we do not distinguish t from t_i.

15.3.1 Likelihood Calculation via Simulation

First of all, the problem of likelihood value calculation is addressed and serves as an introduction of notations. As shown in the model building section, our state-space model consists of two components: one unobserved component, which is intrinsic

asset value process $\{S_t; t \geq 1\}$; one observable component, which is trading price $\{Y_t; t \geq 1\}$. $\{S_t; t \geq 1\}$ and is a Markov process, characterized by its initial density $S_1 \sim \mu_\theta(\cdot)$ and transition probability density

$$S_{t+1} \mid (S_t = S) \sim f_\theta(\cdot \mid S), \tag{15.7}$$

where θ is the parameter vector for the model. Since $\{S_t; t \geq 1\}$ is observed indirectly through trading price $\{Y_t; t \geq 1\}$, their common marginal probability density has the form

$$Y_t \mid (S_1, \ldots, S_t = S, \ldots, S_m) \sim g_\theta(\cdot \mid S). \tag{15.8}$$

The important issues for a state-space model are filtering and parameter estimation. The filtering equations for the state-space model can be written as follows:

$$p(S_t \mid Y_{1:t-1}; \theta) = \int p(S_t \mid S_{t-1}; \theta) p(S_{t-1} \mid Y_{1:t-1}; \theta) dS_{t-1}, \tag{15.9}$$

$$p(S_t \mid Y_{1:t}; \theta) = \frac{p(Y_t \mid S_t; \theta) p(S_t \mid Y_{1:t-1}; \theta)}{p(Y_t \mid Y_{1:t-1}; \theta)}, \tag{15.10}$$

$$p(Y_t \mid Y_{1:t-1}; \theta) = \int p(Y_t \mid S_t; \theta) p(S_t \mid Y_{1:t-1}; \theta) dS_t. \tag{15.11}$$

Equations (15.9)–(15.11) enable us to filter for a given θ and evaluate marginal likelihood of observation $\{Y_{1:t}\}$. The likelihood function is given by

$$p(Y_{1:t} \mid \theta) = p(Y_1 \mid \theta) \prod_{k=2}^{t} p(Y_k \mid Y_{1:k-1}; \theta). \tag{15.12}$$

In order to obtain the marginal likelihood function value, $\{S_t; t \geq 1\}$ should be integrated out from joint likelihood function (likelihood function assuming $\{S_t; t \geq 1\}$ is observable):

$$\begin{aligned} p(Y_{1:t} \mid \theta) &= \int p_\theta(S_{1:t}, Y_{1:t}) dS_{1:t} \\ &= \int \mu_\theta(S_1) \prod_{k=2}^{t} f_\theta(S_k \mid S_{k-1}) \prod_{k=1}^{t} g_\theta(Y_k \mid S_k) dS_{1:t}. \end{aligned} \tag{15.13}$$

When both observation and state transition equations are linear and Gaussian, the likelihood function can be evaluated analytically by Kalman filtering. However, since our model is a nonlinear, non-Gaussian, and high dimensional[2] state-space model, integration in (15.13) needs to be calculated numerically via Monte Carlo method, which means that $S_{1:t}$ needs to be sampled from a suitable distribution $\pi_t(S_{1:t})$, and the likelihood value can be calculated by

$$\mathbf{E}_{\pi_t}(p(Y_{1:t} \mid \theta)) = \frac{1}{N} \sum_{i=1}^{N} p_\theta(S_{1:t}, Y_{1:t}), \tag{15.14}$$

[2] The number of dimensions is equal to the number of data points in high frequency data set.

where $\{S_{1:t}^i, i = 1, \ldots, N\}$ are N samples drawn from $\pi_t(S_{1:t})$. Sampling from $\pi_t(S_{1:t})$ links particle filtering (sequential importance sampling) with state-space models and we first present importance sampling.

15.3.2 Importance Sampling

Consider a sequence of probability distributions $\pi_{t(t\geq 1)}$ defined on a sequence of measurable spaces $(E_t, \mathscr{F}_t)_{t\geq 1}$, where $E_1 = E, \mathscr{F}_1 = \mathscr{F}$ and $E_t = E_{t-1} \times E, \mathscr{F}_t = \mathscr{F}_{t-1} \times \mathscr{F}$. Each distribution $\pi_t(dS_{1:t}) = \pi_t(S_{1:t})dS_{1:t}$ is known up to a normalizing constant Z_t, i.e.

$$\pi_t(S_{1:t}) = \frac{\gamma_t(S_{1:t})}{Z_t}, \tag{15.15}$$

$$Z_t = \int \gamma_t(S_{1:t})dS_{1:t}, \tag{15.16}$$

where the only requirement for $\gamma_t : E_t \to \mathscr{R}^+$ is that it is known pointwise. The purpose is to sample N independent random variables, $S_{1:t}^i \sim \pi_t(S_{1:t})$ for $i = 1, \ldots, N$. In the high frequency data model, $S_{1:t}$ is the path of intrinsic price process until time t, thus $\pi_t(S_{1:t})$ is a complex high-dimensional probability distribution. It is difficult to draw samples directly from such a distribution. A traditional way is to use importance sampling technique. Importance sampling serves as a fundamental Monte Carlo method and is also the basis of particle filtering algorithms. It relies on an importance density $q_t(S_{1:t})$, such that

$$\pi_t(S_{1:t}) > 0 \Rightarrow q_t(S_{1:t}) > 0. \tag{15.17}$$

In this case, (15.15) and (15.16) are rewritten as follows:

$$\pi_t(S_{1:t}) = \frac{w_t(S_{1:t})q_t(S_{1:t})}{Z_t}, \tag{15.18}$$

$$Z_t = \int w_t(S_{1:t})q_t(S_{1:t})dS_{1:t}, \tag{15.19}$$

where $w_t(S_{1:t})$ is the unnormalized weight

$$w_t(S_{1:t}) = \frac{\gamma_t(S_{1:t})}{q_t(S_{1:t})}, \tag{15.20}$$

$$W_t^i = \frac{w_t(S_{1:t}^i)}{\sum_{i=1}^{N} w_t(S_{1:t}^i)}. \tag{15.21}$$

Importance density $q_t(S_{1:t})$ is often carefully selected from some special distributions, from which it is easy to draw samples.

15.3.3 Sequential Importance Sampling: Particle Filtering

Particle filtering (sequential importance sampling) method chooses a special importance density

$$q_t(S_{1:t}) = q_{t-1}(S_{1:t-1})q_t(S_t \mid S_{1:t-1}) = q_1(S_1)\prod_{k=2}^{t} q_k(S_k \mid S_{1:k-1}). \tag{15.22}$$

In the algorithm, $S^i_{1:t}$ is called one particle at time t. To obtain a particle, first S_1 should be sampled from $\pi_1(S_1)$ and given a weight w_1 at time 1. Then based on the result of time 1, S_2 should be sampled from $\pi_2(S_{1:2})$ and given a weight w_2 at time 2 and so on. The associated unnormalized weights can be calculated recursively according to (15.20):

$$\begin{aligned} w_t(S_{1:t}) &= \frac{\gamma_t(S_{1:t})}{q_t(S_{1:t})} = \frac{\gamma_{t-1}(S_{1:t-1})}{q_{t-1}(S_{1:t-1})}\frac{\gamma_t(S_{1:t})}{\gamma_{t-1}(S_{1:t-1})q_t(S_t \mid S_{1:t-1})} \\ &= w_{t-1}(S_{1:t-1})\alpha_t(S_{1:t}) = w_1(S_1)\prod_{k=2}^{t}\alpha_k(S_{1:k}), \end{aligned} \tag{15.23}$$

where α_k is called as incremental importance weight function, which is given by

$$\alpha_k(S_{1:k}) = \frac{\gamma_k(S_{1:k})}{\gamma_{k-1}(S_{1:k-1})q_k(S_k \mid S_{1:k-1})}. \tag{15.24}$$

In the high frequency data model, let $\pi(S_{1:t}) = p(S_{1:t} \mid Y_{1:t})$, $\gamma(S_{1:t}) = p(S_{1:t}, Y_{1:t})$, $Z_t = p(Y_{1:t})$, then the only thing left is to select an importance distribution $q_t(S_t \mid S_{1:t-1})$. In practice, $q_t(S_t \mid S_{t-1}) = q(S_t \mid Y_t, S_{t-1})$. Particularly, $q_t(S_t \mid S_{1:t-1})$ is chosen as $f_\theta(S_t \mid S_{t-1})$ by Gordon et al. (1993), then $\alpha_t(S_{1:t}) = g_\theta(Y_t \mid S_t)$ in this case due to the Markov property of unobserved process S_t. Other schemes are possible, for instance, Pitt and Shephard (1999) simulate S_t from S_{t-1} using information Y_t, which is the so-called auxiliary particle filtering. The SIR algorithm is simply written as Algorithm 1.

Algorithm 1 Particle filtering method for state-space model

At time $t = 1$
  Select $q_1(S_1) = \mu(S_1)$, and sample $S^i_1, i = 1, \ldots, N$ from $q_1(S_1)$.
  Compute unnormalized weights $w_1(S^i_1) = g_\theta(Y_1 \mid S^i_1)$.
  Compute normalized weights $W^i_1 \propto w_1(S^i_1)$.
  Resample $\{S^i_1\}$ via weight $\{W^i_1\}$ to obtain N particles, denoted as $\{S^i_1\}$.
for iteration $t \geq 2$ **do**
  Sample $S^i_t \sim f_\theta(S_t \mid S^i_{t-1})$.
  Compute unnormalized weights $w_t(S_t) = g_\theta(Y_t \mid S^i_t)$.
  Compute normalized weights $W^i_t \propto w_t(S^i_{1:t})$.
  Resample $\{S^i_t\}$ via weight $\{W^i_t\}$ to obtain N particles, denoted as $\{S^i_t\}$.
end for

Algorithm 1 allows us to estimate sequentially the marginal likelihood function by

$$\hat{p}_\theta(Y_{1:t}) = \hat{p}_\theta(Y_1)\prod_{t=2}^{T}\hat{p}_\theta(Y_t \mid Y_{1:t-1}), \tag{15.25}$$

where

$$\hat{p}_\theta(Y_t \mid Y_{1:t-1}) = \frac{1}{N}\sum_{k=1}^{N} w_t(S_{1:t})^i. \tag{15.26}$$

15.3.4 Particle MCMC

In the particle filtering algorithm, when parameter vector θ is fixed, $\{S_{1:t}\}^i$ is sampled from $p(S_{1:t} \mid Y_{1:t})$. For parameters estimation in PMCMC method, θ should also be sampled from certain distribution. PMCMC can be regarded as particle filtering within MCMC, which allows us to sample from joint density $p(\theta, S_{1:t} \mid Y_{1:t})$ in each iteration by an particular MCMC algorithm. In this chapter, Metropolis–Hastings (M–H) algorithm proposed by Metropolis et al. (1953) is applied. Since Metropolis–Hastings algorithm is widely used and well known, it is not introduced in detail. Refer to Bolstad (2010) for some comprehensive introduction.

By standard decomposition, $p(\theta, S_{1:t} \mid Y_{1:t}) = p(\theta \mid Y_{1:t})p_\theta(S_{1:t} \mid Y_{1:t})$. Consequently it is natural to use a proposal density for an M–H upgrade in the form of

$$\tilde{q}(\theta', S'_{1:t} \mid \theta, S_{1:t}) = \tilde{q}(\theta' \mid \theta)p_{\theta'}(S'_{1:t} \mid Y_{1:t}). \tag{15.27}$$

The M–H acceptance rate $\tilde{\alpha}$ is given by

$$\tilde{\alpha} = \frac{p(\theta', S'_{1:t} \mid Y_{1:t})}{p(\theta, S_{1:t} \mid Y_{1:t})} = \frac{p_{\theta'}(Y_{1:t})}{p_\theta(Y_{1:t})}\frac{\tilde{q}(\theta \mid \theta')}{\tilde{q}(\theta' \mid \theta)}, \tag{15.28}$$

where $p_\theta(Y_{1:t})$ and $p_{\theta'}(Y_{1:t})$ can be calculated via particle filtering method listed in Algorithm 1. The whole algorithm is summarized in Algorithm 2.

Algorithm 2 Particle MCMC for state-space model

At time $t = 1$
  Set initial parameters θ_0.
  Run a particle filtering algorithm, obtaining $\hat{p}_{\theta_0}(Y_{1:t})$, denote as estimation of marginal likelihood value.
for iterations $i \geq 1$ **do**
  Sample $\theta' \sim \tilde{q}(\cdot \mid \theta_{i-1})$.
  Run a particle filtering algorithm, obtaining $\hat{p}_{\theta_0}(Y_{1:t})$, denote as estimation of marginal likelihood value.
  With probability $\tilde{\alpha} = \min(1, \frac{p(\theta', S'_{1:t}|Y_{1:t})}{p(\theta, S_{1:t}|Y_{1:t})} = \frac{p_{\theta'}(Y_{1:t})}{p_\theta(Y_{1:t})}\frac{\tilde{q}(\theta|\theta')}{\tilde{q}(\theta'|\theta)})$,
  set $\theta_i = \theta'$, and $\hat{p}_{\theta_i}(Y_{1:t}) = \hat{p}_{\theta'}(Y_{1:t})$.
end for

The estimation approach (Algorithm 2) relies on particle approximations to the likelihood functions. Under mild regularities on state transition function and the likelihood function, particle approximations to the likelihood functions of Algorithm 1 converge to the true values as the number of particles N increases, refer to Crisan and Doucet (2002) for a summary. Typically, particle filtering will achieve good accuracy when the number of particles N is equal or larger than the number of data points. The convergence of particle filtering also depends on properties of resampling methods. Theories for multinomial resampling, residual resampling, and systematic resampling are established. The extension to branching resampling is given by Xiong and Zeng (2011). Moreover, in Algorithm 1, the resampling procedure is conducted in every iteration; however, it is not necessary. Douc and Moulines (2008) and Del Moral et al. (2012) prove the convergence properties for algorithms in which resampling is conducted at random times, according to coefficient of variation (CV) and effective sample size (ESS) criterion.

15.4 Simulation and Empirical Studies

In this section, first of all we show that as a Monte Carlo integration for (15.13), resampling can reduce variance for likelihood value calculation, because the resampling steps eliminate those particles which have lower probabilities to generate observed values. Then through a simulation experiment, the method is tested and confirmed to be useful, and related numerical issues are discussed. Finally, the algorithm is applied to real transaction data.

15.4.1 Variance Reduction Effect of Particle Filtering Method

In Algorithm 2, the likelihood value $p_\theta(Y_{1:t})$ is also the normalized constant Z_t in (15.15). Then according to Monte Carlo integration theory, $p_\theta(Y_t|Y_{1:t-1})$ is calculated at every time t via particle filtering by (15.26).

Generally speaking, the algorithm falls into the Monte Carlo integration framework, directly using Monte Carlo integration method suffers from large variance. To make it clear, take GBM, for example. In Fig. 15.1, 100 transaction data points are simulated according to GBM dynamics. Since the variance of GBM increases with time, the top-left panel of Fig. 15.1 shows that after a short period of time, the simulated paths will deviate from the true one significantly. To maintain a given degree of accuracy, the number of paths (number of particles) needed may be large and may increase with time. However, when particle filtering with resampling scheme is applied, the paths are able to be constrained near the true one. The basic idea of resampling is to eliminate trajectories which have small normalized importance weights and to concentrate upon trajectories with larger weights. As resampling frequency increases, the simulated paths get to the true one closer, which improves the accuracy of estimation for $p_\theta(Y_{1:t})$.

Fig. 15.1 Effects of particle filtering in variance reduction

As Fig. 15.1 shows, compared to ordinary Monte Carlo method, SIR algorithm can reduce variance significantly.

15.4.2 Simulation Study: GBM Case

To simulate trading price with 1/100 tick size, when underlying intrinsic asset value movements follow a GBM process, the drift parameter μ for GBM is 4.4e−8 per second, and the diffusion parameter is 1.2e−4 per second. The parameter ρ for non-clustering noise is 0.2. Parameter α and η for clustering noise is 0.0093 and 0.0287, respectively. The trading time is assumed to follow a Poisson process with parameter 0.067, thus for 1 day, there is about 2,000 observations. These values are consistent with estimation results in the paper of Zeng (2003), therefore, are able to generate reasonable price dynamics. The 1/100 trading rule is used here. In Sect. 15.4.3, the performance of PMCMC will be compared under both 1/8 and 1/100 trading rule. The length of Markov chain for MCMC is 10,000, and the first 2,000 is ignored as burn-in.

Firstly, PMCMC with SIR particle filtering with 10,000 particles is used. The plots in the first row of Fig. 15.2 show that the Markov chain moves around the parameter space very well. The decreasing pattern of the autocorrelations coefficients of the parameters in the plots in the second row indicates that the mixing properties of Markov chain is good. Moreover, the third row of Fig. 15.2 gives the posteriors

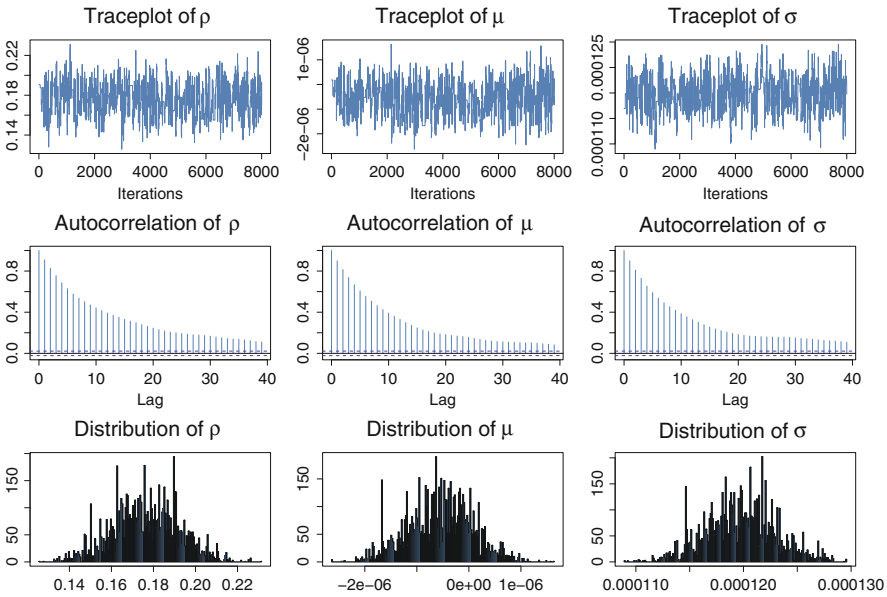

Fig. 15.2 Estimation results via SIR

Table 15.3 Estimation of parameters via SIR

	ρ	μ	σ
Mean	0.1778	−5.1434e−7	1.1981e−4
Standard deviation	0.0152	6.3941e−7	3.2527e−6

of the parameters. In Table 15.3, the estimates of ρ and σ are around the true value with small standard deviations. However, the estimate of μ is less accurate and it is not surprising because μ is a trend parameter, whose estimation accuracy depends on the range of trading time, whereas the accuracy of estimations for ρ and σ mainly depends on the number of observations. In a word, the algorithm works for the model.

However, for some applications in finance, only SIR algorithm may be not enough. One concern is that although resampling step in the algorithm can reduce likelihood value variance to certain level, the magnitude may be still too large. Large variance for likelihood values will reduce accept rate for proposals in MCMC algorithm, making the whole algorithm inefficient. There are several ways to deal with this problem. One brutal force is to increase the number of particles. This method will always work theoretically; however, in some cases, when noise to signal ratio is high, the computational burden introduced by more particles will be beyond practical.

One possible solution is to use auxiliary particle filtering (Pitt and Shephard, 1999). Note that in SIR algorithm, S_t is blindly simulated from S_{t-1}, without using

Y_t, which is known at the time. Briefly, auxiliary particle filtering simulates S_t from S_{t-1}, considering information from Y_t, for detailed illustration, refer to the original paper. The key idea is to pick up S_{t-1} which can generate S_t with larger probability to be consistent with Y_t. This algorithm may further reduce variance of likelihood calculation, as shown by Pitt (2002) and Malik and Pitt (2011) for stochastic variance model estimation. Johannes et al. (2009) and Christoffersen et al. (2010) use auxiliary particle filtering for stochastic volatility models, among others.

Another method helpful for this problem, called implicit particle filter, is illustrated by Chorin et al. (2012). The key idea of the method is still focusing on the high probability regions of the target probability density function approximated by particles. In the algorithm, a transform function F is defined (Eq. (4) in the origin paper), then finding the high probability particles is equivalent to finding region around the global minimum of F, which can be done by standard optimization procedure.

Apart from this variance issue, another issue is that the likelihood calculated by SIR method is not continuous in the unknown parameters. This problem is addressed in Pitt (2002) and Malik and Pitt (2011). The reason is that if $S_{t-1}^i (i = 1, \ldots, N)$ drawn from the filtering density $p(S_{t-1} \mid Y_{1:t-1})$ slightly change, then the proposal sample for $S_t^i (i = 1, \ldots, N)$ will also change slightly; however, the discrete probabilities associated with these proposals will change as well. The implication is that even if we use the same random numbers in each time step, the resampled particle will not be close, then the likelihood function is not continuous with parameters.

This feature may cause problem for numerical optimization procedure based on gradient, if maximum likelihood method is applied. Moreover, usual optimal random walk methods may not perform as well as expected because probability of acceptance does not tend to 1 as a proposed move becomes more local moves or even if parameter does not change at all. To address this issue, Pitt (2002) and Malik and Pitt (2011) propose a continuous approximation resampling method (CSIR). Generally speaking, the resampling method is a stratified bootstrap method. Incorporating this method in particle MCMC method may increase accepting rate and improve mixing property of MCMC chains.

Table 15.4 Estimation of parameters via CSIR

	ρ	μ	σ
Mean	0.1794	−8.004831e−7	1.2014e−4
Standard deviation	0.0161	6.046625e−7	3.603310e−6

Since this continuous resampling is well accepted for financial applications, it is also used for our model with simulated data as a comparison. The result shows that CSIR can help reduce the number of particles needed. Table 15.4 and Fig. 15.3 show that CSIR with 2,000 particles can generate similar and reasonable estimation results for our high frequency data model, in comparison with those calculated by SIR algorithm with 10,000 particles.

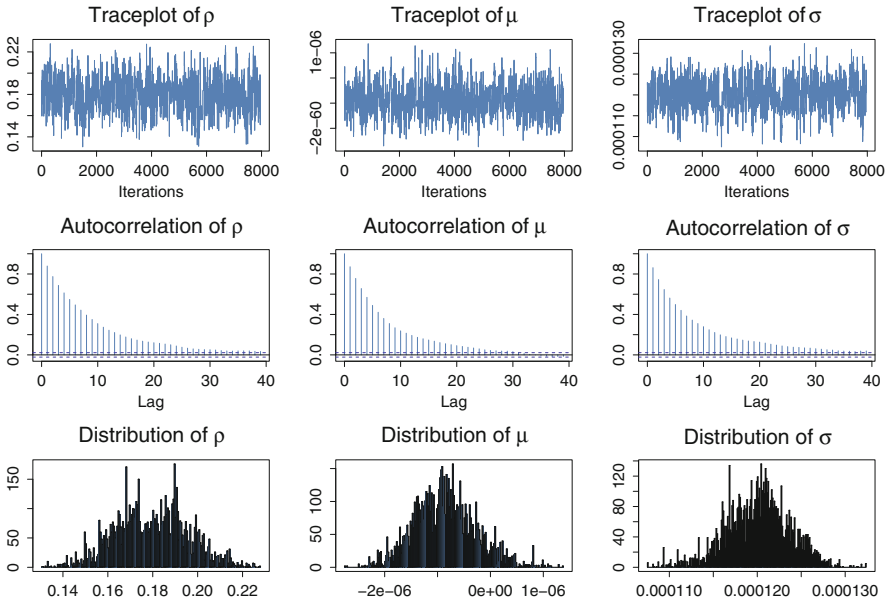

Fig. 15.3 Estimation results via CSIR

15.4.3 Comparison Algorithm Under Trading Rules with 1/8 and 1/100 Tick Size

In market microstructure literature, the effects of tick size on bid-ask spread, price discovery and transaction costs are widely discussed. Zeng (2003) proposes a modeling method for microstructure noise based on 1/8 tick size, which can be compared with 1/100 tick size in the simulation study.

In this section, the same underlying asset price S_t is simulated, with the same value of μ, σ, and ρ. Then the trading prices Y_t are generated via these two trading rules with different minimal tick sizes. The parameter estimation results are listed in Table 15.5. The properties of Markov chains for those parameters are similar with Fig. 15.3. Note that only the likelihood function needs to be modified for 1/8 tick size model, which is an advantage of PMCMC algorithm. For detailed numerical study of 1/8 tick size model via PMCMC method, refer to Zhu (2011).

Table 15.5 Estimation of parameters via SIR for model with 1/8 tick size

	ρ	μ	σ
Mean	0.2012	−5.1434e−7	1.22e−04
Standard deviation	0.0103	3.693e−07	4.915e−06

One interesting question is whether smaller tick size scheme is better for recovering true underlying process from trading noise.

Table 15.6 Summary of log-likelihood value calculated using Algorithm 1

	Mean	Standard deviation
1/8 tick size	−2,228.48	0.9176
1/100 tick size	−3,632.61	2.2457

Particle filtering algorithm with the true parameters is applied and likelihood values for different tick size model are repeatedly calculated 200 times. The simulated trading price has 2,000 data points, hence 2,000 particles are used. As Table 15.6 illustrates, the variance of likelihood values under 1/100 is larger, and mean is smaller.

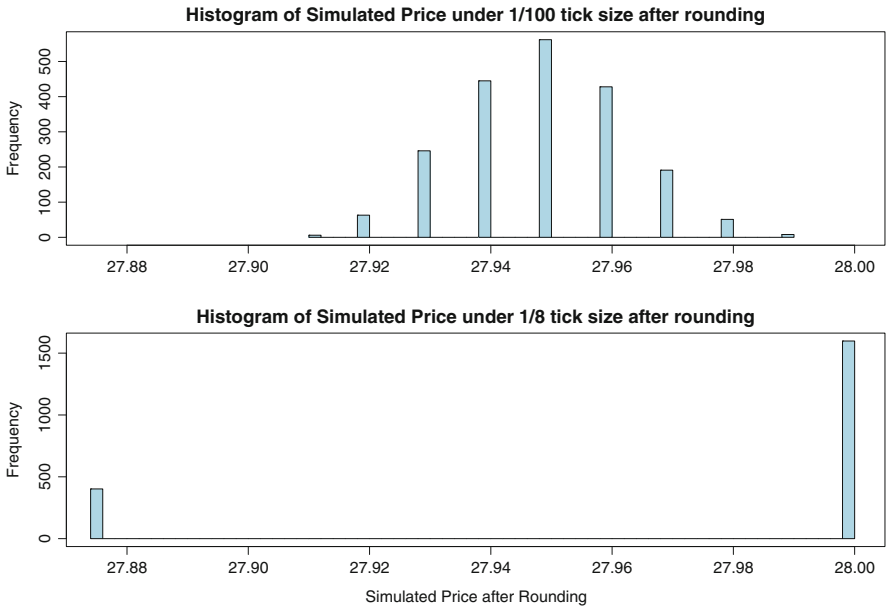

Fig. 15.4 Example: Distribution of particles (prices) in 60th step after rounding

The reason lies in Fig. 15.4. The figure shows distribution of particles after rounding in 60th out of 2,000 steps in the particle filtering algorithm. The simulated intrinsic asset value is 27.9485. The generated observed transaction price is 27.95 and 28.00 according to 1/100 tick size and 1/8 tick size rules, respectively. The rounding procedure (discreteness of trading price) in the model tends to eliminate the diversity of particles. The concentration effect is much stronger under 1/8 trading rule. In the model with 1/8 tick size, although different particles are generated for

likelihood calculation, the strong discreteness in trading price will eliminate such difference and generate similar likelihood values. Therefore, the variance is smaller. Besides, the strong discreteness will distort the simulated price away from intrinsic value and more close to trading price. Hence the mean of likelihood is larger. The simulation shows that 1/8 model should be less sensitive to parameters as long as they are kept in a reasonable range, and the trading rule will distort the underlying value a lot. Figure 15.5 confirm this argument by showing filtered intrinsic price path via true parameters is closer to the true one under 1/100 tick size rule. The total square error between filtered path and true path is only 0.0765 under 1/100 tick size rule, compared with 2.9742 under 1/8 tick size rule. The same problem of trading noise is studied under particle filtering framework by Duan and Fulop (2009). The authors show that even with daily data, the trading noise is still so significant and has an innegligible effect on credit risk model estimation.

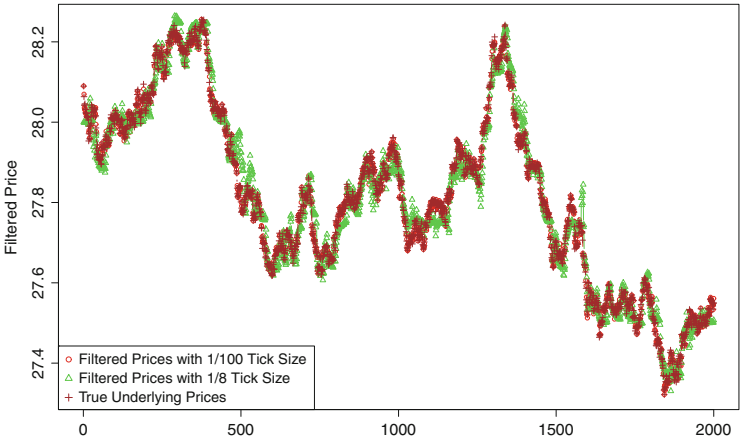

Fig. 15.5 True path and filtered paths for different trading rules

However, note that the large variance for likelihood value calculation under 1/100 tick size makes it more difficult to estimate true parameters with MCMC, because accept rate of proposed parameters will decrease as the variance increase. Therefore, the variance reduction via resampling for particle filtering and some extensions, such as continuous approximation resampling method, which are demonstrated in Sects. 15.4.1 and 15.4.2, are important in practice.

15.4.4 Simulation Study: Jump-Diffusion Case

In studies based on low frequency data, such as daily data, jumps are assumed to have relatively large jump size and small jump intensity, which is caused by macroeconomic information and financial statement of particular companies. When daily

returns are used, one well-accepted assumption for jump-diffusion process is that there should be at most one jump in a given day. Therefore, $\Delta N_t = N_t - N_{t-1}$ follows a Bernoulli distribution. However, there are studies indicating that there could be some small jumps in 1 day, and these small jumps may give rise to an appearance of infrequent large jumps if one only uses low frequency data such as daily, weekly, and monthly data. According to empirical study of Duan and Fulop (2007) concerning jumps in high frequency data, estimations of jump intensity and jump size depend on sampling frequency. As one increases the sampling frequency from once every hour to once every 10 min, the estimated mean number of jumps in price per trading session rises. Since tick-by-tick data are used in the simulation study, it is reasonable to assume jumps have relatively small size and arrive more frequently.

In this section, for simulation experiment, parameters except for jumps are set to be the same as Sect. 15.4.2: $\rho = 0.2$, $\alpha = 0.225$, $\eta = 0.066$, $\gamma = 0.3$, $\mu = 4.4\text{e}^{-8}$ per second, $\sigma = 1.2\text{e}^{-4}$ per second. The minimal tick size is 1/8. One day's data is generated with three different sets for jump parameters $(\mu_J,\ \sigma_J,\ \lambda)$ in seconds: $\Theta_1 = (4.4e-5,\ 1.2e-5,\ 0.01)$, $\Theta_2 = (4.4e-3,\ 1.2e-3,\ 0.0001)$ and $\Theta_3 = (4.4e-3,\ 1.2e-3,\ 0.001)$. The particle number is 2000, and the length of Markov chain is 45,000. The last 35,000 data is used for analysis.

Table 15.7 Estimation results for jump-diffusion process with different parameter sets

Panel A. Estimation results for parameter set Θ_1						
	ρ	μ	σ	μ_J	σ_J	λ
True value	0.200	4.400e−8	1.200e−4	4.400e−5	1.200e−5	0.01
Estimated mean	1.877e−1	3.399e−7	1.160e−4	4.180e−5	1.092e−5	1.084e−2
Estimated error	4.380e−5	2.037e−7	5.925e−6	5.056e−6	2.879e−6	4.633e−3
Panel B. Estimation results for parameter set Θ_2						
	ρ	μ	σ	μ_J	σ_J	λ
True value	0.200	4.400e−8	1.200e−4	4.400e−3	1.200e−3	0.0001
Estimated mean	1.937e−1	3.564e−7	1.185e−4	4.324e−3	1.154e−3	1.158e−4
Estimated error	8.484e−3	2.080e−7	6.171e−6	4.842e−4	2.803e−4	4.217e−5
Panel C. Estimation results for parameter set Θ_3						
	ρ	μ	σ	μ_J	σ_J	λ
True value	0.200	4.400e−8	1.200e−4	4.400e−3	1.200e−3	0.001
Estimated mean	1.856e−1	3.980e−7	1.192e−4	4.452e−3	1.180e−3	1.103e−3
Estimated error	8.893e−3	2.456e−7	6.284e−6	4.082e−4	2.582e−4	2.251e−4

Table 15.7 lists estimates for different parameter sets. Θ_1 means smaller but more frequent jumps. Θ_3 means larger but less frequent jumps. Θ_2 denotes the middle case. Results show that PMCMC method can get reasonable estimates of parameters for jump-diffusion process with all three sets of parameters. Figure 15.6 shows the similar pattern with Fig. 15.2. The plots in the first row show that the Markov chain moves around the parameter space well. The decreasing pattern of the autocorrelations coefficients of the parameters in the plots in the second row indicates that the mixing properties of Markov chain is good. Moreover, the third row gives the posteriors of the parameters. All the estimates except for μ are around the true value with small standard deviations.

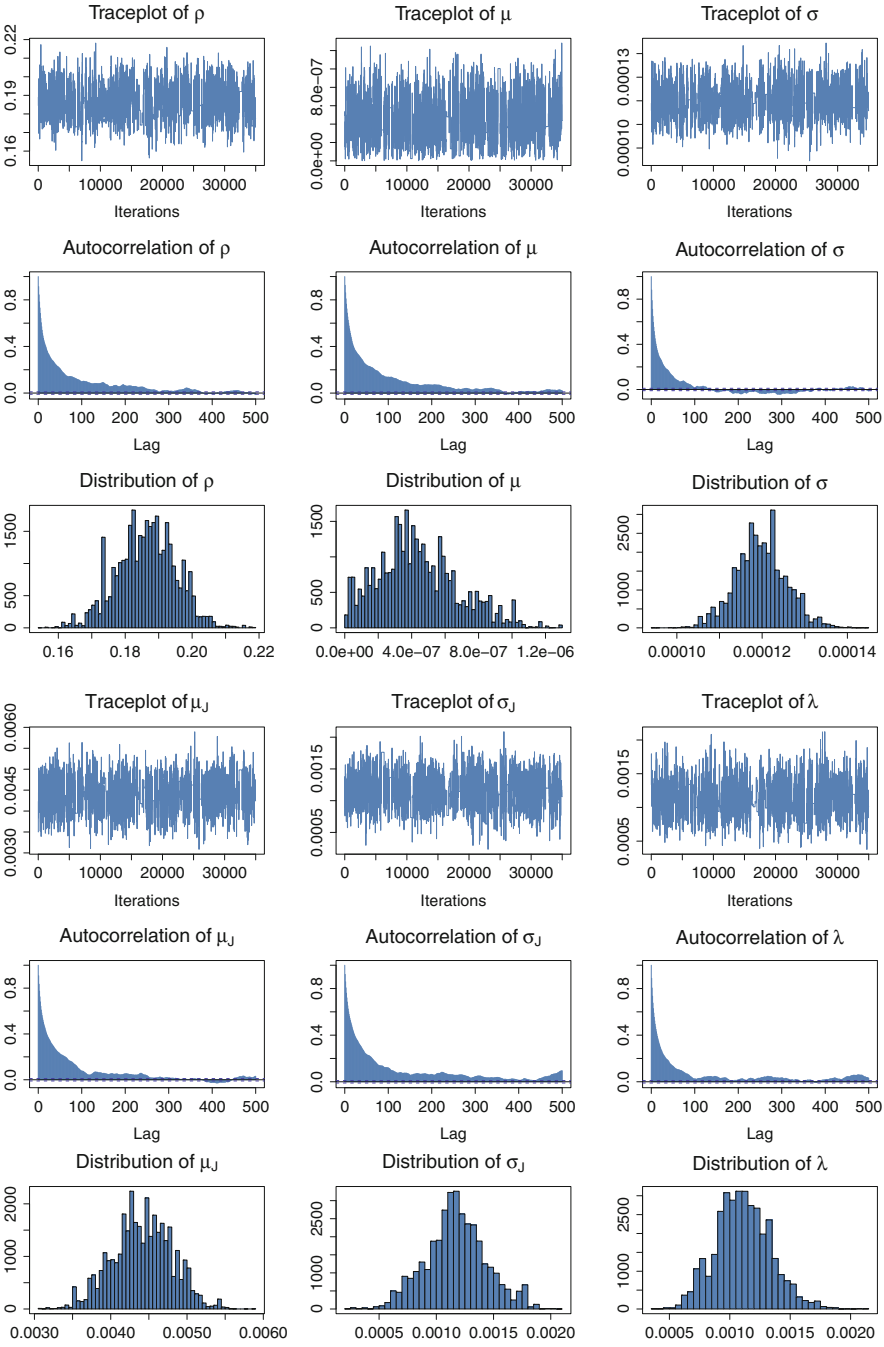

Fig. 15.6 Estimation results for jump-diffusion process via SIR

15.4.5 Real Data Application

This section applies the mode and the algorithm to a real data set. The data set consists of all transaction price of MSFT (MicroSoft) from NYSE, NASDAQ, and AMEX on January 3, 2011. The dynamics of intraday price is very different from daily price dynamics which we are familiar with. As is shown in Fig. 15.7, the price discreteness is obvious: the path is not so "smooth" as daily price dynamics, and the jump size is in unit of minimal tick size, which is 0.01.

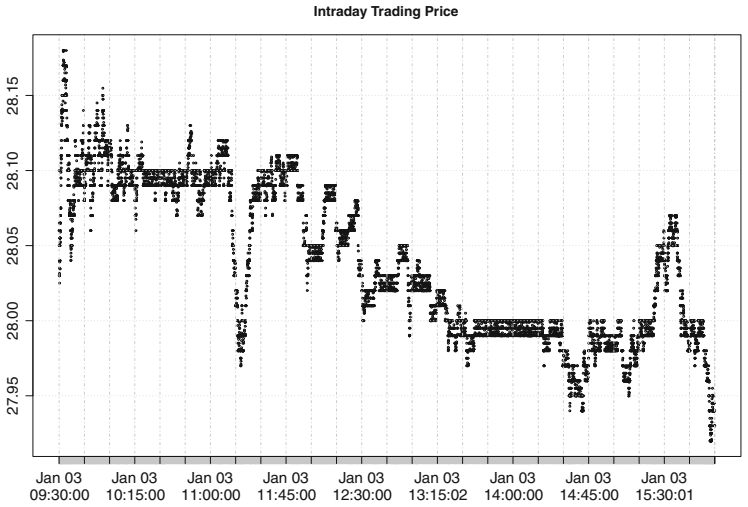

Fig. 15.7 Intraday trading price: MSFT, January, 03, 2011

For parameter estimation, all 12,093 observations are used, and particle number is 13,000. The underlying asset value dynamics is assumed to follow a GBM. CSIR algorithm is used to estimate model parameters from real data, for CSIR is shown to be more efficient than SIR method. The estimation results are reasonable, shown in Table 15.8. The accept rate of MCMC proposal is around 30%. Figure 15.8 demonstrates similar patterns for estimates, autocorrelation coefficients and posteriors with those in simulation studies.

Table 15.8 Estimation of parameters via CSIR

	ρ	μ	σ
Mean	0.0678	−2.5551e−07	7.9339e−05
Standard deviation	0.0034	5.4830e−07	1.2250e−06

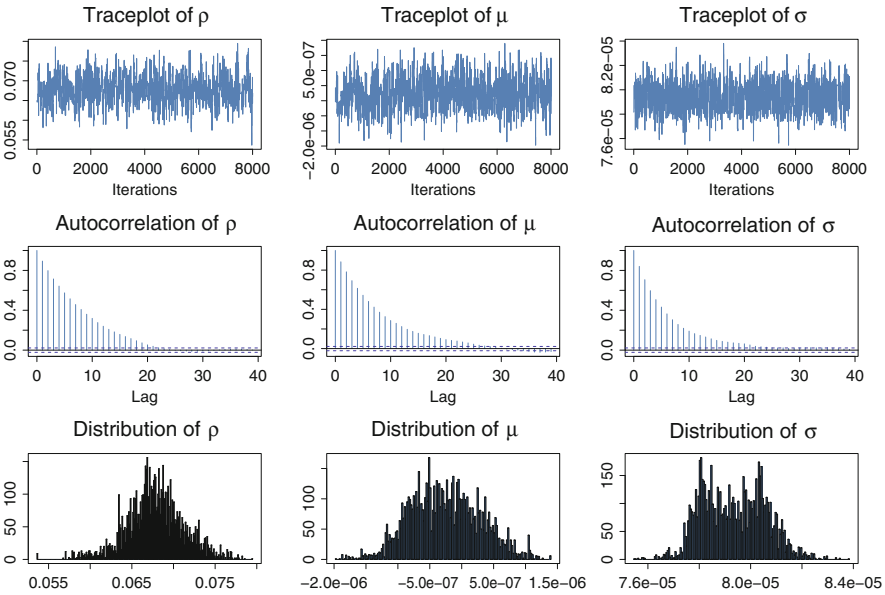

Fig. 15.8 Real data estimation via CSIR: MSFT, January, 03, 2011

15.5 Conclusion

One important issue in ultra-high frequency data study is how to model microstructure noise. Although summarizing all kinds of noise in a Gaussian random variable will make the estimation procedure much easier, this method will neglect many stylized facts from trading process. Zeng (2003) model is able to capture various kinds of noise into three categories, which are related to price discreteness, price clustering, and non-clustering noise. When considering noise explicitly, the model is a state-space model, making it difficult to estimate parameters using likelihood-based methods. To address this issue, particle Markov Chain Monte Carlo algorithm is applied to estimate the model. Moreover, methods that can enhance efficiency of this algorithm are discussed through numerical studies, such as the continuous approximation resampling method (CSIR) proposed by Pitt (2002) and Malik and Pitt (2011). Results show that PMCMC method generates reasonable estimates for model parameters when the underlying asset value follows GBM or jump-diffusion process. It is interesting to further extend PMCMC method for more complicated underlying dynamics, for instance, stochastic volatility models with leverage effect.

Acknowledgments We thank Professor Yong Zeng and one anonymous referee for numerous suggestions and insights; thank Louis Schenck and Taisuke Nakata for the R-codes to produce Fig. 15.1, and Eric Goldlust for providing the formulae for Table 15.1. The second author acknowledges the support from Hong Kong RGC Earmaked grant 500909.

References

1. Aït-Sahalia Y., Mykland P. A. and Zhang L. How often to sample a continuous-time process in the present of market microstructure noise. *Review of Financial Studies*, 18: 351-416. 2005.
2. An S. and Schorfheide F. Bayesian analysis of DSGE models. *Econometric Reviews*, 26: 113-172. 2007.
3. Andrieu C., Doucet A. and Holenstein R. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B*, 72: 1-33. 2010.
4. Asparouhova E. N., Bessembinder H. and Kalchevab I. Liquidity biases in asset pricing tests. *Journal of Financial Economics*, 96: 215-237. 2010.
5. Bandi F. M. and Russell J. R. Separating microstructure noise from volatility. *Journal of Financial Economics*, 79: 655-692. 2006.
6. Black F. Noise. *Journal of Finance*, 41: 529-543. 1986.
7. Bolstad W. M. *Understanding computational Bayesian Statistics*. New York: Wiley. 2010.
8. Carvalho C. M. and Lopes H. F. Simulation-based sequential analysis of Markov switching stochastic volatility models. *Computational Statistics & Data Analysis*, 51: 4526-4542. 2007.
9. Chorin A. J., Morzfeld M. and Tu X. M. A survey of implicit particle filters for data assimilation. this volume, 63–88. 2013.
10. Christoffersen P., Jacobs K. and Mimouni K. Volatility dynamics for the S&P500: evidence from realized volatility, daily returns, and option prices. *Review of Financial Studies*, 23: 3141-3189. 2010.
11. Cohen K., Hawawini G., Maier S., Schwartz R. and Whitcomb D. Implications of microstructure theory for empirical research on stock price behavior. *Journal of Finance*, 35: 249-257. 1980.
12. Creal D. A survey of sequential Monte Carlo methods for economics and finance. Working Paper. 2009.
13. Crisan D. and Doucet A. A Survey of convergence results on particle filtering methods for practitioners. *IEEE Transactions on Signal Processing*, 50: 736-746. 2002.
14. Del Moral P., Doucet A. and Jasra A. On adaptive resampling procedures for sequential Monte Carlo methods. *Bernoulli*, 18: 252-278. 2012.
15. Duan, J.C. and Fulop, A. How Frequently Does the Stock Price Jump? - An Analysis of High-Frequency Data with Microstructure Noises. Working Paper. (2007)
16. Duan J. C. and Fulop A. Estimating the structural credit risk model when equity prices are contaminated by trading noises. *Journal of Econometrics*, 150: 288-296. 2009.
17. Douc R. and Moulines E. Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *The Annals of Statistics*, 36: 2344-2376. 2008.
18. Doucet A. and Johansen A. M. A tutorial on particle filtering and smoothing: Fifteen years later. In Crisan D. and Rozovsky B., editors, *Handbook of Nonlinear Filtering*. Oxford University Press, 2011.
19. Fernández-Villaverde J. and Rubio-RamÍrez J. F. Estimating macroeconomic models: A likelihood approach. *Review of Economic Studies*, 74: 1059-1087. 2007.
20. Gordon N., Salmond D. and Smith A. F. M. Novel approach to nonlinear/nongaussian Bayesian state estimation. *IEEE: Radar and Signal Processing*, 140: 107-113. 1993.
21. Harris L. Stock price clustering and discreteness. *Review of Financial Studies*, 4: 389-415. 1991.
22. Hasbrouck J. Trades, quotes, inventories and information. *Journal of Financial Economics*, 42: 229-252. 1988.
23. Hasbrouck J. Security bid / ask dynamics with discreteness and clustering: Simple strategies for modeling and estimation. *Journal of Financial Markets*, 2: 1-28. 1999.
24. Johannes M., Polson N. and Stroud J. Optimal filtering of jump-diffusions: Extracting latent states from asset prices. *Review of Financial Studies*, 22: 2759-2799. 2009.
25. Liu J. and West M. Combined parameter and state estimation in simulation-based filtering. In Doucet A., De Freitas J. F. G. and Gordon N. J., editors, *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, 2001.

26. Lopes H. F. and Tsay R. S. Particle filters and Bayesian inference in financial econometrics. *Journal of Forecasting*, 30: 169-209. 2011.
27. Malik S. and Pitt M. K. Particle filters for continuous likelihood evaluation and maximisation. *Journal of Econometrics*, 165: 190-209. 2011.
28. Merton R. C. On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, 29: 449-470. 1974.
29. Merton R. C. Option Pricing when Underlying Stock Returns are Discontinuous. *Journal of Financial Economics*, 3: 125-144. 1976.
30. Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., Teller E. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21: 1087-1092. 1953.
31. Pitt M. K. Smooth particle filters for Likelihood Evaluation and Maximization. Working Paper. 2002.
32. Pitt M. K. and Shephard N. Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94: 590-599. 1999.
33. Pitt M. K., Silva R. S., Giordani P. and Kohn R. Auxiliary particle filtering within adaptive Metropolis-Hastings sampling. Working Paper. 2010.
34. Rios M. P. and Lopes H. F. The extended Liu and West filter: Parameter learning in Markov switching stochastic volatility models, Chapter 2, in this volume.
35. Spalding R., Tsui Kam-Wah and Zeng Y. A Micromovement Model with Bayes Estimation via Filtering: Application to Measuring Trading Noises and Trading Cost. *Nonlinear Analysis: Theory, Methods and Applications*, 64: 295-309. 2006.
36. Storvik G. Particle filters in state space models with the presence of unknown static parameters. *IEEE: Transactions of Signal Processing*, 50: 281-289. 2002.
37. Xiong J. and Zeng Y. A branching particle approximation to the filtering problem with counting process observations. *Statistical Inference for Stochastic Processes*, 14: 111-140. 2011.
38. Zeng Y. A partially observed model for micromovement of asset prices with Bayes estimation via filtering. *Mathematical Finance*, 13: 411-444. 2003.
39. Zhu C. Parameters estimation for jump-diffusion process based on low and high frequency data. Master Thesis. The Hong Kong Polytechnic University. 2011.

Index

Y. Zeng and S. Wu (eds.), *State-Space Models: Applications in Economics and Finance*, Statistics and Econometrics for Finance 1, DOI 10.1007/978-1-4614-7789-1,