PARAMETERS ESTIMATION FOR JUMP-DIFFUSION PROCESS BASED ON LOW AND HIGH FREQUENCY DATA

ZHU CAI

M.Phil

The Hong Kong

Polytechnic University

2011

The Hong Kong Polytechnic University

Department of Applied Mathematics

Parameters Estimation for Jump-Diffusion Process Based on Low and High Frequency Data

ZHU Cai

A thesis submitted in partial fulfillment of

the requirements for the degree of Master of Philosophy

June 2011

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ (Signed)

ZHU Cai (Name of student)

Abstract

In this thesis, we develop some parameter estimation methods of jump-diffusion process. The originality of the thesis lies in the fact that the developed estimation methods are different from those commonly-used approaches.

This thesis consists of two parts. In the first part, estimation method for continuous state branching process with immigration (hereafter, CBI) is proposed, which is based on the weighted conditional least square estimators (WCLSE). It is remarkable that the Cox-Ingersoll-Ross model with jumps (JCIR) in the studies of interest rate is a simplified version of our CBI process. Our developed method provides new perspective in parameter estimation for JCIR model. The strength of the method is that it avoids computationally expensive numerical integration which is used in many extant estimation methods.

In the second part, particle Markov chain Monte Carlo method is applied to the estimation of a parametric model for ultra-high frequency stock price data, whereas most existing studies mainly focus on nonparametric estimation methods. Our method has two special features: on the one hand, it can estimate all parameters in the jump-diffusion model whereas nonparametric methods can only provide volatility estimation; on the other hand, it can effectively estimate volatility generated by diffusion component under the influence of jumps with market microstructure noise.

Detailed simulation studies are implemented for both developed methods to evaluate their estimation performance. Results indicate that both these methods lead to reasonable estimations for parameters in the models.

Key Words jump-diffusion process, continuous state branching process with immigration, weighted conditional least square estimator, high frequency data, particle Markov chain Monte Carlo.

Acknowledgements

First and foremost, I am extremely grateful to my chief supervisor Dr. Jianhui, Huang. His enthusiasm and sound suggestions help me greatly. This thesis would not have been possible without him. I am honored to call him my teacher and friend.

I also owe my most sincere gratitude to my co-supervisor Prof. Xiaojun Chen. Although she is very busy, her office is always open for me. Her support, encouragement and inspiration are also sources of my strength during these two years.

Many thanks to the numerous people who have been my teachers within these two years. I would like to express my gratitude especially to: Prof. Yong Zeng and Dr. Chunhua Ma. I am very grateful for their generous support, advice and help.

Furthermore, I would like to thank my research student colleagues and friends for academic discussions and for providing an interesting and stimulating learning environment. The computer technicians of our department deserve my gratitude for providing assistance and administrative support when I use department's Linux cluster.

Last but not least, I am forever grateful to my parents for their unconditioned love and support.

Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

Due to the arrivals of unanticipated news and possible liquidity shocks, financial markets sometimes generate discontinuities, which can be termed as "Jumps" in academic literature. A large number of recent empirical and theoretical studies reveal the existence of jumps and their significant impact on financial modeling, from risk management to derivatives pricing and hedging. See, for example, Merton (1976), Bates (1996), Duffie, Pan and Singleton (2000), Das (2002) and Johannes (2004).

There are several reasons to extend basic models based on geometric Brownian motion to those based on jump-diffusion process. First of all, although jumps do not come to markets regularly, their arrivals tend to depend on market information. Due to this reason, the analysis of jumps will enable us to learn more about the influence of macro-economical information and corporate financial conditions on asset price.

Secondly, previous structural default models for corporate bonds pricing, see Black and Scholes (1973), Merton (1974), and Black and Cox (1976), rely on geometric Brownian motion to model firm value processes. However, it is notable that when such models are applied, the credit spreads vanish for bonds with a short time to maturity, as discussed by Jarrow and Protter (2004). This phenomenon contradicts

the empirical observation of credit spreads which have positive limits at the short end of the term structure. To explain this inconsistency, there are papers which suggest to model firm value process as some jump-diffusion processes, see Zhou (2001) for example.

Thirdly, reasonable specification of interest rate process is essential for accurate valuation of fixed income derivatives. In prior literature, many researchers, such as Vasicek (1977), Cox et al. (1985), Chan et. al. (1992), Ahn and Gao (1999), model interest rate by continuous-path stochastic process. However, there are strong empirical evidence indicating that interest rate should evolve as a jump-diffusion process. Relevant studies can be found in Das (2002), Johannes (2004) and Jarrow et al. (2007).

Moreover, the introduction of jumps may explain the excess kurtosis, skewness of stock return distributions and implied volatility smiles in option price. For this reason, jump-diffusion process is widely used to improve standard Black-Scholes model. Some prior works along this direction include Merton (1976) and Kou (2002). The existence of jumps also makes the markets to be incomplete. The incomplete nature of a market challenges traditional hedging methods. Unlike the complete Black-Scholes model, a continuously rebalanced delta hedge will not lead to an instantaneously risk-free portfolio. He et al. (2006) shows that the degree of market incompleteness depends on the size and intensity of jumps, which determine the magnitude of derivative hedging error.

In summary, it is reasonable and meaningful to find some approaches which estimate parameters in jump-diffusion model accurately and efficiently. Despite the

progresses in statistical inference methods, parameter estimation under influence of jumps is still a challenge. Jump-diffusion process used in defaultable bonds pricing, fixed income derivatives pricing and interest rate modeling largely belong to affine term structure process (Duffie et al., 2000). Methods based on Fourier transform of conditional characteristic function are broadly applied to estimate parameters in such kind processes. However, Fourier transform is computationally expensive and numerically unstable. Furthermore, considerable existing methods are based on daily, weekly and monthly observed data, which are the so-called low frequency data. With emergence of accessible high frequency data, such as intraday option price and intraday stock price, many researchers hope to take advantage over the high volume of information contained in such data. While high frequency data may lead to better estimations, it is notable that many existing methods suffer from market microstructure noise.

This thesis is devoted to developing estimation methods for jump-diffusion process via low and high frequency data respectively. The thesis comprises four chapters. In Chapter 1, there is a brief introduction of background and motivations. In Chapter 2, an estimation method for the continuous state branching processes with immigration (hereafter, CBI) is developed. CBI process is a sub-class of affine process developed by Duffie et al. (2000) and has a simplified version which is coincident with Cox-Ingersol-Ross (Cox et al., 1985) model with jumps in the study of interest rate modeling. As a result, the developed method provides new perspective in interest rate model estimation. The method is based on weight least square estimators and it is suitable for low frequency data. The strength of this

method is that it avoids the computationally expensive numerical integrations in many existing methods. Consistency of the estimators are also proved. The property of this method is illustrated through a detailed simulation study. Besides, an empirical study is conducted via U. S. daily effective Federal Fund rate. In Chapter 3, particle Markov chain Monte Carlo framework is used, for the first time, to estimate parameters of jump-diffusion process via intraday stock price. High frequency data modeling is an emerging researching field. In current studies, various nonparametric methods are applied to estimate volatility for stock price. However, parametric methods are less used. Potential reasons are the difficulties in parameter estimation and the lack of proper modeling methods concerning market microstructure theory. The purpose of Chapter 3 is to provide possible solution to the problem of parameter estimation in high frequency data modeling. Particle Markov chain Monte Carlo method is applied to estimate a parametric model developed by Zeng (2003). Unlike nonparametric methods, estimation of this model identify all parameters in jump-diffusion process, besides volatility. Detailed simulation studies are conducted and a comparison is made with existing nonparametric methods. The results show that this method can estimate volatility generated by diffusion process effectively under the influence of jumps with market microstructure noise. In Chapter 4, conclusions and ideas for future research are provided.

Chapter 2

Estimation of Jump-Diffusion Process via Low Frequency Data

2.1 Motivation

Many existing models assume that interest rates move continuously, and follow some diffusion-type processes. However, recent studies provide empirical evidence that interest rates contain unexpected discontinuous changes of large magnitude. Accordingly, there is an increasing body of literature explicitly incorporating jumps into interest rates modeling. Das (2002) extends the Vasicek (1977) model to integrate jumps and shows that incorporating jumps may capture many empirical features of Federal Funds rate which cannot be explained by continuous diffusion models. Farnsworth and Bass (2003) model Federal Fund rate process with jumps and find their model explaining well the shifts in yield curve. In Johannes (2004), Cox-Ingersoll-Ross (CIR) model (Cox et al., 1985) is also extended to the jump-diffusion case. A nonparametric methods is applied to analyze trend of mean reverting, volatility and jump intensity at different levels of interest rates. Moreover, jumps in the LIBOR rates, found by Jarrow et al. (2007), can well explain the volatility smiles of interest rate caps.

Among many models, the CIR model is the most widely used one in studies of interest rate. However, it is not an easy task to extend this model to the jump-diffusion case, mainly due to the difficulties in parameter estimation. In literature jumps are generally modeled as a compound Poisson process with restrictive forms of distributions for jump magnitude. Most researchers assume that jumps have a constant or affine jump intensity with exponentially or normally distributed jump magnitudes. The reasons for these assumptions are that CIR model is a special form of affine term structure models (Duffie et al., 2000), and analytical forms of characteristic functions are available for exponential and normal distribution. Therefore analytical form of conditional characteristic function (CCF) for CIR model with jumps (JCIR) is also available according to Duffie (2005). These analytical forms are important because they can be used within maximum likelihood method or generalized method of moments to estimate parameters. Many researchers, such as Singleton (2001), Das (2002), use this kind of methods. However, methods based on CCF need to transform characteristic function to density function, by applying either Fast Fourier Transform (FFT) or other numerical integration method, such as Gauss quadrature. The transformation is computationally expensive, which needs specialties in numerical analysis. Because of these difficulties, many studies involving interest rate modeling still use CIR model without jumps.

In this chapter, a new estimation method for the continuous state branching process with immigration (hereafter, CBI) is proposed, which is based on weighted conditional least square estimators (WCLSE). Since the CBI process has JCIR model as its special form, this method leads to consistent estimators of parame-

ters for JCIR model. The originality of this method is as follows. On the one hand, it does not require transformation of characteristic functions, which makes its implementation simpler and faster. On the other hand, it does not specify special forms of jumps, thus it can be applied as a more general estimation method for real problems. This method is developed for drift and diffusion parameters of JCIR model, and it needs to be used together with other methods, such as nonparametric method developed by Bandi and Nguyen (2003), to estimate parameters for jumps.

2.2 Model Specification

As mentioned above, JCIR model is a special form of the affine term structure process and the general CBI process. For a better understanding of the estimation method to be developed in next section, it is necessary to introduce some preliminaries for both of the processes.

The following definition for affine jump-diffusion process is taken from Duffie et al. (2000).

Definition 2.1 (Affine Jump-Diffusion Process)**.** *Fix a probability space* $(\Omega, \mathcal{F}, P)$ *and an information filtration* $\{\mathcal{F}_t\}$*, an affine jump-diffusion process* X *is a Markov process in some state space* $D \in \mathbb{R}$*, satisfying the stochastic differential equation*

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t + dJ_t,$$

where W_t *is a* $(\mathcal{F}_t)$*-standard Brownian motion;* $\mu : D \to \mathbb{R}$*,* $\sigma : D \to \mathbb{R}$*, and* J_t *is a pure jump process whose jumps have a fixed probability distribution* ν *on* $\mathbb{R}$ *and*

arrive with intensity $\{\lambda(X_t) : t \geq 0\}$*, for some* $\lambda : D \to [0, \infty)$*.*

The affine jump-diffusion process X has conditional characteristic function of the form as follows, for any $u \in \mathbb{R}$,

$$\mathbf{E}(e^{iu \cdot X(t)} \mid X(s)) = e^{\varphi(t-s,iu)+\psi(t-s,iu) \cdot X(s)} \tag{2.1}$$

where $\varphi(\cdot, iu)$ and $\psi(\cdot, iu)$ are coefficients. X is regular if the coefficients $\varphi(\cdot, iu)$ and $\psi(\cdot, iu)$ of conditional characteristic function are differentiable and if their derivatives are continuous at 0. This regularity implies that these coefficients satisfy a Riccati ordinary differential equation (ODE). The general analytical solutions for Riccati ODEs are listed in Duffie (2005). These Riccati ODEs play an essential role in currently widely-used estimation methods based on conditional characteristic function, which will be briefly introduced in next section.

Duffie (2005) defines the JCIR model as a "basic affine process". Then the CIR model is treated as a simplified version of JCIR model.

Definition 2.2 (CIR Model)**.**

$$dX_t = (a + bX_t)dt + \sigma\sqrt{X_t}dW_t, \tag{2.2}$$

where W_t *is a standard Brownian motion,* a *and* b *are drift parameters, and* σ *is diffusion parameter.*

Definition 2.3 (JCIR Model)**.**

$$dX_t = (a + bX_t)dt + \sigma\sqrt{X_t}dW_t + dJ_t, \tag{2.3}$$

where J_t *is a compound Poisson process with exponential jump sizes.* W_t*,* a*,* b *and*

σ are of the same meanings as those in Equation (2.2). In addition, W_t and J_t are independent.

The general concept of CBI (continuous state branching processes with immigration) is first introduced by Kawazu and Watanable (1971). In the following, the definition taken from Dawson and Li (2006) is used.

Definition 2.4 (CBI Process)**.** *Let $\mathbb{R}_+ = [0, \infty)$. Let $m(d\xi)$ and $\mu(d\xi)$ be σ-finite measures supported by $\mathbb{R}_+ \backslash \{0\}$, such that*

$$\int_0^\infty \xi m(d\xi) + \int_0^\infty \xi \wedge \xi^2 \mu(d\xi) < \infty. \tag{2.4}$$

Let W_t be a standard Brownian motion, $N_0(ds, d\xi)$ be a Poisson random measure on $(0, \infty) \times \mathbb{R}_+$ with intensity $dsm(d\xi)$ and $N_1(ds, du, d\xi)$ be a Poisson random measure on $(0, \infty) \times (0, \infty) \times \mathbb{R}_+$ with intensity $dsdu\mu(d\xi)$. Supposing that W_t, N_0, N_1 are independent of each other, a CBI process is as follows.

$$\begin{aligned} X_t \;=\; & x_0 + \int_0^t (a + bX_s)ds + \int_0^t \sigma\sqrt{X_s} dW_s \\ & + \int_0^t \int_{\mathbb{R}_+} \xi N_0(ds, d\xi) + \int_0^t \int_0^{X_{s-}} \int_{\mathbb{R}_+} \xi \tilde{N}_1(ds, du, d\xi), \end{aligned} \tag{2.5}$$

where $x_0 \in \mathbb{R}_+, a \geq 0, b \in \mathbb{R}, \sigma \geq 0$, and $\tilde{N}_1(ds, du, d\xi) = N_1(ds, du, d\xi) - dsdu\mu(d\xi)$. The above process is also called CB process without immigration if $a = 0$ and $m = 0$ (i.e. $N_0 = 0$).

The CBI process X_t will be called supercritical, subcritical or critical respectively if $b > 0, b < 0,$ or $b = 0$. This classification also corresponds to different asymptotic behaviors of X_t as $t \to \infty$. Roughly speaking, if $b > 0$, the process explodes exponentially with rate b. If $b < 0$, the process without immigration goes to 0

almost surely, but immigration will prevent extinction, thus will yield an ergodic behavior under mild moment conditions. In the critical case, X_t has linear growth rate according to a weak convergence result proved by Huang, Ma and Zhu (2011). Although CBI process is firstly applied in population biology and demography, a comparison of Equation (2.2), Equation (2.3) and the general form of CBI process Equation (2.5) indicates that CIR model and JCIR model are all special cases of CBI process. When Poisson random measure $N_0(ds, d\xi)$ and $N_1(ds, du, d\xi)$ are set to be zero, CBI process becomes the CIR model. When only Poisson random measure $N_1(ds, du, d\xi)$ is set to be zero, CBI process becomes the JCIR model. Filipović (2001) further proves that all non-negative Markov short rate processes with affine term structure are coincident with the class of CBI processes.

Consequently, estimation method developed for CBI process can be applied discretely for JCIR model. In this chapter, consistency of parameter estimators for CBI process are proved for supercritical, subcritical and critical cases. Note that interest rate modeling requires b in Equation (2.5) should be negative, thus both of simulation and empirical studies only focus on this subcritical case.

2.3 Estimation Method

Real interest rate data is often obtained as discretely data points with equidistant time interval Δt $\{t_k = k\Delta t, k = 0, \ldots, n\}$, such as daily, weekly and monthly data. The target of this section is to estimate a, b and σ in Equation (2.5) based on the observed data stream $\{X_{t_k}\}_{k=0}^{n}$.

2.3.1 A Brief Review of Existing Estimation Methods

Before introducing our new estimation method, a review of existing estimation methods is given as follows. In general, there are mainly three kinds of estimation methods for JCIR model (Singleton, 2001): maximum likelihood method, Bayesian estimation method and generalized method of moments. These methods are all based on conditional characteristic function (CCF). Let $\phi_{X_t}(u, \Theta)$ denote the known CCF of X_{t+1} given X_t, which has the form of Equation (2.1), where Θ is parameter vector. By definition, $\phi_{X_t}(u, \Theta)$ is the Fourier transform of the density function of X_{t+1} conditioned on X_t.

$$\phi_{X_t}(u, \Theta) = \int f(X_{t+1} \mid X_t; \Theta) e^{iuX_{t+1}} dX_{t+1}. \tag{2.6}$$

Therefore, conditional density function of X_{t+1} can be obtained by an inverse Fourier transform of $\phi_{X_t}(u, \Theta)$.

$$f(X_{t+1} \mid X_t; \Theta) = \frac{1}{\pi} \int \mathrm{Re}[e^{-iuX_{t+1}} \phi_{X_t}(u, \Theta)] du. \tag{2.7}$$

Based on Equation (2.7), the log-likelihood function has the form

$$\ell(\Theta) = \log \ell(X_{1:T} \mid \Theta) = \sum_{t=1}^{T-1} \log \Big\{ \int \mathrm{Re}[e^{-iuX_{t+1}} \phi_{X_t}(u, \Theta)] du \Big\}. \tag{2.8}$$

Maximum likelihood method is implemented by maximizing Equation (2.8).

$$\hat{\Theta} = \underset{\Theta}{\mathrm{argmax}} \log \ell(\Theta). \tag{2.9}$$

In Bayesian setup, the parameters are viewed as random variables and the following

posterior density is investigated:

$$p(\Theta \mid X_{1:T}) = \frac{\ell(X_{1:T} \mid \Theta)\pi(\Theta)}{\int \ell(X_{1:T} \mid \Theta) d\Theta}, \tag{2.10}$$

where $\pi(\Theta)$ is prior density, which expresses econometrician's personal beliefs and knowledge.

Conditional moments method calculates conditional moments of X_{t+1} given X_t from derivatives of CFF at $u = 0$, thus gives a particular conditional moments,

$$\frac{\partial^j \phi_{X_t}(u, \Theta)}{\partial u^j}\Big|_{u=0} = i^j \mathbf{E}[X_{t+1}^j \mid X_t], \tag{2.11}$$

orthogonality conditions for generalized method of moments (GMM) can be constructed from the moment restrictions

$$\mathbf{E}\Big(X_{t+1}^j - \frac{\partial^j \phi_{X_t}(u, \Theta)}{i^j \partial u^j} \Big|_{u=0} \Big| X_t\Big) = 0. \tag{2.12}$$

Although the three methods above are quite intuitive, the numerical implementations are not easy tasks. Numerical integration is often needed, which is computationally expensive, and coefficient matrix of the system of equations in GMM method is sometimes singular.

2.3.2 New Estimation Method Based on WCLSE

Following the estimation approach in discrete-time and discrete-space branching processes with immigration in Wei and Winnicki (1990), the weighted conditional least square estimators (WCLSE) is introduced to avoid the problems mentioned above in methods based on conditional characteristic functions. The related esti-

mation procedure has four steps as follows.

Throughout this chapter, the following condition is needed, which means X_t has finite second moment.

$$\int_1^\infty \xi^2 m(d\xi) + \int_1^\infty \xi^2 \mu(d\xi) < \infty. \tag{2.13}$$

Let $l_{01} = \int_0^\infty \xi m(d\xi)$, $l_{02} = \int_0^\infty \xi^2 m(d\xi)$, $l_{11} = \int_0^\infty \xi \mu(d\xi)$ and $l_{12} = \int_0^\infty \xi^2 \mu(d\xi)$. Additionally, the following condition is also assumed to be true.

$$a + l_{01} > 0 \text{ and } \sigma^2 + l_{12} > 0. \tag{2.14}$$

Step 1. Write X_k for X_{t_k}. By Equation (2.5),

$$X_k = \gamma_0 + \gamma_1 X_{k-1} + \varepsilon_k, \tag{2.15}$$

where $\varepsilon_k = X_k - \mathbf{E}[X_k \mid X_{k-1}], \gamma_1 = e^{b\Delta t}$, $\gamma_0 = \frac{a+l_{01}}{b}(e^{b\Delta t} - 1)$ if $b \neq 0$, otherwise $\gamma_0 = (a + l_{01})\Delta t$ if $b = 0$. Then ε_k is a martingale difference w.r.t. $\{\mathcal{F}_k\}$, where $\mathcal{F}_k = \sigma\{X_0, X_1, \ldots, X_k\}$.

Step 2. Rewrite Equation (2.15) as follows,

$$\frac{X_k}{(X_{k-1}+1)^{\frac{1}{2}}} = \gamma_1 (X_{k-1}+1)^{\frac{1}{2}} + (\gamma_0 - \gamma_1)(X_{k-1}+1)^{-\frac{1}{2}} + \delta_k, \tag{2.16}$$

where $\delta_k = \varepsilon_k/(X_{k-1}+1)^{\frac{1}{2}}$. Note that $\mathbf{E}[\delta_k^2 \mid \mathcal{F}_{k-1}] = \frac{\eta_1 X_{k-1} + \eta_0}{X_{k-1}+1}$, where, if $b \neq 0$,

$$\eta_0 = \frac{(\sigma^2 + l_{12})(a + l_{01})}{2b^2}(e^{b\Delta t} - 1)^2 + \frac{l_{02}}{2b}(e^{2b\Delta t} - 1),\ \eta_1 = \frac{\sigma^2 + l_{12}}{b} e^{b\Delta t}(e^{b\Delta t - 1}). \tag{2.17}$$

if $b = 0$, $\eta_0 = (\sigma^2 + l_{12})(a + l_{01})(\Delta t)^2/2 + l_{02}\Delta t$, $\eta_1 = (\sigma^2 + l_{12})\Delta t$.

Step 3. Based on Equation (2.16), the following WCLSEs are obtained for a, b.

$$\hat{b}_n = \frac{1}{\Delta t}\log\frac{\sum_{k=1}^{n} X_k \sum_{k=1}^{n} \frac{1}{X_{k-1}+1} - n\sum_{k=1}^{n}\frac{X_k}{X_{k-1}+1}}{\sum_{k=1}^{n}(X_{k-1}+1)\sum_{k=1}^{n}\frac{1}{X_{k-1}+1} - n^2}, \tag{2.18}$$

$$\hat{a}_n = \frac{\frac{1}{n}(\sum_{k=1}^{n} X_k - e^{\hat{b}_n \Delta t}\sum_{k=1}^{n} X_{k-1})}{e^{\hat{b}_n \Delta t} - 1}\hat{b}_n - l_{01}. \tag{2.19}$$

Step 4. Once a and b are obtained, σ^2 can be estimated by least square estimation method according to Ludger and Rydén (1997). Let's consider conditional variance of ε_k : $\mathbf{Var}(\varepsilon_k \mid \mathcal{F}_{k-1})$.

$$\begin{aligned}
\mathbf{Var}(\varepsilon_k \mid \mathcal{F}_{k-1}) &= \mathbf{E}[(\varepsilon_k - \mathbf{E}[\varepsilon_k \mid \mathcal{F}_{k-1}])^2 \mid \mathcal{F}_{k-1}] \\
&= \mathbf{E}[\varepsilon_k^2 \mid \mathcal{F}_{k-1}] - \mathbf{E}^2[\varepsilon_k \mid \mathcal{F}_{k-1}] \\
&= \mathbf{E}[\varepsilon_k^2 \mid \mathcal{F}_{k-1}]. \qquad (2.20)
\end{aligned}$$

The last equation holds because ε_k is martingale difference.

$$\mathbf{E}[\varepsilon_k^2 \mid \mathcal{F}_{k-1}] = \eta_1 X_{k-1} + \eta_0 = (K_1 X_{k-1} + K_0)\sigma^2 + K_3 X_{k-1} + K_2, \tag{2.21}$$

where

$$\begin{aligned}
K_0 &= \frac{a + l_{01}}{2b^2}(e^{b\Delta t - 1} - 1)^2, && (2.22)\\
K_1 &= \frac{e^{b\Delta t}}{b}(e^{b\Delta t} - 1), && (2.23)\\
K_2 &= \frac{l_{12}(a + l_{01})}{2b^2}(e^{b\Delta t - 1} - 1)^2 + \frac{l_{02}}{2b}(e^{2b\Delta t} - 1), && (2.24)\\
K_3 &= \frac{l_{12}}{b}e^{b\Delta t}(e^{b\Delta t - 1}). && (2.25)
\end{aligned}$$

Equation (2.21) is considered as a regression equation: $Y_k = Z_k\sigma^2 + e_k$, where

$$Y_k = (X_k - \gamma_0 - \gamma_1 X_{k-1})^2 - K_3 X_{k-1} - K_2, \tag{2.26}$$

$$Z_k = K_1 X_{k-1} + K_0, \tag{2.27}$$

and e_k is an error process with unknown form. Therefore σ^2 can be calculated by standard least square estimation process:

$$\hat{\sigma}^2 = \frac{\sum_{k=1}^{n} Z_k Y_k}{\sum_{k=1}^{n} Z_k^2}. \tag{2.28}$$

Note that a and b must be substituted by corresponding estimators, and then $\hat{\eta_1}, \hat{\eta_0}$, $\hat{K}_0, \hat{K}_1, \hat{K}_2, \hat{K}_3, \hat{Y}_k, \hat{Z}_k$ can be calculated.

2.4 Asymptotic properties of the WCLSE

Theorem 2.1. *Assume that $b < 0$, and condition (2.13) - (2.14) hold. Then $\hat{a}_n$ and $\hat{b}_n$ are strongly consistent. Furthermore, $\Big(\sqrt{n}(\hat{b}_n - b), \sqrt{n}(\hat{a}_n - a)\Big) \xrightarrow{d} N(0, (UV)W(UV)')$, where*

$$U = \begin{pmatrix} \frac{1}{e^{b\Delta t}\Delta t} & 0 \\ \frac{a+l_{01}}{be^{b\Delta t}\Delta t} - \frac{a+l_{01}}{e^{b\Delta t}-1} & \frac{b}{e^{b\Delta t}-1} \end{pmatrix},$$

$$V = \Big(\mathbf{E}[X+1]\mathbf{E}\big[\frac{1}{X+1}\big] - 1\Big)^{-1} \begin{pmatrix} \mathbf{E}\big[\frac{1}{X+1}\big] & -1 \\ -\mathbf{E}\big[\frac{X}{X+1}\big] & \mathbf{E}[X] \end{pmatrix},$$

$$W = \begin{pmatrix} \eta_0 + \eta_1 \mathbf{E}[X] & \mathbf{E}\big[\frac{\eta_0+\eta_1 X}{X+1}\big] \\ \mathbf{E}\big[\frac{\eta_0+\eta_1 X}{X+1}\big] & \mathbf{E}\big[\frac{\eta_0+\eta_1 X}{(X+1)^2}\big] \end{pmatrix},$$

and X is a random variable with the stationary distribution of X_n as $n \to \infty$.

Proof. By Pinsky (1971) and condition (2.13), if $b < 0$, we have $X_t \xrightarrow{d} X$. If we further assume that $\sigma > 0$, then $X(\cdot)$ is ergodic. From the Birkhoff's ergodic theorem (see Theorem 7.2.1 in Durrett, 2010), we have

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{a.s.} -\frac{a+l_{01}}{b}, \quad \frac{1}{n}\sum_{k=1}^{n} \frac{1}{1+X_{k-1}} \xrightarrow{a.s.} \mathbf{E}\Big[\frac{1}{1+X}\Big],$$
$$\frac{1}{n}\sum_{k=1}^{n} \frac{X_k}{X_{k-1}+1} \xrightarrow{a.s.} \gamma_1 + (\gamma_0 - \gamma_1)\mathbf{E}\Big[\frac{1}{1+X}\Big].$$

Then by Equation (2.18) - (2.19), $\hat{b}_n \xrightarrow{a.s.} b$ and $\hat{a}_n \xrightarrow{a.s.} a$. Let $\hat{\theta}_{1n} = e^{\hat{b}_n \Delta t}$ and $\hat{\theta}_{2n} = \frac{\hat{a}_n + l_{01}}{\hat{b}_n}(e^{\hat{b}_n \Delta t} - 1)$. By Taylor's theorem, $\hat{b}_n - b = \frac{1}{\Delta t}(\hat{\theta}_{1n} - \gamma_1)(e^{-b\Delta t} + o_p(1))$ and

$$\begin{aligned} \hat{a}_n - a \;=\; & \Big[\frac{\gamma_0}{e^{b\Delta t}\Delta t(e^{\hat{b}_n \Delta t} - 1)} - \frac{\gamma_0 b}{(e^{\hat{b}_n \Delta t} - 1)(e^{b\Delta t} - 1)} + o_p(1)\Big](\hat{\theta}_{1n} - \gamma_1) \\ & + \frac{\hat{b}_n}{e^{\hat{b}_n \Delta t} - 1}(\hat{\theta}_{2n} - \gamma_0). \end{aligned} \tag{2.29}$$

As in Wei and Winnicki (1990), $\Big(\sqrt{n}(\hat{\theta}_{1n} - \gamma_1), \sqrt{n}(\hat{\theta}_{2n} - \gamma_0)\Big) = V_n Z_n$, where

$$V_n = \Big(\frac{1}{n^2}\sum_{k=1}^{n}(X_{k-1}+1)\sum_{k=1}^{n}\frac{1}{X_{k-1}+1} - 1\Big)^{-1} \begin{pmatrix} \frac{1}{n}\sum_{k=1}^{n}\frac{1}{X_{k-1}+1} & -1 \\ -\frac{1}{n}\sum_{k=1}^{n}\frac{X_{k-1}}{X_{k-1}+1} & \frac{1}{n}\sum_{k=1}^{n} X_k \end{pmatrix}, \tag{2.30}$$

$$Z_n' = \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\varepsilon_k & \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\frac{\varepsilon_k}{X_{k-1}+1} \end{pmatrix}.$$

By the ergodic theorem, $V_n \xrightarrow{a.s.} V$. From condition (2.14), X is not degenerate. It follows from Jensen's inequality and non-degeneracy of X, $\det(V) > 0$. By Equation (2.29) - (2.30), it suffices to prove that $Z_n \xrightarrow{d} N(0, W)$. Note that

$$\frac{1}{n}\sum_{k=1}^{n}\mathbf{E}[\varepsilon_k^2|\mathcal{F}_{k-1}] \xrightarrow{p} \eta_0 + \eta_1\mathbf{E}[X], \quad \frac{1}{n}\sum_{k=1}^{n}\mathbf{E}\Big[\frac{\varepsilon_k^2}{(X_{k-1}+1)^2}|\mathcal{F}_{k-1}\Big] \xrightarrow{p} \mathbf{E}\Big[\frac{\eta_0+\eta_1 X}{(X+1)^2}\Big],$$

$$\frac{1}{n}\sum_{k=1}^{n}\mathbf{E}\Big[\frac{\varepsilon_k^2}{(X_{k-1}+1)}|\mathcal{F}_{k-1}\Big] \xrightarrow{p} \mathbf{E}\Big[\frac{\eta_0+\eta_1 X}{X+1}\Big].$$

For any $\epsilon > 0$, $\frac{1}{n}\sum_{k=1}^{n}\mathbf{E}[\varepsilon_k^2 1_{\{|\varepsilon_k|>\sqrt{n}\epsilon\}}|\mathcal{F}_{k-1}] \leq \frac{1}{n}\sum_{k=1}^{n}\mathbf{E}[\varepsilon_k^2 1_{\{|\varepsilon_k|>\sqrt{k}\epsilon\}}|\mathcal{F}_{k-1}]$. Note that $\varepsilon_k^2 = (X_k - \gamma_1 X_{k-1} - \gamma_0)^2$ is stationary since (X_{k-1}, X_k) is stationary. As $k \to \infty$, $\mathbf{E}[\varepsilon_k^2] \to \frac{a+l_{01}}{b}\eta_1 + \eta_0$. Then ε_k^2 is uniformly integrable and $\frac{1}{n}\sum_{k=1}^{n}\mathbf{E}[\varepsilon_k^2 1_{\{|\varepsilon_k|>\sqrt{n}\epsilon\}}|\mathcal{F}_{k-1}] \to 0$. Similarly,

$$\frac{1}{n}\sum_{k=1}^{n}\mathbf{E}\Big[\frac{\varepsilon_k^2}{(X_{k-1}+1)^2}1_{\{|\varepsilon_k/(X_{k-1}+1)|>\sqrt{n}\epsilon\}}|\mathcal{F}_{k-1}\Big] \to 0.$$

As a result, the Linderberg conditions are satisfied. The remaining proof follows from the martingale central limit theorem, see Hall and Heyde (1980). □

Now we concentrate on the supercritical case. By Pinsky (1971) and condition (2.13) - (2.14), if $b > 0$, then as $t \to \infty$, $X_t/e^{bt} \xrightarrow{a.s.} L$ for some positive variable L and $0 < L < \infty$ $a.s.$ Let $\mathbb{R}^\infty$ denote the space of real sequences $x = (x_1, x_2, \cdots)$ with metric $d(x, y) = \sum_{j=1}^{\infty} 2^{-j}|x_j - y_j|/(1 + |x_j - y_j|)$. Recall that δ_k is defined in Equation (2.16) for $k \geq 1$. We consider the $\mathbb{R}^\infty$-valued random variables $\Gamma_n =$

$\{\varrho_{nj}\}$ and $\Gamma = \{\varrho_j\}$ defined by

$$\varrho_{nj} = \begin{cases} \delta_{n-j+1} & j = 1, 2, \ldots, n, \\ 0 & \text{otherwise}, \end{cases} \tag{2.31}$$

and ϱ_j are i.i.d. random variables distributed as $N(0, \eta_1)$, where η_1 is defined in Equation (2.17).

Lemma 2.1. *Assume that $b > 0$, conditions (2.13) - (2.14) hold. Then Γ_n converges weakly to Γ in $\mathbb{R}^\infty$.*

Proof. Recall that ε_k is defined in Equation (2.15) and $X_k = X_{t_k}$. For $j \leq n$, we note that

$$\begin{aligned} \varepsilon_{n-j+1} \;=\; & \int_{t_n-j}^{t_n-j+1} \sigma e^{b(t_{n-j+1}-s)} \sqrt{X(s)} dW_s \\ & + \int_{t_{n-j}}^{t_{n-j+1}} \int_0^{X(s-)} \int_{\mathbb{R}_+} e^{b(t_{n-j+1}-s)} \xi \tilde{N}_1(ds, du, d\xi) \\ & + \int_{t_{n-j}}^{t_{n-j+1}} \int_{\mathbb{R}_+} e^{b(t_{n-j+1}-s)} \xi \tilde{N}_0(ds, d\xi). \end{aligned}$$

Let $\phi_{n-j}(s) = e^{b(s-t_n-j)}(X_{n-j} + 1)$. Define

$$\delta'_{n-j+1} = \int_{t_{n-j}}^{t_{n-j+1}} \sigma e^{\frac{b}{2}(t_{n-j+2}-s)} dW_s + \int_{t_{n-j}}^{t_{n-j+1}} \int_0^{\phi_{n-j}(s)} \int_{\mathbb{R}_+} \frac{e^{b(t_{n-j+1}-s)}\xi}{(X_{n-j}+1)^{\frac{1}{2}}} \tilde{N}_1(ds, du, d\xi).$$

It is not hard to see that for fixed j and some positive constant C,

$$\mathbf{E}[(\delta_{n-j+1} - \delta'_{n-j+1})^2] \leq C\mathbf{E}\big[\int_0^{\Delta t} e^{2b(\Delta t - s)} | \frac{X_{s+t_n-j}}{X_{t_{n-j}} + 1} - e^{bs} | ds \big] + C \frac{\int_0^{\Delta t} e^{2b(\Delta - s)} ds}{\mathbf{E}[X_{t_{n-j}} + 1]},$$

which converges to 0 as $n \to \infty$. Define ϱ'_{nj} as in Equation (2.31) with δ_{n-j+1} replaced by δ'_{n-j+1}. Then Lemma 2.1 is equivalent to the weak convergence of

$\{\varrho'_{nj}, 1 \le i \le k\}$ for all $k \ge 1$. Without loss of generality, we only consider the case $k = 2$. For $\lambda_j \in \mathbb{R}$ $(j = 1, 2)$ and $i^2 = -1$,

$$\mathbf{E}[e^{i\lambda_1\varrho_{n1}+i\lambda_2\varrho_{n2}}] = e^{-\frac{1}{2}\sigma^2\lambda_1^2\int_0^{\Delta t} e^{b(2\Delta t-s)}ds}\mathbf{E}[e^{i\lambda_2\varrho_{n2}}e^{\upsilon_n}],$$

where

$$\begin{aligned}\upsilon_n &= \int_0^{\Delta t}\int_{\mathbb{R}_+} e^{bs}(X_{n-1}+1)[e^{ie^{b(\Delta t-s)}\xi(X_{n-1}+1)^{-\frac{1}{2}}} - 1 - ie^{b(\Delta t-s)}\xi(X_{n-1}+1)^{-\frac{1}{2}}]\mu(d\xi)ds \\ &= -\frac{l_{12}}{2}\int_0^{\Delta t} e^{b(2\Delta t-s)}ds + u_n.\end{aligned}$$

Here, u_n is some random variable, $\sup_n |u_n|$ is bounded and $|u_n| \xrightarrow{a.s.} 0$ as $n \to \infty$. Note that $\left|e^{\upsilon_n} - e^{-\frac{l_{12}}{2}\int_0^{\Delta t} e^{b(2\Delta t-s)}ds}\right| \le K|u_n|$ for some positive constant K. We have

$$\begin{aligned}\mathbf{E}[e^{i\lambda_1\rho_{n1}+i\lambda_2\rho_{n2}}] &= e^{-\frac{1}{2}(\sigma^2+l_{12})\lambda_1^2\int_0^{\Delta t} e^{b(2\Delta t-s)}ds}\mathbf{E}[e^{i\lambda_2\varrho_{n2}}] + o(1) \\ &= e^{-\frac{1}{2}[(\sigma^2+l_{12})\lambda_1^2+(\sigma^2+l_{12})\lambda_2^2]\int_0^{\Delta t} e^{b(2\Delta t-s)}ds} + o(1),\end{aligned}$$

which converges to $e^{-\frac{1}{2}\eta_1(\lambda_1^2+\lambda_2^2)}$, as $n \to \infty$. □

Lemma 2.2. *Under the conditions of Lemma 2.1, we have as $n \to \infty$,*

$$\Big[\sum_{j=1}^{n}(X_{j-1}+1)\Big]^{-\frac{1}{2}}\sum_{j=1}^{n}\varepsilon_j \xrightarrow{d} N(0, \eta_1). \tag{2.32}$$

Proof. We first claim that as $n \to \infty$,

$$V_n = \Big(\frac{e^{bn\Delta t}}{e^{b\Delta t}-1}\Big)^{-\frac{1}{2}}\sum_{j=1}^{n} e^{b(j-1)\Delta t/2}\delta_j \xrightarrow{d} N(0, \eta_1). \tag{2.33}$$

In fact, for any $k \ge 1$ let $M_{nk} = (e^{b\Delta t}-1)^{\frac{1}{2}}\sum_{j=1}^{k} e^{-bj\Delta t/2}\varrho_{nj}$. Then we have

$V_n = M_{nn}$. By Lemma 2.1, for any fixed k, $M_{nk} \xrightarrow{d} M_k = (e^{b\Delta t-1})^{\frac{1}{2}} \sum_{j=1}^{k} e^{-bj\Delta t/2}\varrho_j$. It is easy to see that $M_k \xrightarrow{d} N(0, \eta_1)$ as $k \to \infty$. Also note that for any $\epsilon > 0$,

$$\mathbf{P}(|M_{nk} - M_{nn}| > \epsilon) = \frac{1}{\epsilon^2}\mathbf{E}\big[\sum_{j=k+1}^{n} e^{-bj\Delta t}\gamma_{nk}^2\big] \le \frac{\eta_1 + \eta_0}{\epsilon^2} \sum_{j=k+1}^{\infty} e^{-bj\Delta t}.$$

Thus it follows from Theorem 3.2 in Billingsley (1999) that Equation (2.33) holds. Using the same method as in Theorem 3.5 in Wei and Winnicki (1990), we have

$$\sum_{j=1}^{n}[(X_{j-1}+1)^{\frac{1}{2}} - (e^{b(j-1)\Delta t}L)^{\frac{1}{2}}]\delta_j = o_p(e^{bn\Delta t/2}). \tag{2.34}$$

Also note that

$$e^{-bn\Delta t}\sum_{k=1}^{n}(X_{k-1}+1) \xrightarrow{a.s.} \frac{1}{e^{b\Delta t}-1}L. \tag{2.35}$$

By Equation (2.33), we obtain Equation (2.32). □

Theorem 2.2. *Under the conditions of Lemma 2.1, $\hat{b}_n$ is strongly consistent while $\hat{a}_n$ is not weakly consistent. Furthermore,*

$$\Big[\sum_{j=1}^{n}(X_{j-1}+1)\Big]^{\frac{1}{2}}(\hat{b}_n - b) \xrightarrow{d} \frac{1}{\Delta t e^{b\Delta t}}N(0, \eta_1). \tag{2.36}$$

Proof. Since $X_j/X_{j-1} \xrightarrow{a.s.} e^{b\Delta t}$ as $j \to \infty$, we have $n^{-1}\sum_{j=1}^{n}\frac{X_j}{X_{j-1}+1} \xrightarrow{a.s.} e^{b\Delta t}$. Furthermore,

$$\sum_{j=1}^{\infty}\frac{1}{X_{j-1}+1} < \infty \ a.s. \tag{2.37}$$

By Equation (2.18), it is easy to see that $\hat{b}_n \xrightarrow{a.s.} b$. Note that $\hat{b}_n - b = \frac{1}{\Delta t}(\hat{\theta}_{1n} - \gamma_1)(e^{-b\Delta t} + o_p(1))$ and

$$\Big[\sum_{j=1}^{n}(X_{j-1}+1)\Big]^{\frac{1}{2}}(\theta_{1n} - \gamma_n) = \frac{A_n - B_n}{1 - C_n},$$

where

$$
\begin{aligned}
A_n &= \Big[\sum_{j=1}^{n}(X_{j-1}+1)\Big]^{-\frac{1}{2}}\sum_{j=1}^{n}\varepsilon_j,\\
B_n &= \Big[n\sum_{j=1}^{n}\frac{\varepsilon_j}{X_{j-1}+1}\Big]\Big/\Big[\sum_{j=1}^{n}\frac{1}{X_{j-1}+1}(\sum_{j=1}^{n}(X_{j-1}+1))^{\frac{1}{2}}\Big],\\
C_n &= n^2\Big[\sum_{j=1}^{n}(X_{j-1}+1)\sum_{j=1}^{n}\frac{1}{X_{j-1}+1}\Big]^{-1}.
\end{aligned}
$$

Note that $\sum_{j=1}^{\infty}\mathbf{E}\big[\frac{\varepsilon_j^2}{(X_{j-1}+1)^2}\mid\mathcal{F}_{j-1}\big]=\sum_{j=1}^{\infty}\frac{\eta_1X_{j-1}+\eta_0}{X_{j-1}+1}<\infty$. It follows from Theorem 2.17 in Hall and Heyde (1980) that $\sum_{j=1}^{n}\frac{\varepsilon_j}{X_{j-1}+1}$ converges a.s. By Equation (2.35) and Equation (2.37), $B_n\overset{a.s.}{\to}0$ and $C_n\overset{a.s.}{\to}0$. Thus Equation (2.36) follows from Lemma 2.1. By Equation (2.29) and Equation (2.36), it is not hard to see that

$$
n\Big[\sum_{j=1}^{n}(X_{j-1}+1)\Big]^{-\frac{1}{2}}(\hat{a}_n-a)\overset{d}{\to}\frac{2b}{e^{b\Delta t}-1}N(0,\eta_1)
$$

which implies that $\hat{a}_n-a\overset{p}{\to}\infty$, as $n\to\infty$. □

Theorem 2.3. *Let $Y_n(t)=X_{[nt]}/n$. Assume that $b=0$ and condition (2.13) holds. Then $Y_n(\cdot)$ converges in distribution on $D([0,\infty),\mathbb{R}_+)$ to a CBI process defined by*

$$
Y(t)=(a+l_{01})t+\int_0^t\sqrt{(\sigma^2+l_{12})Y(s)}\,dW(s), \tag{2.38}
$$

where $Y(0)=0$ and $W(\cdot)$ is a standard Brownian motion.

Proof. This proof is divided into four steps. We may rewrite Equation (2.5) as

follows.

$$\frac{X_{nt}}{n} = \frac{x_0}{n} + at + \int_0^{nt} \frac{\sigma\sqrt{X_s}}{n} dW(s) + \int_0^{nt} \int_{\mathbb{R}_+} \frac{\xi}{n} N_0(ds, d\xi) + \int_0^{nt} \int_0^{X_{s-}} \int_{\mathbb{R}_+} \frac{\xi}{n} \tilde{N}_1(ds, du, d\xi). \quad (2.39)$$

Step 1. Applying Doob's inequality to the martingale terms in Equation (2.39), we obtain

$$\mathbf{E}\Big[\frac{1}{n}\sup_{0\le s\le t} X_{ns}\Big] \le \frac{x_0}{n} + (a + l_{01})t + 4\sigma\Big(\int_0^{nt} \frac{\mathbf{E}[X_s]}{n^2} ds\Big)^{\frac{1}{2}} + 4\Big(l_{12}\int_0^{nt} \frac{\mathbf{E}[X_s]}{n^2} ds\Big)^{\frac{1}{2}}.$$

Let $C(t) := 1 + \limsup_{n\to\infty} \mathbf{E}\big[\frac{1}{n}\sup_{0\le s\le t} X_{ns}\big]$. Since $\mathbf{E}[X_t] = (a + l_{01})t$, $C(t)$ is a locally bounded function of $t \ge 0$. Similarly, $\limsup_{n\to\infty} \mathbf{E}\big[\frac{1}{n^2}\sup_{0\le s\le t} X_{ns}^2\big]$ is also locally bounded.

Step 2. Tightness. Since $C(t)$ is locally bounded, X_{nt}/n is a tight sequence of random variables for every $t \ge 0$. Let $\{\tau_n\}$ be a sequence of stopping times bounded by T and let $\{\delta_n\}$ be a sequence of positive constants such that $\delta_n \to 0$ as $n \to 0$. By the properties of independent increments of Brownian motion and Poisson process we have

$$\mathbf{E}\big[\frac{1}{n}\big|X_{n(\tau_n+\delta_n)} - X_{n\tau_n}\big|\big] \le \frac{x_0}{n} + (a + l_{01})\delta_n + 4\sigma\Big(\int_0^{\delta_n} C(T + s)ds\Big)^{\frac{1}{2}} + 4\Big(l_{12}\int_0^{\delta_n} C(T + s)ds\Big)^{\frac{1}{2}},$$

which converges to 0 as $n \to \infty$. Then $\frac{X_{nt}}{n}$ is tight in $D([0, \infty), \mathbb{R}_+)$ by the criterion of Aldous (1978).

Step 3. Limiting. Let $Y(\cdot)$ be any limit point of $\frac{X_{nt}}{n}$. Without loss of generality, by

Skorokhod's theorem, we can assume that on some Skorokhod's space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbf{P})$, $\frac{X_{nt}}{n} \xrightarrow{a.s.} Y(t)$ in the topology of $D([0,\infty), \mathbb{R}_+)$. We claim that for any fixed $\lambda \in \mathbb{R}$,

$$M(t) = e^{-\lambda Y(t)} - 1 - \int_0^t e^{-\lambda Y(s)} \Big[\frac{1}{2}(\sigma^2 + l_{12})\lambda^2 Y(s) - (a + l_{01})\lambda \Big] ds \quad (2.40)$$

is a square integrable $\mathcal{F}_t$-martingale. In fact, by Itô's formula, it is not hard to show that

$$M_n(t) = e^{-\lambda \frac{X_{nt}}{n}} - e^{-\lambda \frac{X_0}{n}} - \int_0^t e^{-\lambda \frac{X_{ns}}{n}} A_n(\frac{X_{ns}}{n}) ds$$

is a square integrable martingale, where $A(x) = \frac{1}{2}x\sigma^2\lambda^2 + xn^2 \int_0^\infty (e^{-\lambda\xi/n} - 1 + \lambda\xi/n)\mu(d\xi) - a\lambda + n\int_0^\infty (e^{-\lambda\xi/n} - 1)m(d\xi)$. By condition (2.13), the tightness of X_{nt}/n, Problem 13 (P.151) in Ethier and Kurtz (1986) and Proposition 1.23 (P.293) in Jacod and Schiryaev (1987), we have

$$M_n(t) \xrightarrow{a.s.} M(t) \quad \text{in } D([0,\infty), \mathbb{C}), \qquad \text{as } n \to \infty. \quad (2.41)$$

Then for all $t \geq 0$, $M_n(t) \xrightarrow{a.s.} M(t)$ in $\mathbb{R}_+$. Since $\limsup_{n\to\infty} \mathbf{E}\big[\frac{1}{n^2} \sup_{0\leq s\leq t} X_{ns}^2\big]$ is locally bounded, $\sup_n \mathbf{E}[M_n^2(t)] < \infty$. Then for any $t \geq 0$, $M_n(t) \xrightarrow{L_2} M(t)$. Thus $M(t)$ is a martingale.

Step 4. It follows from (2.40) and Theorem 2.42 (P86) in Jocod and Schiryaev (1987) that $Y(\cdot)$ is a semi-martingale and it admits the canonical representation $Y(t) = (a+l_{01})t + Y_c(t)$, where $Y_c(t)$ is a continuous local martingales with quadratic covariation process $\int_0^t (\sigma^2 + l_{12}) Y_s \, ds$. Therefore, $Y(\cdot)$ is the solution of the stochastic equation (2.38). In addition, by Step 2, we have the weak convergence for X_{nt}/n. Since $[nt]/n \to t$ as $n \to \infty$, we have Proposition 2.3. □

Remark 2.1. By the above theorem and the continuous mapping theorem,

$$\big(X_n/n, n^{-2}\sum_{k=1}^{n} X_{k-1}\big) \xrightarrow{d} \big(Y(1), \int_0^1 Y(t)dt\big).$$

As calculated in Pitman and Yor (1982), the Laplace transform of $(Y(1), \int_0^1 Y(t)dt)$ is given by

$$\begin{aligned}&\mathbf{E}[e^{-2\lambda_1 Y(1)-2\lambda_2\int_0^1 Y(t)dt}]\\ &= \left(\cosh(2(\sigma^2+l_{12})^2\lambda_2^{\frac{1}{2}}) + 2\lambda_1\lambda_2^{-\frac{1}{2}}\sinh(2(\sigma^2+l_{12})\lambda_1^{\frac{1}{2}})\right)^{-2(a+l_{01})/(\sigma^2+l_{12})}.\end{aligned}$$

Theorem 2.4. *Assume that $b = 0$ and condition (2.13) - (2.14) hold. Then $\hat{b}_n \xrightarrow{p} b$ and*

$$n(\hat{b}_n - b) \xrightarrow{d} \frac{Y(1) - (a + l_{01})}{\Delta t e^{\Delta t}\int_0^1 Y(t)dt}, \tag{2.42}$$

where $Y(\cdot)$ is defined by Equation (2.38).

Proof. By Proposition 2.3, $X_n/n \xrightarrow{d} Y(1)$. It follows from Remark 2.1 that $Y(1) > 0$ and $\int_0^1 Y(t)dt > 0$, *a.s.* Then $X_n \xrightarrow{p} \infty$. We can find some subsequence $\{n_k\}$ such that $X_{n_k} \xrightarrow{a.s.} \infty$. Define

$$M_n = (X_n + 1)\big/\big[\Pi_{k=1}^{n}(1 + \eta_0/(X_{k-1} + 1))\big].$$

We see that M_n is a positive $\mathcal{F}_n$-martingale. By the martingale convergence theorem, M_n converges a.s. Then $\Pi_{j=1}^{n_k}(1 + \eta_0/(X_{j-1} + 1)) \xrightarrow{a.s.} \infty$. This implies that $\sum_{j=1}^{n_k} \eta_0/(X_{j-1}+1) \xrightarrow{a.s.} \infty$ and thus $\sum_{j=1}^{n} \eta_0/(X_{j-1}+1) \xrightarrow{a.s.} \infty$. On the other hand,

$\sum_{j=1}^{n} \varepsilon_k/(X_{j-1}+1)$ is a $\mathcal{F}_n$-martingale and

$$\sum_{j=1}^{n} \mathbf{E}\Big[\big(\frac{\varepsilon_j}{X_{j-1}+1}\big)^2\Big|\mathcal{F}_{j-1}\Big] = \sum_{j=1}^{n} \frac{\eta_1 X_{j-1}+\eta_0}{(X_{j-1}+1)^2} \le (\eta_1+\eta_0)\Big(\sum_{j=1}^{n} \frac{1}{X_{j-1}+1}\Big).$$

By the local martingale convergence theorem (see Theorem 2.17 in Wei and Winnicki, 1989), we have

$$\sum_{j=1}^{n} \frac{\varepsilon_j}{X_{j-1}+1} = o\Big(\sum_{j=1}^{n} \frac{1}{X_{j-1}+1}\Big) \quad a.s. \tag{2.43}$$

Now we first consider

$$n(\theta_{1n}-\gamma_1) = \frac{A_n - B_n}{1-C_n}, \tag{2.44}$$

where

$$A_n = \frac{n\sum_{j=1}^{n}\varepsilon_j}{\sum_{j=1}^{n}(X_{j-1}+1)}, \quad B_n = n^2 \sum_{j=1}^{n} \frac{\varepsilon_j}{X_{j-1}+1}\Big(\sum_{j=1}^{n}(X_{j-1}+1)\sum_{j=1}^{n}\frac{1}{X_{j-1}+1}\Big)^{-1},$$

$$C_n = n^2\Big(\sum_{j=1}^{n}(X_{j-1}+1)\sum_{j=1}^{n}\frac{1}{X_{j-1}+1}\Big)^{-1}.$$

By Remark 2.1 and Equation (2.43), we have $\sum_{j=1}^{n}\varepsilon_j/n \xrightarrow{d} Y(1)-(a+l_{01})$, $B_n \xrightarrow{p} 0$ and $C_n \xrightarrow{p} 0$. Thus

$$n(\hat{\theta}_{1n}-\gamma_1) \xrightarrow{d} \frac{Y(1)-(a+l_{01})}{\int_0^1 Y(t)dt}, \tag{2.45}$$

and then $\hat{\theta}_{1n} \xrightarrow{p} \gamma_1$. By Taylor's theorem, $\hat{b}_n - b = \frac{1}{\Delta t}(\hat{\theta}_{1n}-\gamma_1)(e^{-b\Delta t}+o_p(1))$ and (2.42) follows from (2.45). □

Remark 2.2. In the critical case, Theorem 2.4 shows that the WCLSEs $\hat{b}_n$ has

a non-Gaussian asymptotic distribution with normalizing factor n. However, it is not known what is the asymptotic distribution for $\hat{a}_n$, which may depend on the limiting behavior of $\sum_{j=1}^{n} \frac{1}{x_{j-1}+1}$ as $n \to \infty$.

Theorem 2.5. *Assume that $b < 0$, and condition (2.13) - (2.14) hold, then $\hat{\sigma}^2$ in Equation (2.28) is strongly consistent.*

Proof. As in Ludger and Rydén (1997), let θ be parameter vector. Fix θ, let $h(x, y; \theta) = (K_1 x + K_0)\{(y - \gamma_1 x - \gamma_0)^2 - K_3 x - K_2\}$, and let $U \subset \Theta$ be a neighborhood of θ such that $\mathbf{E}_\theta\Big[\sup_{\theta \in U}|h(X_0, X_1; \theta)|\Big] < \infty$. By Theorem 2.1 and ergodicity of $\{X_k\}$, the inequality

$$\begin{aligned}
&\limsup_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}(K_1 X_{k-1} + K_0)\{(X_k - \gamma_1 X_{k-1} - \gamma_0)^2 - K_3 X_{k-1} - K_2\} \\
\leq\ &\limsup_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\limsup_{\theta\in U}(K_1 X_{k-1} + K_0)\{(X_k - \gamma_1 X_{k-1} - \gamma_0)^2 - K_3 X_{k-1} - K_2\} \\
=\ &\mathbf{E}_\theta\Big[\sup_{\theta\in U}|h(X_0, X_1; \theta)|\Big],
\end{aligned}$$

holds almost surely. Then let $U \downarrow \{\theta\}$, because $K_0, K_1, K_2, K_3, \gamma_1, \gamma_0$ are all continuous functions of θ, according to the dominated convergence theorem, the conclusion is drawn that

$$\begin{aligned}
&\limsup_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}(K_1 X_{k-1} + K_0)\{(X_k - \gamma_1 X_{k-1} - \gamma_0)^2 - K_3 X_{k-1} - K_2\} \\
\leq\ &\mathbf{E}_\theta[(K_1 X_0 + K_0)\{(X_1 - \gamma_1 X_0 - \gamma_0)^2 - K_3 X_0 - K_2\}] \\
=\ &\sigma^2\mathbf{E}_\theta[(K_1 X_0 + K_0)^2],
\end{aligned}$$

holds almost surely. An entirely analogous argument with lim inf instead of lim sup proves convergence to the right hand side. In a similar fashion, the denominator

in Equation (2.28) can be shown in converge to $\mathbf{E}_\theta[(K_1X_0 + K_0)^2]$ almost surely, which completes the proof. □

2.5 Simulation Study

There are two types of jumps widely used in the study of interest rate modeling: jumps with exponentially distributed jump sizes and jumps with normally distributed jump sizes. Therefore simulation studies here apply Duffie's basic affine process with these two types of jumps.

For each case, 300 sample paths are simulated. The range of time is assumed from 1 to 2500 days, with intervals equal to 1 day. This is analogous to real daily interest rate data. To our best knowledge, there is no exact simulation schemes for JCIR model, so simulation is conducted using Euler method. For parameters of continuous part of Equation (2.3), $a = 0.012$, $b = -0.6$, and $\sigma = 0.141$. For parameters of jumps in Equation (2.3), different values are set according to different types of jumps. For exponentially distributed jumps, there are only two parameters: Poisson jump intensity λ and mean of jump sizes $\frac{1}{K}$. Jump intensity is set from 0.1 to 1, with interval 0.1 and K is set from 10 to 100, with interval 10. For normally distributed jumps, there are three parameters: Poisson jump intensity λ, mean of jump size μ and standard deviation of jump size δ. Jump intensity is also set from 0.1 to 1, with interval 0.1 and standard derivation of jump size is set from 0.01 to 0.5, with interval 0.05. Mean parameter is fixed to be 0, because there is no tendency for direction of jumps.

Some numerical results are listed in Table 2.1 and Table 2.2. For different com-

binations of (λ, K) and (λ, δ), the sample means, bias, and standard deviations of 300 replicates of estimator are listed. SSE is sampling standard error, SEE is mean of standard error estimator. SEE for σ^2 is calculated via standard least square regression theory. Since there are too many combinations to be listed, only a part of results are reported in the tables, but all results are shown in Figure 2.1 and Figure 2.2. Generally speaking, estimation errors get larger when jump intensity and jump magnitude become larger. Recalling the estimation method presented in Section 2.3, effects of jumps on estimating a, b and σ are considered through calculating l_{01} and l_{02}. However, because path simulation is biased, when influence of jumps become significant, errors tend to be larger. In addition, errors for $\hat{\sigma^2}$ are larger than those for $\hat{a}$ and $\hat{b}$. This is because the three parameters are not estimated simultaneously. σ^2 is estimated based on a and b, therefore errors of a and b may influence evaluation of σ^2.

2.6 Empirical Study

In this section, parameters of JCIR model are estimated by combining our estimation method with the nonparametric method developed by Bandi and Nguyen (2003). The nonparametric method is firstly used to identify parameters related to jumps in JCIR model, and then the results are used for estimating drift and diffusion parameters of JCIR model.

Table 2.1. Estimation Results for JCIR Model with Exponential Jump Size

parameter	a	b	σ^2
true value	0.012	-0.6	0.01988
		$\lambda = 0.1$, $K = 10$	
Mean	0.01308	-0.6565	0.02458
Bias	0.001079	-0.05650	0.00470
SSE	0.001787	0.04535	0.007638
SEE	0.001732	0.04485	0.007333
		$\lambda = 0.2$, $K = 20$	
Mean	0.01297	-0.6511	0.02439
Bias	0.0009784	-0.05113	0.004511
SSE	0.001575	0.04510	0.003422
SEE	0.001538	0.04278	0.003235
		$\lambda = 0.3$, $K = 30$	
Mean	0.01301	-0.6521	0.02420
Bias	0.001010	-0.05206	0.004316
SSE	0.001450	0.044527	0.002428
SEE	0.001456	0.041730	0.002088

Table 2.2. Estimation Results for JCIR Model with Normal Jump Size

parameter	a	b	σ^2
true value	0.012	-0.6	0.01988
		$\lambda = 0.1$, $\delta = 0.05$	
Mean	0.01781	-0.8497	0.03920
Bias	0.005805	-0.2497	0.01932
SSE	0.001300	0.06147	0.003010
SEE	0.001350	0.06239	0.001636
		$\lambda = 0.2$, $\delta = 0.02$	
Mean	0.01711	-0.8311	0.03648
Bias	0.005106	-0.2311	0.01660
SSE	0.001223	0.06509	0.002342
SEE	0.001224	0.05946	0.001102
		$\lambda = 0.3$, $\delta = 0.01$	
Mean	0.01692	-0.8189	0.03499
Bias	0.004917	-0.2189	0.01510
SSE	0.001049	0.05473	0.001981
SEE	0.001182	0.05864	0.001003

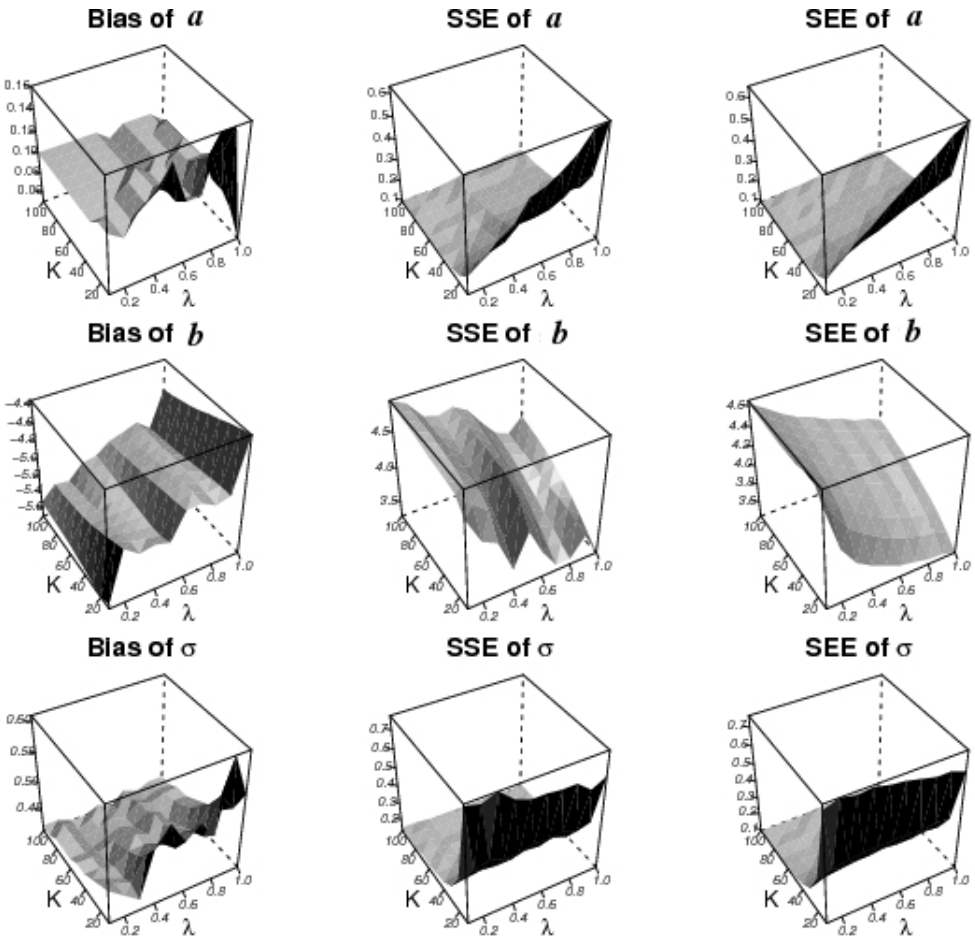

Figure 2.1. Estimation Results for JCIR Model with Exponential Jump Size

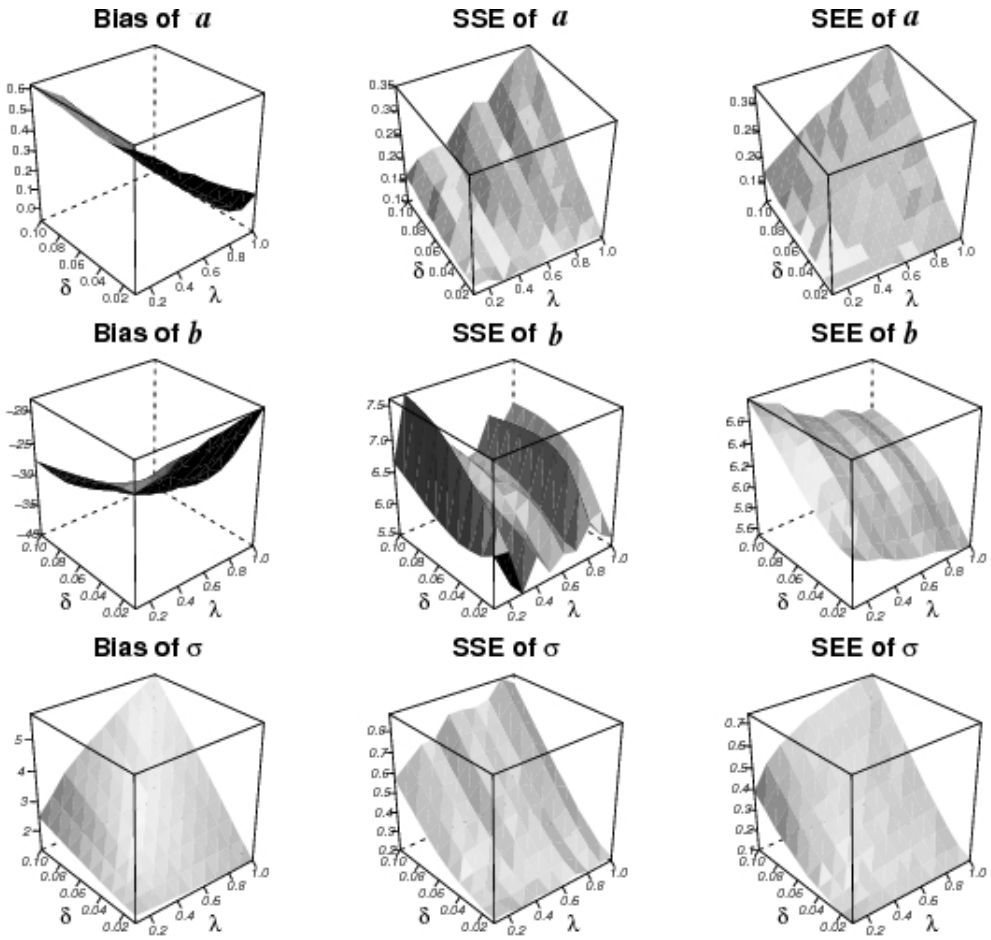

Figure 2.2. Estimation Results for JCIR Model with Normal Jump Size

2.6.1 Data Description

The real interest rate data used is Federal Fund rate (FFR) of the United States[1]. FFR is the interest rate at which private depository institutions (in usual cases, banks) lend balances (Federal Funds) at the Federal Reserve to other depository institutions, usually overnight. The daily effective Federal Funds rate is a weighted average of rates on brokered trades. The weekly data are averages of 7 calendar days ending on Wednesday of the current week. The monthly data are averages including each calendar day in the month. The Federal Funds target rate (FFTR) is determined by a meeting of the members of the Federal Open Market Committee which normally occurs eight times a year about seven weeks apart. The committee may also hold additional meetings and implement target rate changes outside of its normal schedule. The FFR and FFTR have connections as follows. Firstly, through meetings, committee determines FFTR, then Federal Reserve Bank makes open markets operations about selling or purchasing government securities to adjust the FFR to the level of FFTR.

2.6.2 Test for Jumps

To justify the necessity of applying JCIR model to FFR data, one test for existence of jumps is conducted. The test follows procedures developed by Johannes (2004). The null assumption is that the data generating process satisfies a continuous CIR model with no jumps.

1. Estimate test statistic (unconditional kurtosis $\hat{k}$) from a sample of interest rate

[1] The data is downloaded from official website of Federal Reserve Bank of New York: www.newyorkfed.org/markets/omo/dmm/fedfundsdata.cfm.

data (e.g. daily effective Federal Fund rate).

2. Estimate parameters assuming that data are generated from a continuous CIR model without jumps. This estimation is straightforward, by setting jump related parameters to be 0 in our estimation procedures in Section 2.3.2.

3. Simulate N paths of continuous CIR model without jumps, using parameters estimated in Step 2. N needs to be large, and in this test, $N = 5000$, which is larger than $N = 1000$ used by Johannes (2004). For each path, calculate test statistic (unconditional kurtosis $\hat{k}_i, i = 1, \ldots, N$).

4. Use quantiles of the empirical distribution of $\hat{k}_i$ to obtain critical values of test statistic under the null, and compare the statistic estimated from observed data with the critical values. Table 2.3 summarizes statistical tests for FFR data of

Table 2.3. Statistical Test for Existence of Jumps

Data Frequency	Daily	Weekly	Monthly
Sample Statistic	49.35	8.03	2.83
50% quantile	0.88	1.72	1.36
75% quantile	1.10	3.96	3.04
90% quantile	1.32	6.88	5.01
95% quantile	1.49	9.24	6.55
99% quantile	1.89	15.14	10.16

different frequency. For daily data, there is a strong rejection for null assumption of CIR model without jumps. As data frequency becomes lower, the test is less significant. This is quite natural, because low frequency data tends to "smooth" sample path and it is more difficult to identify jumps. Therefore in empirical study, daily effective FFR data is applied.

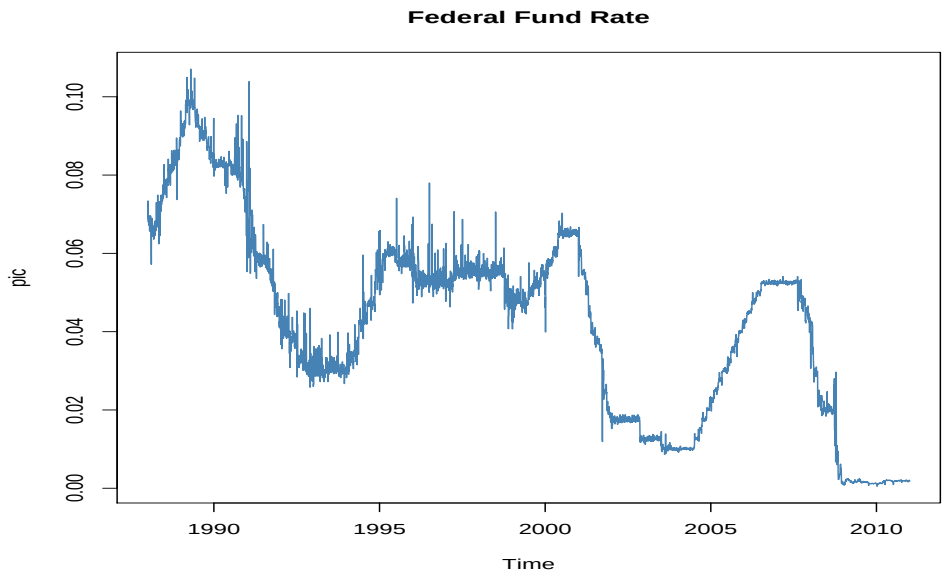

Figure 2.3. Federal Fund Rate: Daily Data

2.6.3 Estimation JCIR Model

Before estimating drift and diffusion parameters for JCIR model, parameters for jumps should be identified first. According to Gikhman and Skorohod (1972), the JCIR model is supposed to have the formula:

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t + J_t, \tag{2.46}$$

where $\mu(X_t) = a + bX_t$, $\sigma(X_t) = \sigma X_t$, W_t is a standard Brownian motion, J_t is a compound Poisson process with jump intensity $\lambda(X_t) = \lambda X_t$ and jump magnitude $c(X_t, Y)$. Y is a random variable satisfying certain distribution law. The infinitesimal conditional moments of the changes in solutions to Equation (2.46) can be

written in terms of the functions $\mu(\cdot), \sigma(\cdot), c(\cdot, Y)$ and $\lambda(\cdot)$. In particular

$$\mathbf{M}^1(r) = \lim_{\Delta \to 0} \frac{1}{\Delta} \mathbf{E}[X_{t+\Delta} - X_t \mid X_t = r] = \mu(r). \tag{2.47}$$

$$\mathbf{M}^2(r) = \lim_{\Delta \to 0} \frac{1}{\Delta} \mathbf{E}[(X_{t+\Delta} - X_t)^2 \mid X_t = r] = \sigma^2(r) + \lambda(r)\mathbf{E}_Y[c^2(r, Y)]. \tag{2.48}$$

$$\mathbf{M}^k(r) = \lim_{\Delta \to 0} \frac{1}{\Delta} \mathbf{E}[(X_{t+\Delta} - X_t)^k \mid X_t = r] = \lambda(r)\mathbf{E}_Y[c^k(r, Y)], \forall\ k > 2. \tag{2.49}$$

Under some mild conditions like those discussed in Bandi and Nguyen (2003) and Johannes (2004), $\mathbf{M}^k(r)$ can be estimated consistently as follows:

$$\widehat{\mathbf{M}}^1_{n,T}(r) = \frac{\frac{1}{h_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}\Big(\frac{X_{i\Delta_{n,T}}}{h_{n,T}}\Big)[X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}]}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}\Big(\frac{X_{i\Delta_{n,T}} - r}{h_{n,T}}\Big)}. \tag{2.50}$$

$$\widehat{\mathbf{M}}^2_{n,T}(r) = \frac{\frac{1}{h_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}\Big(\frac{X_{i\Delta_{n,T}}}{h_{n,T}}\Big)[X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}]^2}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}\Big(\frac{X_{i\Delta_{n,T}} - r}{h_{n,T}}\Big)}. \tag{2.51}$$

$$\widehat{\mathbf{M}}^k_{n,T}(r) = \frac{\frac{1}{h_{n,T}} \sum_{i=1}^{n-1} \mathbf{K}\Big(\frac{X_{i\Delta_{n,T}}}{h_{n,T}}\Big)[X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}]^k}{\frac{\Delta_{n,T}}{h_{n,T}} \sum_{i=1}^{n} \mathbf{K}\Big(\frac{X_{i\Delta_{n,T}} - r}{h_{n,T}}\Big)}. \tag{2.52}$$

According to the basic rules of nonparametric regression, $\widehat{\mathbf{M}}^k$ is estimated exactly by classical Nadaraya-Watson kernel estimation, and $h_{n,T}$ is bandwidth. In basic nonparametric regression, suppose that (X, Y) is a pair of random variables which take values in $\mathbb{R} \times \mathbb{R}$, the target is to estimate $\Psi(\cdot)$ in the nonparametric regression model $Y = \Psi(X) + \epsilon$, given some data: $(x_1, y_1), \ldots, (x_n, y_n)$ of i.i.d. realizations of (X, Y).

The Nadaraya-Watson kernel estimation mentioned above has been introduced and studied by Watson (1964) and Nadaraya (1964). The method provides a

smoothing estimation of the regression function $\Psi(\cdot)$ by local weighted averaging of the y_i values,

$$\Psi^n(x) = \frac{\sum_{i=1}^{n} \mathbf{K}\Big(\frac{x-x_i}{h}\Big) y_i}{\sum_{i=1}^{n} \mathbf{K}\Big(\frac{x-x_i}{h}\Big)}, \tag{2.53}$$

where $\mathbf{K}$ is the kernel, i.e., $\mathbf{K} : \mathbb{R} \to \mathbb{R}$ is a bounded, integrable function. h is the bandwidth parameter, controlling the smoothness of the estimate $\Psi^n(\cdot)$.

Generally, h is selected through a cross validation procedure. In cross-validation, the data are divided into subsets, then the model is successively fitted by omitting each subset in turn. The fitted model is used to 'predict' the response for the left-out subset. Trying this procedure for different values of bandwidth parameter h will suggest one that minimizes the cross-validation estimation of the mean-squared error. Other approaches have been proposed including local choice for bandwidth, see Schucany (2004) for an overview.

In JCIR model, $[X_{(i+1)\Delta_{n,T}} - X_{i\Delta_{n,T}}]^k$ is regarded as y_i in Equation (2.53), and $X_{i\Delta_{n,T}}$ is regarded as x_i in Equation (2.53). Therefore, $\mathbf{M}^k$ can be calculated according to procedure of nonparametric regression. However, choosing optimal bandwidth is still an open problem. Those existing methods, such as cross validation, are generally designed for standard regression models, and may not be the proper choices. Some further discussions about this topic are in Bandi and Phillips (2003) and Johannes (2004). Here the same bandwidth (0.015) in Bandi and Nguyen (2003) is used.

In empirical study, the JCIR model with constant jump intensity λ and normally distributed jump magnitudes is used. As in simulation study, the mean of jump

sizes is 0, and standard deviation of jump sizes is δ. Therefore according to basic statistics,

$$\mathbf{E}_Y[y^{2r}] = \delta^{2r} \prod_{n=1}^{r} (2r-1), \tag{2.54}$$

$$\mathbf{E}_Y[y^{2r-1}] = 0. \tag{2.55}$$

Substituting these moments into formulas of $\mathbf{M}^k(r)$ leads results for parameters

$$(\delta^2)_{n,T} = \frac{1}{n} \sum_{i=1}^{n} \frac{\widehat{\mathbf{M}}^6(X_{i\Delta_{n,T}})}{5\widehat{\mathbf{M}}^4_{n,T}(X_{i\Delta_{n,T}})}, \tag{2.56}$$

$$(\lambda)^2_{n,T} = \frac{1}{n} \sum_{i=1}^{n} \frac{(\widehat{\mathbf{M}}^4/3)^6}{(\widehat{\mathbf{M}}^6/15)^4}. \tag{2.57}$$

The estimation results are $\lambda = 0.026$, and $\delta = 0.0090$. For average, there is one jump around six weeks. Results for other parameters are as follows. Table 2.4 shows that b and σ^2 can be estimated with small standard errors. However, standard error for a is large. The reason is that a is related to mean reverting level, which is a trend parameter, whose estimation accuracy depends on the range of trading time, whereas the accuracy of estimations for b and σ^2 depends on the number of observations.

Table 2.4. Estimation Results for JCIR Model

parameter	a	b	σ^2
Estimated Value	1.1136e-05	-0.002650	6.3882e-05
Standard Error	3.2727e-05	0.0008453	5.4730e-06

2.7 Summary and Remarks

In this chapter, a new estimation method for continuous state branching process with immigration (hereafter, CBI) is proposed, which is based on weighted conditional least square estimators. The method can be regarded as an extension of estimation approach for discrete time and discrete state space branching process with immigration. Because JCIR model is a special form of CBI process, this method is ready to be used for parameter estimation of JCIR model. The strength and originality of this method lies in the fact that it avoids computationally expensive numerical integration in many existing estimation methods for JCIR model. Proofs for consistency of the developed method are provided. Simulation and empirical studies show that it is easy to combine this method with existing nonparametric method to get accurate estimations for parameters in JCIR model, and the calculations are fast and simple.

Chapter 3

Estimation of Jump-Diffusion Process via High Frequency Data

3.1 Motivation

As mentioned in the introduction, jumps have significant impact in many aspects of financial management. With the emergence of available high frequency data, many researchers turn to estimate parameters in jump-diffusion process via such data, instead of traditional low frequency data. Compared with low frequency data, such as daily data, high frequency data has several distinguishable features. Firstly, the intraday observations occur at random trading time. Secondly, the trading price is contaminated by market microstructure noise. Thirdly, high frequency data has no data aggregation effect. Therefore, it contains more information concerning market microstructure.

Because of market microstructure noise, analysis methods for high frequency data are very different from those for low frequency data. Generally speaking, modeling methods for high frequency data can be divided into two main categories: nonparametric methods and parametric methods.

Nonparametric methods are based on integrated volatility, realized volatility and

quadratic variation. The underlying theory for those methods assumes that log prices follow a continuous-time semi-martingale process so that the sum of successively finer sampled high-frequency squared returns converges to the quadratic variation of the price process. Prior works include Andersen and Bollerslev (1998), Barndorff-Nielsen and Shephard (2002). However, these estimators might be inconsistent. Aït-Sahalia et al. (2005a) and Bandi and Russell (2006) show that market microstructure noise may introduce a bias in the realized volatility estimation using high frequency data. As a sequel, many researchers, such as Aït-Sahalia et al. (2005b), Bandi and Russell (2008) and Barndorff-Nielsen et al. (2008), develop various methods based on optimal sampling, subsampling and realized kernels, which may overcome this difficulty. The five-minute sample frequency used in many papers is a tradeoff between estimation accuracy and market microstructure noise. Besides methods for calculating realized volatility, there are also tests for identifying jumps using high frequency data, developed by Aït-Sahalia and Jacod (2009, 2011), Jiang and Oomen (2008) and Lee and Mykland (2008).

Although there is a large volume of literature focusing on nonparametric methods, relatively few works focus on parametric modeling of high frequency data. Parametric Methods treat high frequency data, which contains transaction time together with trading price and trading volume, as a Marked Point Process in which a point process describes the transaction times and marks represent price and volume observed at trading time. See Martin (2006) for a comprehensive introduction to market point process. Based on different statistical foundations, two kinds of parametric models have been developed: time-series-based models and

filtering-based models.

The time-series-based models are originally developed by Engle (2000). In this kind of models, high frequency data is treated as an irregularly-spaced time series and modeled via Autoregressive Conditional Duration (ACD) model proposed by Engle and Russell (1998). There are many researchers using this model to analyze high frequency data, such as McCulloch and Tsay (2001), Grammig and Wellner (2002), Ghysels et al. (2004), Zhang et al. (2009). Refer to Engle and Russell (2005) and Pacurar (2008) for reviews about ACD model and its extensions.

Although ACD model is widely used in financial econometrics, it has some limitations. For example, market microstructure noise is not considered in this kind of models. Black (1986) and Stoll (2000) address noise as an important factor when considering the market microstructure. Although the influence on low frequency data is not so significant due to its short-term impact, noise plays an essential role in high frequency data models. The filtering-based models are proposed by Zeng (2003) and Duan (2009) by integrating the noise. In Zeng's model, tick size effect, clustering noise and non-clustering noise are explicitly modeled. The data is treated as a collection of counting process points, a special case of MPP observations. Within this framework, the model can be formulated by some filtering problem, and it is applied to stochastic volatility estimation by Zeng (2004), Bayesian model selection by Kouritzin and Zeng (2005) as well as option hedging by Lee and Zeng (2010). It is further extended by Hu et al. (2010). In Duan's model, noise is implicitly modeled by introducing a Gaussian white noise. The model is also considered as a filtering problem, and calibrated via a particle filter-

ing based maximum likelihood method. Duan and Fulop (2007) also use a similar model to study jumps in high frequency data. Along with Johannes et al. (2009), these papers are prior studies about jump-diffusion process via high frequency data.

The purpose of this chapter is to develop a general estimation framework for Zeng's high frequency data model (Zeng, 2003)[1], and to make the method more feasible and efficient. The method proposed here is based on particle Markov chain Monte Carlo method (Andrieu et al., 2010).

3.2 Model Specification

Zeng's model (2003) is based on the intuition that trading price should arise from an intrinsic price process in combination with market noise from trading activities. According to market microstructure theory, intraday price movements suffer from three important kinds of noise: discrete, clustering, and non-clustering noise. Discrete noise exists because intraday prices move discretely, that is, tick by tick. Clustering noise exits because prices gather more on integer and half ticks, instead of distributing evenly on all ticks. Non-clustering noise exits due to change of distribution of prices and outliers in prices formulation.

A high frequency data model should reflect these three kinds of noise. Assume $\{Y_t, t \geq 1\}$ is observed intraday price, and $\{S_t, t \geq 1\}$ is unobserved intrinsic price.

[1] In early research papers, the definitions of ultra-high frequency data (not aggregated, transaction-by-transaction data with random trading time) and high frequency data (aggregated, equally-spaced data) are different, see Engle (2000). All of the nonparametric methods mentioned above use high frequency data, including parametric method of Duan (2007). However, models of Engle (2000) and Zeng (2003) is designed for ultra-high frequency data, which makes them different from other models. In the rest of this thesis, for coherence in narrative, only the phrase "high frequency data" is used. But readers should notice the difference.

First of all, $\{S_t, t \geq 1\}$ should be modeled by certain stochastic process. Then some function which reflects all three kinds of noise should be used to map from intrinsic price S_t to intraday trading price Y_t. The mapping procedure consists of three steps.

Step 1. Incorporate discrete noise by rounding off S_t to its closest tick: Round$[S_t, \frac{1}{M}]$, 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 $\$\frac{1}{8}$ in June 24, 1997 and then further adjusted to \$0.01 beginning from January 29, 2001.

Step 2. Incorporate non-clustering noise by adding V: $Y_t' = \text{Round}[S_t + V, \frac{1}{M}]$, where V is a random variable satisfying doubly geometric distribution with parameter ρ.

$$P(V = v) = \begin{cases} (1-\rho) & \text{if } v = 0, \\ \frac{1}{2}(1-\rho)\rho^{M|v|} & \text{if } v = \pm\frac{1}{M}, \pm\frac{2}{M}, \ldots. \end{cases} \tag{3.1}$$

Step 3. Incorporate clustering noise by biasing Y_t'. The biasing function $b(\cdot)$ moves Y_t' to some close ticks according to certain probability defined by parameters α, β, γ. The construction of biasing function is related to tick size. For example, when tick size is $\$\frac{1}{8}$, $b(\cdot)$ is constructed by following rules: if the fractional part of Y_t' is an even eighth, then Y_t' stays on Y_t' with probability one; if the fractional part of Y_t' is an odd eighth, then Y_t' stays on Y_t' with probability $1-\alpha-\beta-\gamma$, Y_t' moves to the closest odd quarter with probability α, moves to the closet half with probability β, and moves to the closed integer with γ. This modeling is based on the clustering phenomenon: integers and halves are most likely trading prices and

have about the same frequencies; odd quarters are the second most likely and have about the same frequencies; odd eighthes are least likely and have about the same frequencies. The three parameters α, β, γ can be estimated via relative frequency methods. The details are illustrated in Zeng (2003).

The conditional likelihood function $P(Y_t \mid S_t)$, which is similar to the one in Kouritzin and Zeng (2005), is listed in Table 3.1, where

$$D = M * |Y_t - \text{Round}[S_t, \frac{1}{M}]|,$$

and

$$R = \begin{cases} 3 & \text{if the fractional part of } Y_t \text{ is } \frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}, \\ 2 & \text{if the fractional part of } Y_t \text{ is } \frac{1}{4}, \frac{3}{4}, \\ 1 & \text{if the fractional part of } Y_t \text{ is } \frac{1}{2}, \\ 0 & \text{if } Y_t \text{ is an integer.} \end{cases}$$

In general, the intrinsic price S_t could be modeled by any stochastic process. It is natural to apply geometric Brownian motion (GBM), jump-diffusion process and stochastic volatility process in the model. In this chapter, GBM and a simple jump-diffusion process are used. GBM is used firstly to study the method in details, then results are extended to the jump-diffusion case.

The stochastic differential equation for Geometric Brownian Motion is as follows:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t, \tag{3.2}$$

where μ is the drift parameter, σ is the positive diffusion parameter, and W_t is a standard Brownian motion.

Table 3.1. Likelihood Function for High Frequency Data Model

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	$(1-\rho)*(1+\beta*\rho*(1+\rho^2))$
1	1	$0.5*(1-\rho)*(\rho+\beta*(2+2*\rho^2+\rho^4))$
1	2	$0.5*(1-\rho)*\rho*(\rho+\beta*(2+\rho^2+\rho^4))$
1	3	$0.5*(1-\rho)*(\rho^3+\beta*(2+\rho^2+\rho^4+\rho^6)$
1	≥ 4	$0.5*(1-\rho)*\rho^{D-3}*(\rho^3+\beta*(1+\rho^2+\rho^4+\rho^6))$
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	$(1-\rho)*(1-\alpha-\beta-\gamma)$
3	≥ 1	$0.5*(1-\alpha-\beta-\gamma)*(1-\rho)*\rho^D$

The stochastic differential equation for Merton's jump-diffusion process (Merton, 1976) is as follows:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t + J_t dN_t, \tag{3.3}$$

where μ, σ and W_t are of the same meanings as those in Equation (3.2), 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. These two processes yield the same conditional likelihood formula in Table 3.1. The target is to estimate all parameters in both stochastic processes together with ρ in Equation (3.1).

3.3 Estimation Method

Traditionally, estimation methods for parametric model are built upon likelihood function. In Engle's model, trading time and trading price are supposed to be

directly observed. Therefore, the estimation via maximum likelihood method is workable. However, the estimation is more difficult and more computationally expensive in Zeng's model. One reason is that the intrinsic prices are not observed directly due to the existence of market microstructure noise, which makes the model to be a state-space model. Another reason is the high randomness of market microstructure noise makes the state-space model nonlinear and non-Gaussian. Because of these challenges in model estimation, filtering based models are usually difficult to incorporate complicated noise structures which reflect stylized facts of intrinsic price process, such as leverage, jump and stochastic volatility. Therefore an efficient estimation method should be developed in such case.

In this chapter, an alternative method will be derived, which is based on particle Markov Chain Monte Carlo (PMCMC) method. This method combines particle filtering with Markov Chain Monte Carlo (MCMC). To be precise, MCMC is used to propose new values for parameters in the model, then particle filtering is used to calculate values of marginal likelihood functions in the state-space model based on those proposed parameters.

The idea of PMCMC method has been used in economics by several scholars for estimating 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 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. Secondly, approximation to likelihood functions by particle filtering is proved to be unbiased by Pitt et al. (2010). Thirdly, the Markov Chain Monte Carlo step of the method approximates to the posterior distributions of parameters, containing more information to conduct statistical inference than just point estimations. Moreover, due to the nature of simulation, parallel programming is able to be used in calculation, which makes this method more efficient when dealing with large data set in practice.

The state-space model consists of two components: one unobserved component, which is intrinsic price process $\{S_t; t \geq 1\}$; one observable component, which is trading price $\{Y_t; t \geq 1\}$. $\{S_t; t \geq 1\}$ is a Markov process, characterized by its initial density $S_1 \sim \mu_\theta(\cdot)$ and transition probability density

$$S_{t+1}|(S_t = S) \sim f_\theta(\cdot|S), \tag{3.4}$$

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

$$Y_t|(S_1, \ldots, S_t = S, \ldots, S_m) \sim g_\theta(\cdot|S). \tag{3.5}$$

The most important issues for a state-space model are filtering and parameter

estimation. The filtering equations for the state-space model are as follows:

$$p(S_t|Y_{1:t-1};\theta) = \int p(S_t|S_{t-1};\theta)p(S_{t-1}|Y_{1:t-1};\theta)dS_{t-1}, \tag{3.6}$$

$$p(S_t|Y_{1:t};\theta) = \frac{p(Y_t|S_t;\theta)p(S_t|Y_{1:t-1};\theta)}{p(Y_t|Y_{1:t-1};\theta)}, \tag{3.7}$$

$$p(Y_t|Y_{1:t-1};\theta) = \int p(Y_t|S_t;\theta)p(S_t|Y_{1:t-1};\theta)dS_t. \tag{3.8}$$

Equations (3.6) - (3.8) enable us to filter for a given θ and evaluate marginal likelihood of observation $\{Y_{1:t}\}$. The likelihood function is as follows:

$$p(Y_{1:t}|\theta) = p(Y_1|\theta)\prod_{k=2}^{t} p(Y_k|Y_{1:k-1};\theta). \tag{3.9}$$

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):

$$\begin{aligned} p(Y_{1:t}|\theta) &= \int p_\theta(S_{1:t}, Y_{1:t})dS_{1:t} \\ &= \int \mu_\theta(S_1)\prod_{k=2}^{t} f_\theta(S_k|S_{k-1})\prod_{k=1}^{t} g_\theta(Y_k|S_k)dS_{1:t}. \end{aligned} \tag{3.10}$$

When both observation and state transition equations are linear and Gaussian, the likelihood function can be evaluated analytically by Kalman filtering (see Welch and Bishop, 1995). However, since the model in this chapter is a nonlinear, non-Gaussian and high dimensional[2] state-space model, integration in Equation (3.10) needs to be calculated numerically via Monte Carlo method, which means that $S_{1:t}$ need to be sampled from distribution $\pi_t(S_{1:t})$, and the likelihood value can be

[2] The number of dimensions is equal to the number of data points in high frequency data set.

calculated by

$$\mathbf{E}_{\pi_t}(p(Y_{1:t}|\theta)) = \frac{1}{N}\sum_{i=1}^{N} p_\theta(S_{1:t}, Y_{1:t}), \tag{3.11}$$

where $\{S^i_{1:t}, i = 1, \ldots, N\}$ 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. According to Doucet and Johansen (2011), intuitive explanation of particle filtering algorithm is as follows.

Initially, consider a sequence of probability distributions $\pi_{t(t\geq 1)}$ defined on a sequence of measurable spaces $(E_t, \mathcal{F}_t)_{t\geq 1}$, where $E_1 = E, \mathcal{F}_1 = \mathcal{F}$ and $E_t = E_{t-1}\times E$, $\mathcal{F}_t = \mathcal{F}_{t-1} \times \mathcal{F}$. 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.

$$\pi_t(S_{1:t}) = \frac{\gamma_t(S_{1:t})}{Z_t}, \tag{3.12}$$

$$Z_t = \int \gamma_t(S_{1:t})dS_{1:t}. \tag{3.13}$$

The purpose is to sample N independent random variables, $S^i_{1:t} \sim \pi_t(S_{1:t})$ 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

$$\pi_t(S_{1:t}) > 0 \Rightarrow q_t(S_{1:t}) > 0. \tag{3.14}$$

In this case, Equation (3.12) - (3.13) are rewritten as follows:

$$\pi_t(S_{1:t}) = \frac{w_t(S_{1:t})q_t(S_{1:t})}{Z_t}, \tag{3.15}$$

$$Z_t = \int w_t(S_{1:t})q_t(S_{1:t})dS_{1:t}, \tag{3.16}$$

where $w_t(S_{1:t})$ is the unnormalized weight

$$w_t(S_{1:t}) = \frac{\gamma_t(S_{1:t})}{q_t(S_{1:t})}, \tag{3.17}$$

$$W_t^i = \frac{w_t(S_{1:t}^i)}{\sum_{i=1}^{N} w_t(S_{1:t}^i)}. \tag{3.18}$$

Importance density $q_t(S_{1:t})$ is often carefully selected from some special distributions, from which it is easy to draw samples. Particle filtering (sequential importance sampling) method chooses a special importance density

$$q_t(S_{1:t}) = q_{t-1}(S_{1:t-1})q_t(S_t|S_{1:t-1}) = q_1(S_1)\prod_{k=2}^{t} q_k(S_k|S_{1:k-1}). \tag{3.19}$$

In the algorithm, $S_{1:t}^i$ 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 Equation (3.17):

$$\begin{aligned} w_t(S_{1:t}) &= \frac{\gamma_t(S_{1:t})}{q_t(S_{1:t})} = \frac{\gamma_{t-1}(S_{1:t-1})}{q_{t-1}(S_{1:t-1})}\frac{\gamma_t(S_{1:t})}{\gamma_{t-1}q_t(S_t|S_{1:t-1})} \\ &= w_{t-1}(S_{1:t-1})\alpha_t(S_{1:t}) = w_1(S_1)\prod_{k=2}^{t}\alpha_k(S_{1:k}), \end{aligned} \tag{3.20}$$

where α_k is called as incremental importance weight function, which is given by

$$\alpha_k(S_{1:k}) = \frac{\gamma_k(S_{1:k})}{\gamma_{k-1}(S_{1:k-1})q_k(S_k|S_{1:k-1})}. \tag{3.21}$$

In the high frequency data model, let $\pi(S_{1:t}) = p(S_{1:t}|Y_{1:t})$, $\gamma(S_{1:t}) = p(S_{1:t}, Y_{1:t})$, $Z_t = p(Y_{1:t})$, then the only thing left is to select an importance distribution $q_t(S_t|S_{1:t-1})$. In practice, $q_t(S_t|S_{t-1}) = q(S_t|Y_t, S_{t-1})$. Particularly, $q_t(S_t|S_{1:t-1})$ is chosen as $f_\theta(S_t|S_{t-1})$ by Gordon (1993), then $\alpha(S_{1:t}) = g_\theta(Y_t|S_t)$ in this case due to the Markov property of unobserved process S_t. The 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 $S_1^i, i = 1, \ldots, N$ from $q_1(S_1)$.
  Compute unnormalized weights $w_1(S_1^i) = g_\theta(Y_1|S_1^i)$.
  Compute normalized weights $W_1^i \propto w_1(S_1^i)$.
  Resample $\{S_1^i\}$ via weight $\{W_1^i\}$ to obtain N particles, denoted as $\{S_1^i\}$.
for iteration $t \geq 2$ **do**
  Sample $S_t^i \sim f_\theta(S_t|S_{t-1}^i)$.
  Compute unnormalized weights $w_t(S_t) = g_\theta(Y_t|S_t^i)$.
  Compute normalized weights $W_t^i \propto w_t(S_{1:t}^i)$.
  Resample $\{S_t^i\}$ via weight $\{W_t^i\}$ to obtain N particles, denoted as $\{S_t^i\}$.
end for

Algorithm 1 allows us to estimate sequentially the marginal likelihood function by

$$\hat{p}_\theta(Y_{1:t}) = \hat{p}_\theta(Y_1) \prod_{t=2}^{T} \hat{p}_\theta(Y_t|Y_{1:t-1}), \tag{3.22}$$

where

$$\hat{p}_\theta(Y_t|Y_{1:t-1}) = \frac{1}{N} \sum_{k=1}^{N} w_t(S_{1:t})^i. \tag{3.23}$$

In the particle filtering algorithm, when parameter vector θ is fixed, $\{S_{1:t}\}^i$ is

sampled from $p(S_{1:t}|Y_{1:t})$. 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 $p(\theta, S_{1:t}|Y_{1:t})$ 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 details. Refer to Bolstad (2010) for some comprehensive introduction.

By standard decomposition, $p(\theta, S_{1:t}|Y_{1:t}) = p(\theta|Y_{1:t})p_\theta(S_{1:t}|Y_{1:t})$. Consequently it is natural to use a proposal density for an M-H upgrade in the form of

$$q(\theta', S'_{1:t}|\theta, S_{1:t}) = q(\theta'|\theta)p_{\theta'}(S'_{1:t}|Y_{1:t}). \tag{3.24}$$

The M-H acceptance rate α is given by

$$\alpha = \frac{p(\theta', S'_{1:t}|Y_{1:t})}{p(\theta, S_{1:t}|Y_{1:t})} = \frac{p_{\theta'}(Y_{1:t})}{p_\theta(Y_{1:t})}\frac{q(\theta|\theta')}{q(\theta'|\theta)}, \tag{3.25}$$

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 $\theta' \sim q(\cdot \mid \theta_{i-1})$.
  Run a particle filtering algorithm, obtaining $\hat{p}_{\theta_0}(Y_{1:t})$, denote as estimation of marginal likelihood value.
  With probability $\alpha = \min(1, \frac{p(\theta', S'_{1:t}|Y_{1:t})}{p(\theta, S_{1:t}|Y_{1:t})} = \frac{p_{\theta'}(Y_{1:t})}{p_\theta(Y_{1:t})}\frac{q(\theta|\theta')}{q(\theta'|\theta)})$.
  Set $\theta_i = \theta'$, and $\hat{p}_{\theta_i}(Y_{1:t}) = \hat{p}_{\theta'}(Y_{1:t})$.
end for

Algorithm 2 is sufficient for most state-space high frequency data models. In the particle filtering part, the time intervals for particle generation steps are equal. However, one characteristic of Zeng's high frequency data model is that the sequence of trading times has irregular intervals. That's why trading times are modeled as a Poisson process. The generalization of particle filtering algorithm in such situation is straightforward. The only thing needed 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.

The estimation approach 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, see 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, Douc and Moulines (2008) and Del Moral et al. (2011) 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.

3.4 Simulation Study for Geometric Brownian Motion

3.4.1 Simulation

In simulation experiment, parameters are set as follows: $M = 8$, $\rho = 0.2$, $\alpha = 0.225$, $\beta = 0.066$, $\gamma = 0.3$, $\mu = 4.4\text{e}^{-8}$, $\sigma = 1.2\text{e}^{-4}$. The simulated data is shown in Figure 3.1. The figure shows clearly that when market microstructure noise is considered, the observed trading prices present different characteristics compared with the underlying intrinsic prices.

Due to the irregular intervals among trading times, a Poisson process with mean λ' is needed to model the trading time process. Then GBM is simulated according to the trading times. λ' is set to be 0.067, which means there is one trading every 16.67 seconds in average. Since there are totally 6.5 hours in one trading day, sample size in one day should be around 1500. For the sake of simplicity, 2000 simulated trading data are used for estimation. Particle number is 4000, and the length of Markov chain is 100000. The last 50000 data from the chain is used for analysis.

3.4.2 Variance Reduction Effect of Particle Filtering Method

In Algorithm 2, the likelihood value $p_\theta(Y_{1:t})$ is also the normalize constant Z_t in Equation (3.12), then according to standard Monte Carlo integration theory, $p_\theta(Y_t|Y_{1:t-1})$ is calculated at every time t via particle filtering by Equation (3.23). As is known, directly using Monte Carlo integration method suffers from large variance. To make it clear, take GBM for example. Since the variance of GBM

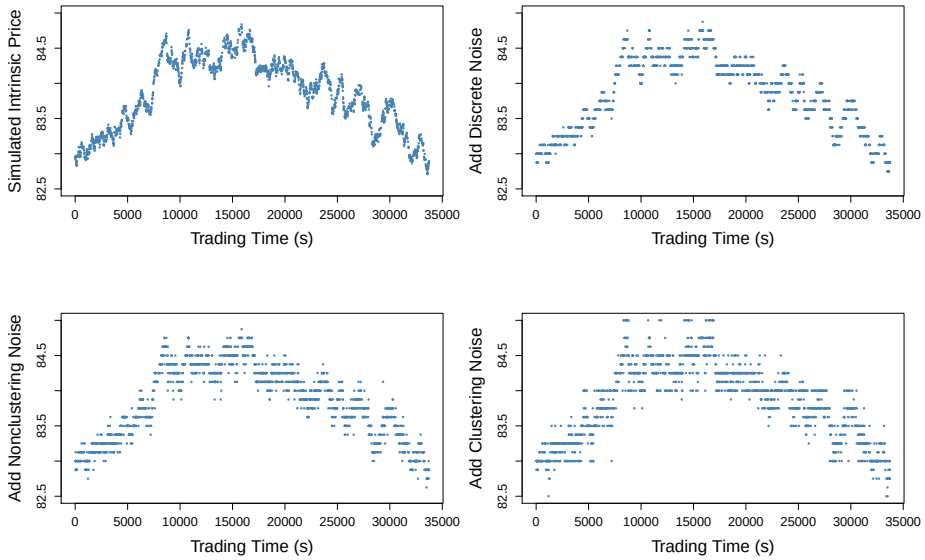

Figure 3.1. Simulated Trading Price for Geometric Brownian Motion

increases with time, Figure 3.2 shows that after very short time, the simulated paths will deviate from the true one far away. To maintain a given degree of accuracy, the number of paths needed may be large and may increase with time. However, when particle filtering with certain kind of resampling scheme is applied, the paths are able to be constrained near the true one. The basic idea of resampling methods 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 $p_\theta(Y_{1:t})$.

3.4.3 Choice of Resampling Schemes

There are four common resampling methods in particle filtering: Multinomial Resampling, Residual Resampling, Stratified Resampling and Systematic Resampling. Douc et al. (2005) gives a brief introduction of these four schemes. Multi-

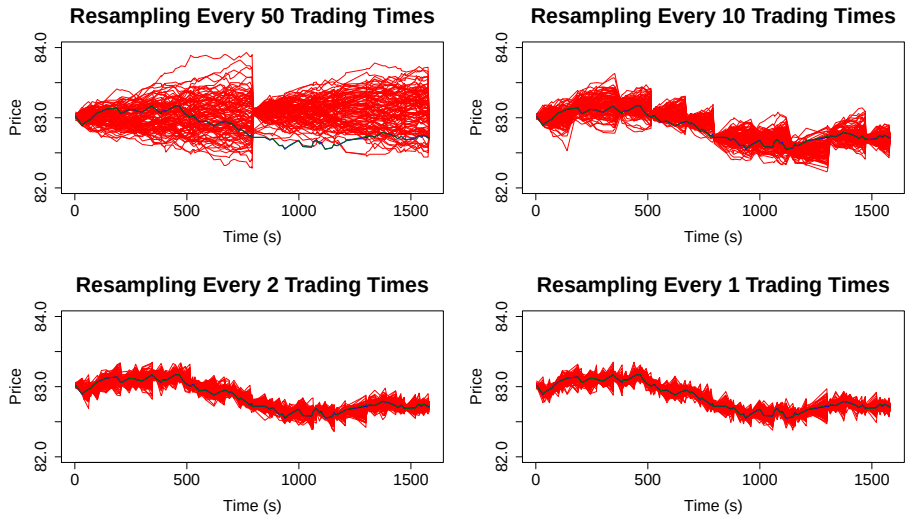

Figure 3.2. Effects of Particle Filtering in Variance Reduction

nomial resampling is a classical and stable algorithm, and is used in particle filtering by Gordon (1993). With pre-sorted weights, the complex of this algorithm can be $O(N)$. Residual Resampling is used in particle filtering by Liu and Chen (1998). This method comprises two main steps: in the first step, particles are replicated deterministically according to their weights and total number of particles; in the second step, particles are randomly sampled via Multinomial Resampling method with processed weights, which is calculated according to certain formulas. Stratified resampling is based on ideas used in survey sampling and conducted by pre-partitioning the (0, 1] interval into N disjoint sets, and is used in particle filtering by Carpenter et al. (1999). Systematic resampling takes one step further by deterministically linking all the variables drawn from the N pre-partitioned sub-intervals in stratified resampling. According to Doucet and Johansen (2011), among these four resampling schemes mentioned, the most widely used one is systematic resampling due to its easy implementation and better performances. The pseudo codes of these four resampling schemes are listed in Appendix A.

Except for the four resampling methods mentioned above, a novel particle filtering algorithm is proposed by Xiong and Zeng (2011), which uses branching resampling scheme. The algorithm is different from classical particle filtering in three aspects. First of all, in classical particle filtering (Algorithm 1), the time intervals between every two steps is equal to time intervals of observed trading time. However, in Xiong's algorithm, the time intervals are determined by users. Secondly, in classical particle filtering (Algorithm 1), the number of particles in the system is unchanged during the calculation. However, in branching particle filtering, the number of particles varies in different steps. Thirdly, Xiong and Zeng (2011) prove that this method has uniform convergence property when applied upon Zeng's high frequency data model, which is better than mean square error convergence property of classical particle filtering. The details about this algorithm is described in Xiong and Zeng (2011). Pseudo code for branching resampling scheme is listed in Appendix A.

A comparison about numerical accuracy of likelihood function calculation is made by using these five resamlping schemes. Calculations of likelihood function value are repeated 1000 times, with 2000 simulated data and different particle numbers. The true parameter values are set to be the same as those in Section 3.4.1. Table 3.2 shows that when the number of particles is equal to the number of data points, systematic resampling yields smallest variance, which is important for implementation of particle filtering. Since particle filtering algorithm is computationally expensive, it is a customary in practice to set the number of particles equal to the number of data points. It is a tradeoff between accuracy and program exe-

cuting time. Note that although multinomial resampling needs least CPU time, the variance is much larger. In further simulation study, systematic resampling is applied. Branching resampling scheme is also used because the algorithm is novel, and there are no numerical study upon it.

Table 3.2. Comparison of Different Resampling Schemes

	Multinomial	Residual	Stratified	Systematic	Branching
			Particle Number = 1000		
mean	-2705.184	-2704.981	-2704.913	-2704.916	-2705.187
sd	1.789844	1.734905	1.644725	1.554531	1.9928612
time (s)	290	345	307	288	399
			Particle Number = 2000		
mean	-2704.344	-2704.292	-2704.253	-2704.265	-2704.443
sd	1.227276	1.153464	1.163640	1.128178	1.2983432
time (s)	569	649	650	612	751
			Particle Number = 3000		
mean	-2704.117	-2704.104	-2704.071	-2704.057	-2704.137
sd	1.0430071	0.9317736	0.9635022	0.9303675	0.9787504
time (s)	881	967	1072	1017	1107
			Particle Number = 4000		
mean	-2703.994	-2703.950	-2703.931	-2703.976	-2703.968
sd	0.8663533	0.7858741	0.8288907	0.8317786	0.8673299
time (s)	1214	1112	1563	1490	1481
			Particle Number = 5000		
mean	-2703.904	-2703.876	-2703.865	-2703.891	-2703.931
sd	0.7866920	0.7296702	0.7267766	0.7178622	0.7698107
time (s)	1574	1392	2121	2031	1803

∗ mean is sample mean.
∗ sd is standard sample deviation.
∗ time is CPU time of programs.

Although resampling step in particle filtering is an unbiased approximation to the distribution $\pi_t(S_{1:t})$, it does introduce errors. As a result, resampling should only be used when necessary and schemes which lead to less variance are preferable. Consequently, Doucet and Johansen (2011) suggest that in practice it is more

sensible to resample only when the variance of the unnormalized weights is superior to a pre-specified threshold. The suggestion is equal to examining the variability of the weights using Effective Sample Size (ESS) criterion suggested by Liu (2001).

$$\text{ESS} = (\sum_{i=1}^{N}(W_n^i)^2)^{-1} \tag{3.26}$$

The explanation of ESS is that in a simple Monte Carlo integration, variance of calculation based on the N weighted samples is approximately equivalent to that based on ESS perfect samples from the target distribution. ESS takes values from 1 to N and resampling is applied only when ESS falls below a threshold N_T, say $N_T = \frac{N}{2}$. Apart from introducing variance, resampling also prevent efficient parallelization of the whole algorithm, which is discussed in details later.

3.4.4 Estimation Results

Table 3.3. Estimation of Parameters via Systematic Resampling

	ρ	μ	σ
mean	0.2012	5.287e-07	1.22e-04
standard deviation	1.030e-02	3.693e-07	4.915e-06

Table 3.4. Estimation of Parameters via Branching Resampling

	ρ	μ	σ
mean	0.1920	5.225e-07	1.219e-04
standard deviation	1.253e-02	4.007e-07	6.026e-06

Figure 3.3 and Figure 3.4 show that the Markov chain moves around the parameter space very well. Autocorrelations of the parameters indicate that the mixing properties of Markov chain is good. The estimations of ρ and σ is around the true

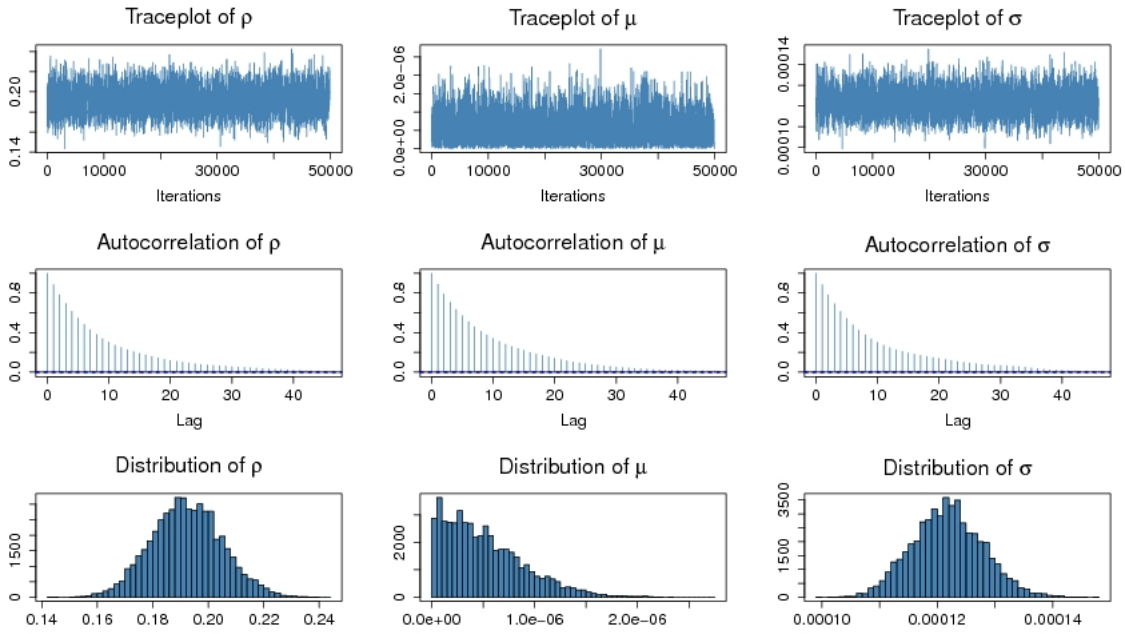

Figure 3.3. Estimation Results via Systematic Resampling

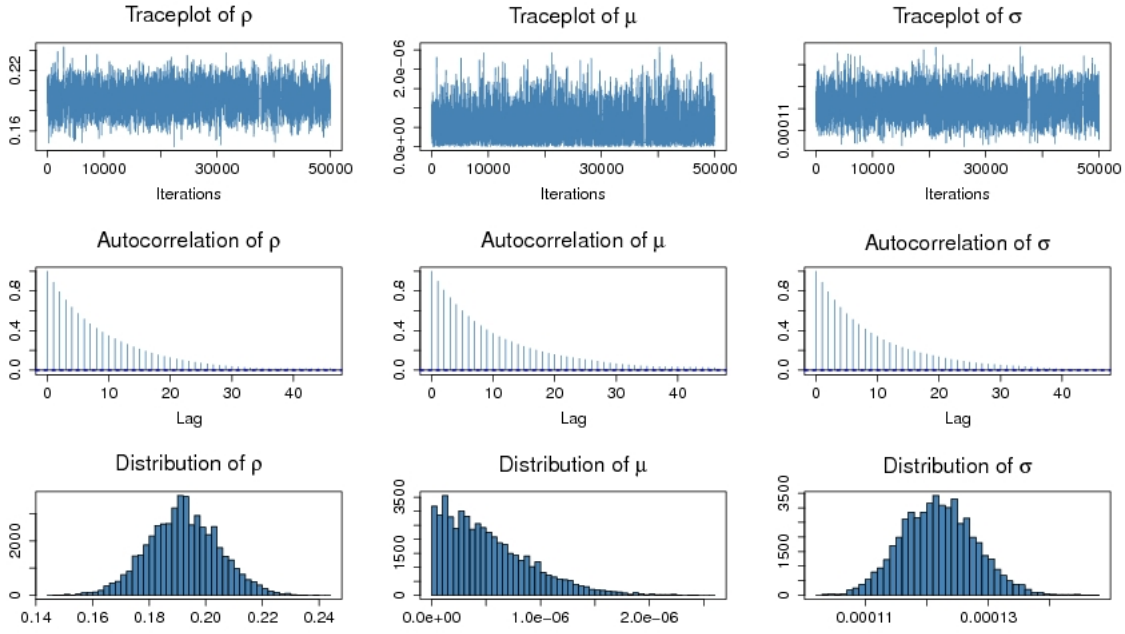

Figure 3.4. Estimation Results via Branching Resampling

value with small standard deviations. However, the result of μ is not so accurate. The reason is that 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. Moreover, there is no significant difference between results via systematic resampling scheme and results via branching resampling scheme.

3.4.5 Advanced Programming Method

One drawback of PMCMC method is its computational burden. However, parallel programming techniques may ease the problem. Since C language is used in the simulation experiment, there are two libraries can be used in parallel programming: OpenMP (OpenMP Architecture Review Board, 2008) and OpenMPI (Gabriel et al., 2004). OpenMP is used in shared memory programming, which can use all the cores in a computer to execute programs. OpenMPI is used in distributed programming, which is able to link several computers with independent processors and memory pools to carry out one program simultaneously. Both libraries are ready to be applied in our algorithm.

It is difficult to parallel MCMC step in PMCMC method, because in most efficient MCMC algorithms, the updates of parameters are related to their previous values. Thus without changing the structure of Algorithm 2, the only parts suitable for parallelization are particle filtering step and resampling step, as is shown in Algorithm 3. In particle filtering step, simulations of different particles are independent. Therefore multi-threads can be created, and each thread is able to undertake simulation tasks for a certain number of particles. However, the next

Algorithm 3 Inner-loop Parallel Programming Mode

```
for L in Markov Chain Length do
  Propose new values for parameters.
  # parallel for loop
  for N in Particle Number do
    Simulation particles.
    Calculate weights.
    Calculate likelihood.
  end for
  # end parallel
  Gather data.
  # parallel for loop
  for N in Particle Number do
    Reampling.
  end for
  # end parallel
end for
```

Algorithm 4 Outer-loop Parallel Programming Mode

```
# parallel for loop
for L in Markov Chain Length do
  Propose new values for parameters.
  for N in Particle Number do
    Simulation particles.
    Calculate weights.
    Calculate likelihood.
  end for
  Gather data.
  for N in Particle Number do
    Reampling.
  end for
end for
# end parallel
```

parallelization for resampling steps cannot be conducted until all simulated data from different threads are gathered. Therefore the parallel regions are short in CPU time, and need to be terminated and restarted frequently, which leads to large overheads when executing parallel version of the program.

Overheads is the extra time needed when running multi-threads programs. For example, when entering into parallel region, threads have to be created and some data structures have to be set up to carry information needed by the system. When a work-sharing directive is implemented, the work to be performed by each thread is usually determined at running time. Time spent on these operations is collectively called parallelization overheads.

In general, there are four types of overheads: Synchronization overheads, load imbalance overheads, limited parallelization overheads and threads management overheads. Imbalance overheads is waiting time due to an imbalanced amount of work in a work-sharing or parallel region. Synchronization overheads is waiting time due to threads which have to synchronize their activities. Limited parallelization overheads is caused by idle threads due to not enough parallelism being exposed by the program. This kind of overheads is related to algorithm design, and does not appear in our program, because parallel regions in our program are all for loops. Thread management overheads is spent in the process of creating and destructing threads. In Table 3.5, a real example of these four kinds of overheads in PMCMC program is given when two cores are applied with OpenMP library.

In Figure 3.5, it is obvious to see that there is indeed acceleration for the program via Algorithm 3. However, the CPU time does not decrease much corresponding to

Table 3.5. Overheads Caused by Multi-Threads (2 Threads)

		Sum	Particle Filtering	Resampling	Other
Total	(%)	5980.05	815.75	5039.77	124.53
Ovhds	(%)	1716.34 (26.44)	12.54 (0.19)	1695.95(26.13)	7.85 (0.12)
Synch	(%)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Imbal	(%)	1703.45 (26.25)	7.48 (1.91)	1692.72 (26.08)	3.07 (0.05)
Limpar	(%)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Mgmt	(%)	12.89 (0.20)	4.87 (0.08)	3.24 (0.05)	4.78 (0.07)

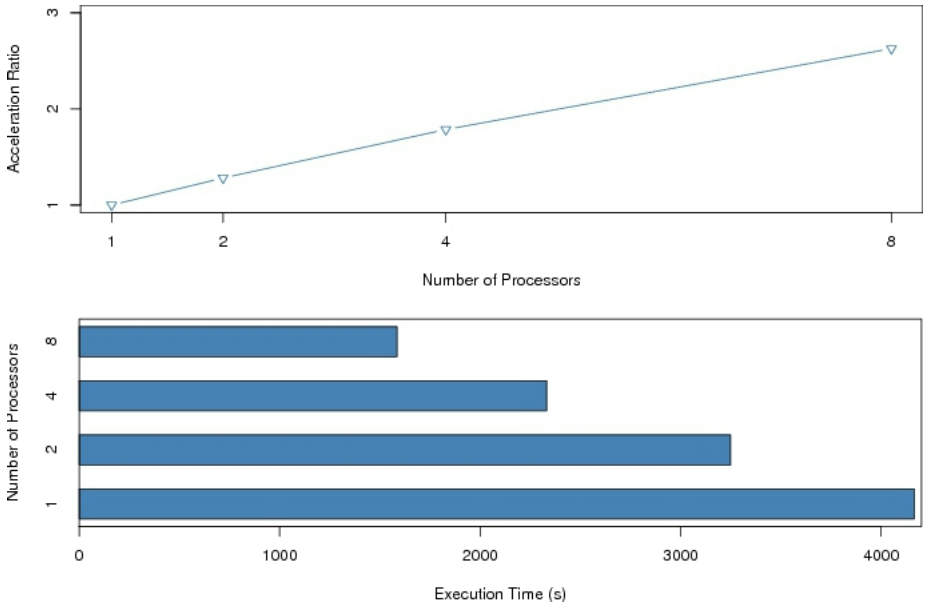

Figure 3.5. Acceleration Effects via Algorithm 3

increasing number of processors. There are two possible reasons. The first reason is small parallel degree. Parallel degree can be simply regarded as CPU time of one parallel region. As is seen in Algorithm 3, the whole program is nested loops. Usually, parallel region should be the outside loop to achieve larger parallel degree. But due to correlation of MCMC chain in the algorithm, it is difficult to do that. Consequently there is much threads management overheads.

The second reason is that load imbalance overheads is huge. As is seen in Table 3.5, imbalance overheads takes up around 26% of executing time. This is caused by the randomness of resampling schemes. It is impossible to determine in advance the number of steps needed to get one sample. As a result, when allocating jobs to different threads, it is impossible to give them equal computational loads. Some threads with less loads have to wait for those with more loads. Therefore, such kind of overheads is inevitable as long as parallelization of resampling steps is required.

Due to these two reasons, simply changing the program from serial to parallel can only partially solve the problem of long CPU time. There is another kind of MCMC algorithm which is not widely used but helpful to obtain larger parallel degree, that is, independent MCMC method (see Liu, 1996). In algorithm of independent MCMC, the proposed parameters are independent of their previous values. Algorithm 2 is revised by letting $q(\theta|\theta') = q(\theta)$, and $q(\theta|\theta') = q(\theta)$, as shown in Algorithm 5.

This modification has two nice features. Firstly, because the parallel region can be placed in the outside loop, it allows larger parallel degree, just as Algorithm 4.

Algorithm 5 Particle Independent MCMC for State Space Model

At time t=0

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 iteration $i \geq 1$ **do**

Sample $\theta' \sim q(\cdot)$.

Run a particle filtering algorithm, obtaining $\hat{p}_{\theta_0}(y_{1:t})$, denote as estimation of marginal likelihood value.

With probability $\alpha = \min(1, \frac{p(\theta', S'_{1:t}|y_{1:t})}{p(\theta, S_{1:t}|y_{1:t})} = \frac{p_{\theta'}(y_{1:t})}{p_{\theta}(y_{1:t})} \frac{q(\theta)}{q(\theta')})$.

Set $\theta_i = \theta'$, and $\hat{p}_{\theta_i}(y_{1:t}) = \hat{p}_{\theta'}(y_{1:t})$.

end for

Secondly, OpenMPI library can be used, which is helpful to break the limitation of number of processors in one computer. As is shown in Figure 3.6, there is significant shrink of executing time of the program via Algorithm 5. When 88 cores are used, the executing time is only $\frac{1}{75}$ of the total CPU time needed for the serial program.

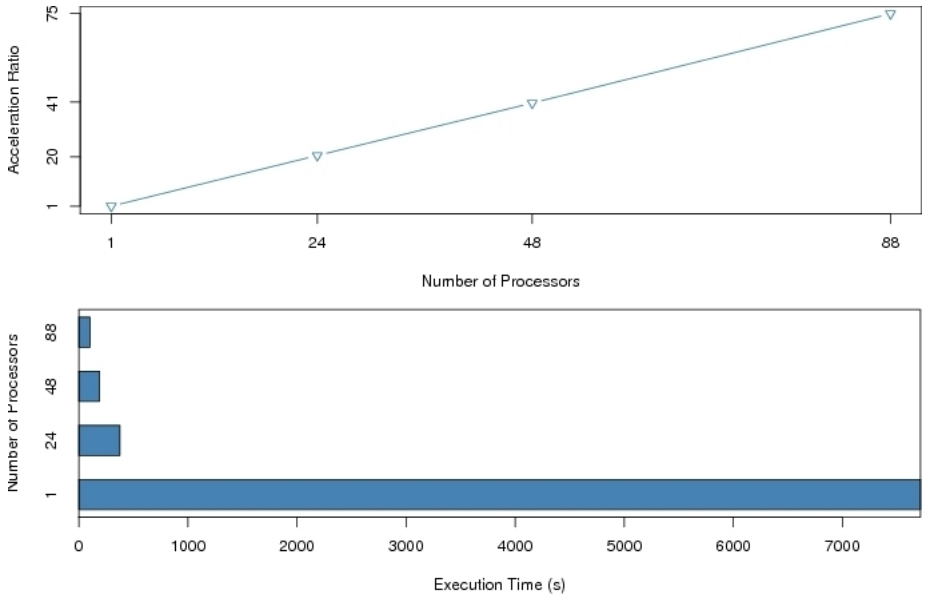

Figure 3.6. Acceleration Effects via Algorithm 5

In addition, a comparison of Figure 3.3 and Figure 3.7 shows that the estimation results are good via the independent MCMC method. But this algorithm also has

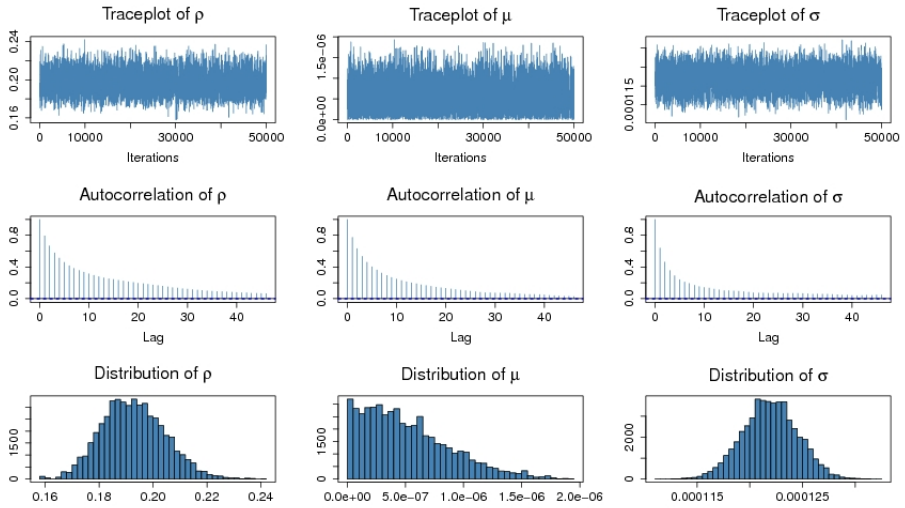

Figure 3.7. Estimation Results via Particle Independent MCMC Method

drawbacks. On the one hand, parameters for proposal density $q(\theta)$ are unknown in real world applications. As a result, a lot of work is needed to find suitable values. On the other hand, how to deal with the order of proposed value from each processor needs further studies. In Algorithm 5, proposed parameters are processed in the order of processors' rank. Jacob et al. (2010) propose other possible choices, which may lead to smaller estimation variance.

3.5 Simulation Study for Jump-Diffusion Process

3.5.1 Simulation

In studies based on low frequency 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 one 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 minutes, 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 simulation experiment, parameters except for jumps are set to be the same as Section 3.4.1: $M = 8$, $\rho = 0.2$, $\alpha = 0.225$, $\beta = 0.066$, $\gamma = 0.3$, $\mu = 4.4\text{e}^{-8}$, $\sigma = 1.2\text{e}^{-4}$. One day's data is generated with three different sets for jump parameters $(\mu_J,\ \sigma_J,\ \lambda)$: $\Theta_1 = (4.4e-5,\ 1.2e-5,\ 0.01)$, $\Theta_2 = (4.4e-3,\ 1.2e-3,\ 0.0001)$ and $\Theta_3 = (4.4e-3,\ 1.2e-3,\ 0.001)$. The number of simulated tradings is 1374, particle number is 2000, and the length of Markov chain is 45000. The last 35000 data is used for analysis.

3.5.2 Estimation Results

Before further discussions, the concepts of quadratic variation, realized volatility and integrated volatility (see Barndorff-Nielsen and Shephard, 2004) should be introduced for a better understanding of following contents.

Definition 3.1 (Quadratic Variation)**.** *Let S_t be a stochastic process. To compute quadratic variation of S_t on [0, T], we choose* $0 = t_0 < t_1 < \ldots < t_n = T$,

denote the set of these times by $\Pi = t_0, t_1, \ldots, t_n$*, denote the length of the largest subinterval by* $\|\Pi\| = \max_j(t_{j+1} - t_j)$*, and defined*

$$Q_\Pi(S_t) = \sum_{j=0}^{n-1}(S_{t_{j+1}} - S_{t_j})^2.$$

The quadratic variation of S_t *on [0, T] is defined to be*

$$QV(T) = \lim_{\|\Pi\|\to 0} Q_\Pi(S_t).$$

Let S_t *be a stochastic process as Equation (3.2), which is a geometric Brownian motion. The quadratic variation of* S_t *on [0, T] is*

$$QV(T) = \int_0^T \sigma^2 ds.$$

Let S_t *be a stochastic process as Equation (3.3), which is a jump-diffusion process. The quadratic variation of* S_t *on [0, T] is*

$$QV(T) = \int_0^T \sigma^2 ds + \sum_{0<s\leq T} \Delta J_s^2.$$

The so called realized volatility in studies of high frequency modeling is an estimation of quadratic variation.

Definition 3.2 (Integrated Volatility)**.** *Let* S *be a stochastic process defined by Equation (3.2) and Equation (3.3). The integrated volatility of* S *on [0, T] is defined to be*

$$IV(T) = \int_0^T \sigma^2 ds.$$

Definitions above indicate that when S_t is a geometric Brownian motion, integrated

volatility can be estimated by realized volatility. However, when S_t is a jump-diffusion process, realized volatility is larger than integrated volatility due to the volatility generated from jumps. The simulation study shows influence of jumps on parameter estimations by PMCMC method. Then a comparison about accuracy of integrated volatility estimation is made among PMCMC method and other nonparametric methods.

Table 3.6. Estimation Results for Parameter Set Θ_1

	True Value	Mean	Standard Deviation
ρ	0.200	1.877e-1	4.380e-5
μ	4.400e-8	3.399e-7	2.037e-7
σ	1.200e-4	1.160e-4	5.925e-6
μ_J	4.400e-5	4.180e-5	5.056e-6
σ_J	1.200e-5	1.092e-5	2.879e-6
λ	0.01	1.084e-2	4.633e-3

Table 3.7. Estimation Results for Parameter Set Θ_2

	True Value	Mean	Standard Deviation
ρ	0.200	1.937e-1	8.484e-3
μ	4.400e-8	3.564e-7	2.080e-7
σ	1.200e-4	1.185e-4	6.171e-6
μ_J	4.400e-3	4.324e-3	4.842e-4
σ_J	1.200e-3	1.154e-3	2.803e-4
λ	0.0001	1.158e-4	4.217e-5

Table 3.8. Estimation Results for Parameter Set Θ_3

	True Value	Mean	Standard Deviation
ρ	0.200	1.856e-1	8.893e-3
μ	4.400e-8	3.980e-7	2.456e-7
σ	1.200e-4	1.192e-4	6.284e-6
μ_J	4.400e-3	4.452e-3	4.082e-4
σ_J	1.200e-3	1.180e-3	2.582e-4
λ	0.001	1.103e-3	2.251e-4

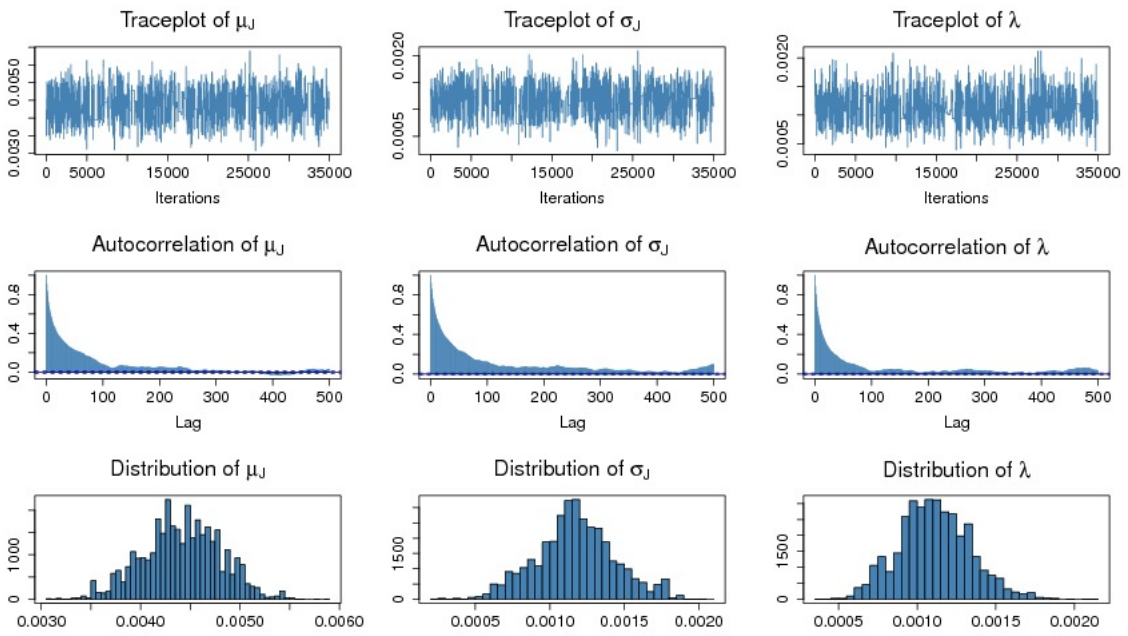

Figure 3.8. Estimation Results for Jump-Diffusion Process of Θ_3

Table 3.6 - 3.8 list estimation results for these three parameter sets, which are reasonable. Mixing properties of Markov chain for parameter set Θ_1 and set Θ_2 are similar to Figure 3.3, thus are not reported here. However, for parameter set Θ_3, the autocorrelation is a little higher and mixing properties of Markov chain is not so good. One possible reason is that jumps generated volatility is substantial in the quadratic variation in the case of Θ_3. When jumps come frequently and jump sizes are large, particle filtering method (Algorithm 1) may generate inaccurate samples, as discussed by Johannes et al. (2009). That's partially because Algorithm 1 simulates new states S_{t+1} blindly from distribution $f_\theta(S_{t+1} \mid S_t)$, and does not use new information contained in Y_{t+1} at this step. An auxiliary particle filtering algorithm developed by Pitt and Shephard (1999) may alleviate the problem. However, due to the heavy computational burden, it is not applicable. Therefore, how to filter jumps efficiently is an important future work.

3.5.3 Comparison with Nonparametric Methods

As mentioned before, integrated volatility and realized volatility are different. Integrated volatility is generated by the continuous component of price process without considering the impact of jumps. However, realized volatility contains integrated volatility and volatility generated from potential jump component. In studies of risk management, it is useful to decompose realized volatility into the contribution from the continuous component and the jumps. For example, in option trading, these two types of volatilities have different hedging requirements and possibilities. In portfolio allocation, the demand for assets subject to both types of risks can be optimized further if such a decomposition is available. Moreover, such a decomposition makes it possible to manage the Brownian risk and large jump risk using Value-at-Risk (VaR) and other tail statistics respectively. This section will compare accuracy of estimations for integrated volatility via PMCMC method and nonparametric methods. Once we have estimations for integrated volatility and realized volatility, the difference between them is volatility generated by jumps.

Many extant nonparametric methods for high frequency data modeling can only estimate quadratic variation (realized volatility), except for Bipower Method developed by Barndorff-Nielson and Shephard (2004) and its extensions. Four widely used nonparametric methods are applied. The first method is Classical Realized Method developed by Andersen et al. (2001). The second method is Two Timescales Method developed by Aït-Sahalia et al. (2005b). The third method is Kernel Method developed by Hansen and Lunde (2006) and Barndorff-Nielson et al. (2008). The fourth method is Bipower Method developed by Barndorff-Nielson

and Shephard (2004). A detailed review is given by McAleer and Medeiros (2006). Among all these methods, Classical Realized Method is developed to estimate realized volatility before extensive research of high frequency data, therefore it suffers from market microstructure noise. The Two Timescales Method and Kernel Method, which also calculate realized volatility, are robust to market microstructure noise, whereas not robust to jumps. These three methods are used to examine the influence of jumps on estimation of integrated volatility when the model is misspecified. Bipower Method is robust to jumps, which calculates integrated volatility.

Table 3.9. Estimations of Integrated Volatility for Different Parameter Sets

	Θ_1	Θ_2	Θ_3
True Value	**3.370e-4**	**3.370e-4**	**3.370e-4**
PMCMC Method	3.148e-4	3.285e-4	3.325e-4
Bipower Method	5.688e-4	2.966e-4	3.401e-4

Table 3.10. Estimations of Realized Volatility for Different Parameter Sets

	Θ_1	Θ_2	Θ_3
True Value	**3.375e-4**	**4.000e-4**	**9.451e-4**
Classical Realized Method	8.927e-3	9.100e-3	7.811e-3
Two Timescales Method	8.245e-4	8.615e-4	1.859e-3
Kernel Method	8.391e-4	8.740e-4	1.264e-3

In Table 3.10, Classical Realized Method which uses all tick-by-tick data overestimates quadratic variation significantly due to market microstructure noise. Kernel Method and Two Timescales Method bring us closer results for quadratic variation[3].

[3] The calculations of nonparametric methods are simple and fast. However, it is not an easy task to choose proper parameters for these methods. In Appendix B, a method for choosing parameters are given. The calculations are conducted via R software (R Development Core Team, 2011) with the package 'realized' (Payseur, 2011).

In Table 3.9, the jump robust Bipower Method is able to provide close estimation for integrated volatility, whereas results of PMCMC are better. However only accurate estimation of integrated volatility can not make PMCMC method more valuable than bipower method, because bipower method is much less computationally expensive. The strength of PMCMC method is that it can provide reasonable parameter estimations of jump magnitudes and jump intensity, which have many potential applications in asset pricing and risk management.

3.6 Summary and Remarks

In this chapter, the particle Markov chain Monte Carlo method, for the first time, is used to estimate high frequency data model. The implementation of this method is straightforward and based on extensive simulation method. The method can estimate integrated volatility and the parameters of jumps. One future work is to modify this method to estimate arrival times of jumps, which is also essential in financial analysis. Since particle filtering is mostly applied in engineering, theoretical results are not so rich as other statistical methods. Therefore, another future work is to develop convergence theories for more advanced algorithms.

Chapter 4

Conclusion and Future Work

In this thesis, new perspectives for jump-diffusion model estimation are provided. In the first part of thesis, a new estimation method is developed for continuous state branching process with immigration (hereafter, CBI). CBI process has a simplified version as Cox-Ingersoll-Ross models with jumps (JCIR), which is widely applied to model stochastic interest rates. Our method is based on weighted conditional least square estimators (WCLSE), and avoids the computationally expensive numerical integration which exists in many other estimation methods for JCIR model. Simulation study shows that this method leads to satisfying estimations. Empirical study provides an example for application of our method via American Federal Fund rate data.

In future studies, several things can be done for the method. First of all, although some researchers, such as Wei and Winnicki (1990) and Overbeck (1998), develop estimation methods for CBI process, their methods are limited to estimate draft and diffusion parameters in models without jumps. In the thesis, their methods are extended to consider influence of jumps. However, the estimation method is still developed for drift and diffusion parameters in JCIR model, which cannot give us parameters for jumps. As is shown in empirical study, some nonparametric methods, such as method developed by Bandi and Nguyen (2003), should be ap-

plied first to estimate parameters for jumps. Therefore, it is necessary to extend our method in the thesis further to estimate parameters for jumps.

Secondly, in simulation studies, estimation of diffusion parameter has larger bias compared with those of draft parameters. The reason is that these parameters are not estimated simultaneously. Estimation of diffusion parameter depends on drift parameters. Errors of drift parameters may accumulate during estimation, which may influence estimation of diffusion parameter. As a result, alternative procedure which can estimate diffusion parameter independently is preferable.

Moreover, since it is difficult to estimate parameters in models with jumps, in many papers about bond pricing, interest rate is still modeled as a continuous-path stochastic process. The method developed in thesis is simple and fast in calculation, therefore it could be extended to other interest rate models, such as Vasicek model (Vasicek, 1977).

In the second part of this thesis, the particle Markov chain Monte Carlo method, for the first time, is applied to estimate a jump-diffusion high frequency data model. In studies of high frequency data modeling, nonparametric methods are extensively used, and market microstructure noise is often modeled implicitly. In the thesis, according to Zeng (2003), a method explicitly dealing with market microstructure noise is applied to high frequency stock price. Particle Markov chain Monte Carlo method is used to estimate parameters in the model. This estimation method, based on Monte Carlo integration theory, is able to reduce variance of numerical integration and achieve high efficiency via parallel programming when large data set is involved. Detailed simulation studies explore the properties of this

method. A comparison is made among the method and other extant nonparametric methods. Results show that this method is able to estimate parameters in the model accurately. Meanwhile, it can also estimate volatility generated by diffusion component under the influence of jumps with market microstructure noise.

Much remains to be done in the field of particle Markov chain Monte Carlo method and high frequency data modeling. First of all, the estimation approach relies on particle approximation to likelihood function in the model. Under mild regularities on state transition function and the likelihood function, approximation to likelihood function by particle filters of Algorithm 1 converges to the true value, see Crisan and Doucet (2002) for a summary. Del Moral (2004) proves consistency and asymptotic normality results for the particle filtering approximation to the likelihood function for the state-space model. Proof for convergence of branching particle filtering algorithm is provided by Xiong and Zeng (2011). However, since particle filtering is mainly used in engineering, the theoretical results are not so rich as other statistical estimation methods. Therefore, the convergence for more advanced particle filtering algorithms and more complicated models is still an open problem.

Secondly, particle filtering method is computationally expensive. To increase speed of the algorithm through parallel programming, independent MCMC algorithm is needed. In real world applications, parameters for proposal density in independent MCMC algorithm are unknown. As a result, a lot of work is needed to find out suitable values. For some MCMC methods, there are adaptive algorithms which are able to search the parameters automatically, see Andrieu and Thoms (2008) for

an overview. However, for independent MCMC, there is no well-accepted solutions. Therefore it is useful to develop efficient adaptive methods which automatically search parameters for proposal density in independent MCMC algorithm.

Moreover, although being able to estimate all parameters in the high frequency data model, the method cannot give us arrival times of jumps. It is useful if arrivals of jumps can be analyzed together with macro-econometric information. Therefore, it is necessary to develop filtering methods to identify when jumps happen.

Appendix A. Resampling Schemes for Particle Filtering

There are detailed descriptions for five resampling schemes in particle filtering applied in Section 3.4.3. In each resampling schemes, suppose that there are N particles with normalized weights $\{X_i, W_i\}_{i=1}^N$, and samples $\{\widetilde{X}_i, \widetilde{W}_i\}_{i=1}^N$ should be generated.

Algorithm A.1 Multinomial Resampling Algorithm

Initialize CDF: obtain $\{P_i\}_{i=1}^N$, where $P_k = \sum_{i=1}^{k} W_i$.
for iterations $1 \leq i \leq N$ **do**
  Generate an uniformly distributed random number $U_i \sim [0, 1]$.
  while $U_i > P_j$ **do**
    $j = j + 1$.
  end while
  Set $\widetilde{X}_i = X_j$ and $\widetilde{W}_i = 1/N$.
end for

Algorithm A.2 Residual Resampling Algorithm

First Step: Generate deterministic sample (sample one)
for $1 \leq i \leq N$ **do**
  $\widetilde{N}_i = \text{floor}(N * W_i)$.
end for
  Set sample size $R = \sum_{i=1}^{N} \widetilde{N}_i$.
  Repeat the i-th particle with $\widetilde{N}_i$ times.
Second Step: Generate random sample (sample two)
  Set $\widehat{W}_i = W_i - N^{-1}\widetilde{N}_i$.
  Normalized $\widehat{W}_i$ to obtain $\widetilde{W}_i$.
  Multinomial Sampling with $\{X_i, \widetilde{W}_i\}_{i=1}^N$, and sample size $N - R$.
Third Step:
  Merge sample one and sample two, and set final $\widetilde{W}_i = 1/N$.

Algorithm A.3 Stratified Resampling Algorithm

Initialize CDF: obtain $\{P_i\}_{i=1}^N$, where $P_k = \sum_{i=1}^{k} W_i$.
for iteration $1 \leq i \leq N$ **do**
  Generate an uniformly distributed random number $U_i \sim [\frac{i-1}{N}, \frac{i}{N}]$.
  while $U_i > P_j$ **do**
    $j = j + 1$.
  end while
  Set $\widetilde{X}_i = X_j$ and $\widetilde{W}_i = 1/N$.
end for

Algorithm A.4 Systematic Resampling Algorithm

Initialize CDF: obtain $\{P_i\}_{i=1}^N$, where $P_k = \sum_{i=1}^{k} W_i$.
Generate an uniformly distributed random number $U \sim [0, \frac{1}{N}]$.
for iteration $1 \leq i \leq N$ **do**
  Generate $U_i = (i-1)/N + U$.
  while $U_i > P_j$ **do**
    $j = j + 1$.
  end while
  Set $\widetilde{X}_i = X_j$ and $\widetilde{W}_i = 1/N$.
end for

Algorithm A.5 Branching Resampling Algorithm

for iteration $1 \leq i \leq N$ **do**
  Generate an uniformly distributed random number $U_i \sim [0, 1]$.
  Generate integer part of $W_i : [W_i]$.
  Generate fraction part of $W_i : \{W_i\} = W_i - [W_i]$.
  if $U_i < \{W_i\}$ **then**
    X_i will be repeatedly selected $[W_i]$ times in $\{\widetilde{X}_i\}_{i=1}^N$.
  else
    X_i will be repeatedly selected $[W_i] + 1$ times in $\{\widetilde{X}_i\}_{i=1}^N$.
  end if
end for

Appendix B. Parameter Choosing for Calculation of Realized Volatility via Nonparametric Methods

Parameter choosing for nonparametric methods is through signature plot. Signature plot developed by Andersen et.al (2000) displays the sample average realized variance across n days (often one year or one month), as a function of different sampling frequencies. Payseur (2007) uses the concept to help choose proper parameters. Instead of sample average realized variance across n days, realized variance calculated from one day's data is plotted, as a function of different value of parameters. Reasonable values of parameters are chosen when variability of estimations becomes stable. There are some examples using Payseur's method. Figure B.1 is signature plot for realized volatility estimations by Classical Realized Method as a function of sub-sampling frequency. Figure B.2 is signature plot for realized volatility estimations by Two Timescales Method as a function of sub-sampling frequency. Figure B.3 is signature plot for realized volatility estimations by Kernel Method (Barlett kernel) as a function of lag parameter. The realized volatility is the average of estimations when their values become stable.

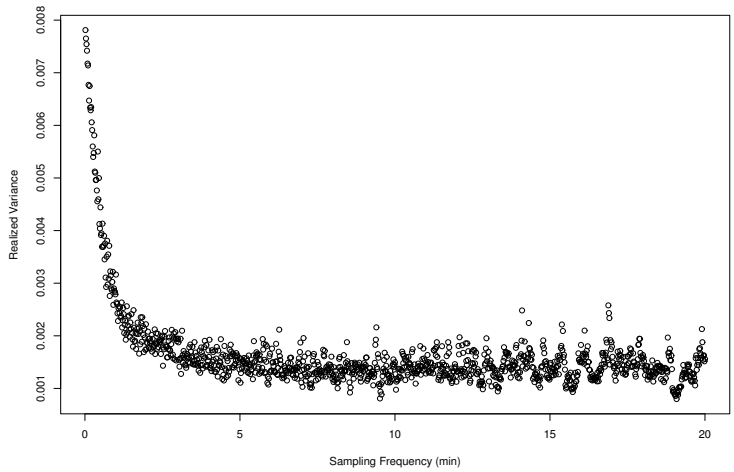

Figure B.1. Signature Plot for Classical Realized Method

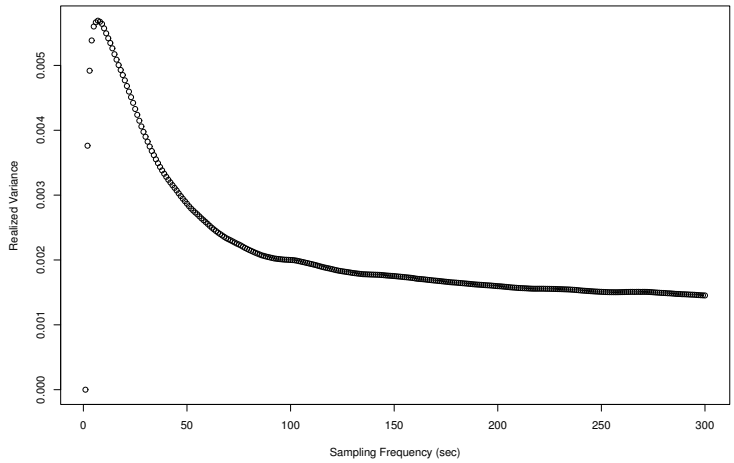

Figure B.2. Signature Plot for Two Timescales Method

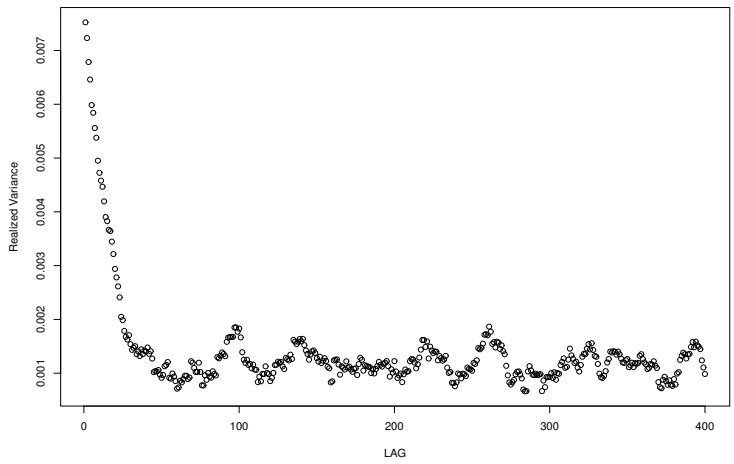

Figure B.3. Signature Plot for Barlett Kernel Method

References

[1] Ahn, D. and Gao, B. (1999). A Parametric Nonlinear Model of Term Structure Dynamics. *Review of Financial Studies* **12**, 721-762.

[2] Aït-Sahalia, Y., Mykland, P. A. and Zhang, L. (2005a). How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise. *Review of Financial Studies* **18**, 351-416.

[3] Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2005b). A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High Frequency Data. *Journal of the American Statistical Association* **100**, 1394-1411.

[4] Aït-Sahalia, Y. and Jacod, J. (2009). Estimating the Degree of Activity of Jumps in High Frequency Financial Data. *The Annals of Statistics* **37**, 2202-2244.

[5] Aït-Sahalia, Y. and Jacod, J. (2011). Testing Whether Jumps Have Finite or Infinite Activity. *The Annals of Statistics* **39**, 1689-1719.

[6] Aldous, D. (1978). Stopping Times and Tightness. *Annals of Probability* **6**, 335-340.

[7] An, S. and Schorfheide, F. (2007). Bayesian Analysis of DSGE Models. *Econometric Reviews* **26**, 113-172.

[8] Andersen, T. G. and Bollerslev, T (1998). Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. *International Economic Review* **39**, 885-905.

[9] Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2000). Great Realizations. *Risk* **13**, 105-108.

[10] Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2001). The Distribution of Exchange Rate Volatility. *Journal of the American Statistical Association* **96**, 42-55.

[11] Andrieu, C. and Thoms, J. (2008). A Tutorial on Adaptive MCMC. *Statistics and Computing* **18**, 343-373.

[12] Andrieu, C., Doucet, A. and Holenstein, R. (2010). Particle Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72**, 269-342.

[13] Bandi, F. M. and Nguyen, T. H. (2003). On the Functional Estimation of Jump-Diffusion Models. *Journal of Econometrics* **116**, 293-328.

[14] Bandi, F. M. and Phillips, P. C. (2003). Fully Nonparametric Estimation of Scalar Diffusion Models. *Econometrica* **71**, 241-283.

[15] Bandi, F. M. and Russel, J. R. (2006). Separating Microstructure Noise from Volatility. *Journal of Financial Economics* **79**, 655-692.

[16] Bandi, F. M. and Russell, J. R. (2008). Microstructure Noise, Realized Volatility, and Optimal Sampling. *Review of Economics and Statistics* **75**, 339-369.

[17] Barndorff-Nielsen, O. E. and Shephard, N. (2002). Estimating Quadratic Variation using Realized Variance. *Journal of Applied Econometrics* **17**, 457-477.

[18] Barndorff-Nielsen, O. E. and Shephard, N. (2004). Power and Bipower Variation with Stochastic Volatility and Jumps. *Journal of Financial Econometrics* **2**, 1-48.

[19] 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.

[20] Bates, D. S. (1996). Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in DM options. *Review of Financial Studies* **9**, 53-56.

[21] Billingsley, P. (1999) *Convergence of Probability Measures*. New York: Wiley.

[22] Black, F. and Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* **81**, 637-654.

[23] Black, F. and Cox, C. (1976). Valuing Corporate Securities: Some Effects of Bond Indenture Provisions. *Journal of Finance* **31**, 351-367.

[24] Black, F. (1986). Noise. *Journal of Finance* **41**, 529-543.

[25] Bolstad, W. M. (2010). *Understanding computational Bayesian Statistics*. New York: Wiley.

[26] Carpenter, J., Clifford, P., and Fearnhead, P. (1999). An Improved Particle Filter for Nonlinear Problems, Radar, Sonar and Navigation. *IEE Proceedings* **146**, 1-7.

[27] Chan, K. C., Karolyi, G. A., Longstaff, F. A. and Sanders, A. B. (1992). An Empirical Comparison of the Short-Term Interest Rate. *Journal of Finance* **47**, 1209-1227.

[28] Cox, J., Ingersoll, J. and Ross, S. (1985). A Theory of the Term Structure of Interest Rates. *Econometrica* **53**, 385-407.

[29] Crisan, D. and Doucet, A. (2002). A Survey of Convergence Results on Particle Filtering Methods for Practitioners. *IEEE Transactions on Signal Processing* **50**, 736-746.

[30] Das, S. R. (2002). The Surprise Element: Jumps in Interest Rates. *Journal of Econometrics* **106**, 27-65.

[31] Dawson, D. A. and Li, Z. H. (2006). Skew Convolution Semigroups and Affine Markov Processes. *Annals of Probability* **34**, 1103-1142.

[32] Del Moral, P. (2004). *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. New York: Springer-Verlag.

[33] Del Moral, P., Doucet, A. and Jasra, A. (2011). On Adaptive Resampling Procedures for Sequential Monte Carlo Methods. To appear in *Bernoulli*.

[34] Douc, R., Cappe, O. and Moulines, E. (2005). Comparison of Resampling Schemes for Particle Filtering. *Proceedings of the Fourth International Symposium on Image and Signal Processing and Analysis*, 64-69

[35] Douc, R. and Moulines, E. (2008). Limit Theorems for Weighted Samples with Applications to Sequential Monte Carlo Methods. *The Annals of Statistics* **36**, 2344-2376.

[36] Doucet, A and Johansen, A. M. (2011). A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later. *Oxford Handbook of Nonlinear Filtering*, eds. Crisan, D and Rozovsky, B. Oxford University Press.

[37] Duan, J.C. and Fulop, A. (2007). How Frequently Does the Stock Price Jump? - An Analysis of High-Frequency Data with Microstructure Noises. Working Paper.

[38] Duan, J. C. and Fulop, A. (2009). Estimating the Structural Credit Risk Model When Equity Prices are Contaminated by Trading Noises. *Journal of Econometrics* **150**, 288-296.

[39] Duffie, D., Pan, J. and Singleton, K. (2000). Transform Analysis and Asset Pricing for Affine Jump-Diffusions. *Econometrica* **68**, 1343-1376.

[40] Duffie, D. (2005). Credit Risk Modeling with Affine Processes. *Journal of Banking & Finance* **29**, 2751-2802.

[41] Durrett, R. (2010). *Probatility: Theory and Examples*. New York: Cambridge University Press.

[42] Engle, R. and Russell, J. R. (1998). Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. *Econometrica* **66**, 1127-1162.

[43] Engle, R. (2000). The Econometrics of Ultra-High-Frequency Data. *Econometirca* **68**, 1-22.

[44] Engle, R. and Russell, J. R. (2005). Analysis of High Frequency Data. *Handbook of Financial Econometrics*, eds. Aït-Sahalia, Y and Hansen L.P. Amsterdam: North-Holland.

[45] Ethier, S. N. and Kurtz, T. G. (1986). *Markov Processes: Characterrization and Convergence*. New York: Wiley. Fernández-Villaverde and Rubio-Ramírez (2007)

[46] Farnsworth, H., and Bass, R. (2003). The Term Structure with Semi-Credible Targeting. *Journal of Finance* **58**, 839-865.

[47] Fernández-Villaverde, J. and Rubio-Ramírez, J. F. (2007). Estimating Macroeconomic Models: A Likelihood Approach. *Review of Economic Studies* **74**, 1059-1087.

[48] Filipović, D. (2001). A General Characterization of One Factor Affine Term Structure Models. *Finance and Stochastics* **5**, 389-412.

[49] Gabriel, E. et al. (2004). Open MPI: Goals, Concept, and Design of a Next Generation MPI Implementation. *11th European PVM/MPI User Group Meeting*, 97-104.

[50] Ghysels, E., Gourieroux, C. and Jasiak, J. (2004). Stochastic Volatility Duration Models. *Journal of Econometrics* **119**, 413-433.

[51] Gikhman, I. I. and Skorohod, A. V. (1972). *Stochastic Differential Equantions.* New York: Springer.

[52] Grammig, J. and Wellner, M. (2002). Modeling the Interdependence of Volatility and Inter-Transaction Duration Processes. *Journal of Econometrics* **106**, 369-400.

[53] Gordon, N. J., Salmond, D. J. and Smith, A. F. M. (1993). Novel Approach to Nonlinear / Non-Gaussian Bayesian State Estimation. *Radar and Signal Processing* **140**, 107-113.

[54] Hall, P. and Heyde, C. (1980). *Martingale Limit Theory and its Applications.* New York: Academic Press.

[55] Hansen, P. R. and Lunde, A. (2006). Realized Variance and Market Microstructure Noise. *Journal of Business and Economic Statistics* **24**, 1-58.

[56] He, C., Kennedy, J. S., Coleman, T. F., Forsyth, P. A., Li, Y., and Vetzal, K. R. (2006). Calibration and Hedging under Jump Diffusion. *Review of Derivatives Research* **9**, 1-35.

[57] Hu, X., Kuipers, D. and Zeng, Y. (2010). Econometric Analysis via Filtering for Ultra-High Frequency Data. Working Paper.

[58] Huang, J. H., Ma, C. H. and Zhu, C. (2011). Estimation for Discretely Observed Continuous State Branching Processes with Immigration. *Statistics & Probability Letters* **81**, 1104-1111.

[59] Jacob, P., Robert, C. P. and Smith, M. H. (2011). Using Parallel Computation to Improve Independent Metropolis-Hastings Based Estimation. Working Paper.

[60] Jacod, J. and Schiryaev, A. N. (1987). *Limit Theorems for Stochastic processes*. New York: Springer-Verlag.

[61] Jarrow, R. and Protter, P. (2004). Structural Versus Reduced Form Models: a New Information Based Perspective. *Journal of Investment Management* **2**, 1-10.

[62] Jarrow, R., Li, H. and Zhao, F. (2007). Interest Rate Caps 'Smile' Too! But Can the LIBOR Market Models Capture It? *Journal of Finance* **61**, 341-378.

[63] Jiang, G. J. and Oomen, R. C.A.. (2008). Testing for Jumps When Asset Prices are Observed with Noise - a "Swap Variance" Approach. *Journal of Econometrics* **144**, 352-370.

[64] Johannes, M. (2004). The Statistical and Economic Role of Jumps in Continuous-Time Interest Rate Models. *Journal of Finance* **59**, 227-260.

[65] Johannes, M. S., Polson, N. G. and Stroud, J. R.. (2009) Optimal Filtering of Jump Diffusions: Extracting Latent States from Asset Prices. *The Review of Financial Studies* **22**, 2559-2599.

[66] Kawazu, K. and Watanabe, S. (1971). Branching Processes with Immigration and Related Limit Theorems. *Theory of Probability and its Applications* **16**, 36-54.

[67] Kou, S. (2002). A Jump Diffustion Model for Option Pricing. *Management Science* **48**, 1086-1101.

[68] Kouritzin, M. and Zeng, Y. (2005). Bayesian Model Selection via Filtering for a Class of Micro-Movement Models of Asset Price. *International Journal of Theoretical and Applied Finance* **8**, 97-121.

[69] Lee, K. and Zeng, Y. (2010). Risk Minimization for a Filtering Micromovement Model of Asset Price. *Applied Mathematical Finance* **17**, 177-199.

[70] Lee, S. and Mykland, P.A. (2008). Jumps in Financial Markets: A New Nonparametric Test and Jump Dynamics. *Review of Financial Studies* **21**, 2535-2563.

[71] Liu, J. S. (1996). Metropolized Independent Sampling with Comparison to Rejection Sampling and Importance Sampling. *Statistics and Computing* **6**, 113-119.

[72] Liu, J. S. and Chen, R. (1998). Sequential Monte Carlo Methods for Dynamical Systems. *Journal of the American Statistical Association* **93**, 1032-1044.

[73] Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*. New York: Springer-Verlag.

[74] Ludger, O. and Rydén, T. (1997). Estimation in the Cox-Ingersoll-Ross Model. *Econometric Theroy* **13**, 430-461.

[75] Martin, J. (2006). *Point Process Theory and Applications*. Boston: Birkhäuser.

[76] McAleer, M. and Medeiros, M. C. (2008). Realized Volatility: a Review. *Econometric Reviews* **27**, 10-45.

[77] McCulloch, R. E. and Tsay, R. S. (2001). Nonlinearity in High-Frequency Financial Data and Hierarchical Models. *Studies in Nonlinear Dynamics and Econometrics* **5**, 1-17.

[78] Merton, R. (1974). On the Pricing of Corporate Debt: the Risk Structure of Interest Rates. *Journal of Finance* **29**, 449-470.

[79] Merton, R. (1976). Option Pricing when Underlying Stock Returns are Discontinuous. *Journal of Financial Economics* **3**, 125-144.

[80] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics* **21**, 1087-1092.

[81] Nadaraya, E. (1964). On Estimating Regression. *Theory of Probability and its Applications* **10**, 186-196.

[82] OpenMP Architecture Review Board. (2008). OpenMP Application Program Interface, 3.0 Edition. URL http://www.openmp.org.

[83] Overbeck, L. (1998). Estimation for Continuous Branching Processes. *Scandinavian Journal of Statistics* **25**, 111-126.

[84] Pacurar, M. (2008). Autoregressive Conditional Duration Models in Finance: A Survey of the Theoretical and Empirical Literature. *Journal of Economic Surveys* **22**, 711-751.

[85] Payseur, S. W. (2007). A One Day Comparison of Realized Variance and Covariance Estimators. Working Paper.

[86] Payseur, S. W. (2011). Realized: Realized. *R package version 1.0.1.*. URL http://CRAN.R-project.org/package=realized.

[87] Pinsky, M. A. (1971). Limit Theorems for Continuous State Branching Processes with Immigration. *Bulletin of the American Mathematical Society* **78**, 242-244.

[88] Pitman, J. and Yor, M. (1982). A Decomposition of Bessel Bridges. *Probability Theory and Related Fields* **59**, 425-457.

[89] Pitt, M. K. and Shephard, N. (1999). Filtering via Simulation: Auxiliary Particle Filters. *Journal of the American Statistical Association* **94**, 590-599.

[90] Pitt, M. K., Silva, R. S., and Giordani, P. (2010). Auxiliary Particle Filtering within Adaptive Metropolis-Hastings Sampling. Working Paper.

[91] R Development Core Team (2011). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http: //www.R-project.org/.

[92] Schucany, W. (2004). Kernel Smoothers: an Overview of Curve Estimators for the First Graduate Course in Nonparametric Statistics. *Statistical Science* **19**, 663-675.

[93] Singleton, K. J. (2001). Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function. *Journal of Econometrics* **102**, 111-141.

[94] Stoll, H. R. (2000). Presidential Address: Friction. *The Journal of Finance* **55**, 1479-1514.

[95] Vasicek, O. (1977). An Equilibrium Characterisation of the Term Structure. *Journal of Financial Economics* **5**, 177-188.

[96] Watson, G. (1964). Smooth Regression Analysis. *Sankhya Series, A* **26**, 359-372.

[97] Wei, C. Z. and Winnicki, J. (1989). Some asymptotic results for branching processes with immigration. *Stochastic Processes and their Applications* **31**, 261-282.

[98] Wei, C. Z. and Winnicki, J. (1990). Estiamtion of the Means in the Branching Processes with Immigration. *The Annals of Statistics* **18**, 1757-1773.

[99] Welch, G. and Bishop, G. (1995). An Introduction to the Kalman Filter. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.79.6578&rep=rep1&type=pdf.

[100] 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.

[101] Zeng, Y. (2003). A Partially Observed Model for Micro-Movement of Asset Prices with Bayes Estimation via Filtering. *Mathematical Finance* **13**, 411-444.

[102] Zeng, Y. (2004). Estimating Stochastic Volatility via Filtering for the Micro-Movement of Asset Prices. *IEEE Transactions on Automatic Control* **49**, 338-348.

[103] Zhang, Q., Cai, C. X. and Keasey, K. (2009). Forecasting Using High-Frequency Data: A Comparison of Asymmetric Financial Duration Models. *Journal of Forcasting* **28**, 371-386.

[104] Zhou, C. (2001). The Term Structure of Credit Spreads with Jump Risk. *Journal of Banking and Finance* **25**, 2015-2040.