University of Alberta

Microstructure Filtering in Financial Market

by

Jianhui Huang

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

Department of Mathematical and Statistical Sciences

Edmonton, Alberta
Spring, 2008

Your file Votre référence
ISBN: 978-0-494-45448-0
Our file Notre référence
ISBN: 978-0-494-45448-0

Abstract

We first propose a nonlinear filtering problem and then derive some novel stochastic filtering equation as well as Bayes factor equations. Next, we obtain robust versions of these equations and present a novel particle filtering algorithm to implement these equations. Moreover, we discuss the fractional Ornstein-Uhlenbeck (FOU) process which is driven by the fractional Brownian motion (FBM). We aim to find some martingale problem for the historical process of FOU. As the result, we provide a sequence of discrete-parameter, computer-workable Markov chains which converge to the historical process of FOU and the martingale problems of these Markov chains are explicitly given.

The second part of our work is concerned with the following problems: Is stochastic volatility (SV) present in the microstructure stock market? If so, which of the classical SV model best represents the stock prices? We first propose some novel microstructure model which is more general and reasonable than the existing models. Indeed, our model enables us to capture most statistical properties of the price process in microstructure market such as cycles, momentum, mean-reversion as well as discreteness and clustering (biasing). With the presence of microstructure noise, it is not clear if the SV plays an essential role when modeling the stock price in market. We use Bayes estimation to show that the SV remains important and model selection to establish that Heston's model represents our stock data significantly better than

the other classical models. The prominent feature of our work is that we provide a common framework where different SV models could be tested and by Bayes factor, their performances could be evaluated in a consistent way.

Acknowledgements

Foremost, I am sincerely grateful to Professor Michael Kouritzin, the supervisor of my thesis, for his directions in the completion of this thesis. I really appreciate his constant encouragements, continuing guidance, insightful and foreseeable comments which are all invaluable assets for me. My collaboration with him is the highlight of my time spent at University of Alberta and I will cherish it the most.

I would like to thank the members of my candidacy committees for their constructive suggestions. I would like to express my special gratitude to Dr. Jia Liu for her data analysis endeavors, Dr. Biao Wu for many beneficial discussions. Mr. Calvin Chan, Surrey Kim deserve a special note of thanks for providing a warm and friendly working environments.

My heartfelt gratitude goes to my parents and brother for all their unfailing love and support over the past years. They encouraged and assisted my efforts from the very beginning. They always believe me and stand by me. My deep appreciation to them, is inexpressible in any language.

Finally and most importantly, I would like to thank my dear and wonderful wife, Dr. Amory Li. My deepest appreciation goes to her for her unconditional understanding, enormous amount of patience and consistent support in all circumstances, for sharing my moments of joy. My life has been enriched by her presence.

Contents

1 Overview **1**

1.1 Background and Motivation 1

1.2 Contribution and Outline 4

2 The Nonlinear Filtering and Bayesian Model Selection **8**

2.1 Notation and Assumption 8

2.2 The DMZ Equation and Regular Filtering 11

2.3 The Evolution Equation and Robust Filtering 21

2.4 Bayes Model Selection 33

2.5 Robust Bayes Model Selection 36

2.6 Particle Filtering to Bayesian Estimation and Model Selection 39

2.6.1 Particle Filtering to Bayesian Estimation 39

2.6.2 Particle Filtering to Model Selection 44

2.7 Historical Process and Long Memory Process 46

2.7.1 Fractional Ornstein-Uhlenbeck Process 46

2.7.2 Historical Process of FOU 51

3 Microstructural Stock Market **57**

3.1 Introduction 57

3.2 The Partially-Observed Microstructure Model 58
3.2.1 Construction of the State Process 58
3.2.2 Stochastic Volatility Models 59
3.2.3 Construction of the Observation Process 63
3.3 Nonlinear Filtering and Bayes Estimation 72
3.3.1 Numerical Results 82
3.4 Bayes Model Selection of Stochastic Volatility Models 88
3.4.1 Bayes Factor and Model Selection 88
3.4.2 Numerical Results 90

Bibliography **92**

Chapter 1

Overview

1.1 Background and Motivation

Stochastic filtering theory is of paramount importance to many fields such as mathematical finance, electronic engineering, environmental and geographical science, as well as biostatistics. In general, the filtering problem can be formulated through a pair of processes (X, Y), where X represents the state of the system while Y denotes its observations. (X, Y) constitutes the state-observation model which typically exhibits a strongly nonlinear property. The primary goal of the filtering problem is to estimate X given the information obtained by Y up to the current time. The stochastic filtering theory provides a sequential Bayesian estimation for the unknown state X when they are partially observed with noisy distorted partial observations Y.

In the last two decades, the study of microstructure market has undergone a tremendous growth from both the academic circle and the finance industry. The research activities mainly focus on the impacts of trading mechanism to the price behaviors. However, the comparisons of competing structure models raised in microstructure market are also investigated. Much of the emphasis has been focused on issues such as: the actual behavior of short-run market; the institutional and structural influences on market dynamics; the impacts of trader's heterogeneity to the price formation process; the major statistical properties of the microstructural data set, etc. These issues have important implications to the market regulation and for the design of trading mechanisms. One of the strongest attractions to the microstructure market models is the availability of high (medium)-frequency data sets which contain complete trading activities of the market participants. It contains all the relevant information about each trade such as transaction times, transaction prices and trading volumes. Such data sets make possible an unprecedented view of the detailed investigation of financial markets. One advantage of our work is that we study the high-frequency data in microstructure market within a continuous-time econometric modeling setup. Compared to the traditional discrete time-series setup, our set up provides more important insights into the functioning of microstructure market. In fact, as discussed in Hasbrouck (1991, 1999), Engle, Russell (1998), Dufour, Engle (2000), the time series analysis of high-frequency data involves an undesirable information loss as it ignores somewhat the evolution of the market that are economically valuable. For example, the time intervals between the trades are informative but they are ignored in the time series analysis. In addition, most practical applications of microstructure models such as volatility measuring and optimal order strategy design also prefer the continuous and real-time setup.

The motivation of this thesis stems from the price formation of microstructure markets to model high (medium)-frequency data. For securities traded in the microstructure market, a market price can always be directly observed at every transaction. However, such price deviates from the underlying intrinsic value process, namely, it is contaminated by the trading noises. The trading noises maybe be defined as a random resource of the price process, which is only caused by the microstructure effects (Black (1986), Hansen, Lunde (2006), Duan, Fulop (2007), Grothe, Müller (2007)). Some of the commonly seen microstructure effects include: non-synchronous trading, the discreteness of the price, the bid-ask spreads, liquidity ratios, and asymmetric information. It follows the trading noises are closely related to the market efficiency and liquidity. Mathematically, we have:

$$Y_{t_i} = F(X_{t_i}, N_{t_i}, t_i), \tag{1.1}$$

where t_i denotes the trading time of i^{th} transaction, Y is the market price process which is based on the intrinsic value process X but corrupted by some trading noises N. The observation mechanism F is some nonlinear functional.

The above formulation (1.1) is obviously a nonlinear filtering problem, where the market price process Y is the partial observation to the latent value process X in the presence of trading noises N. This formulation is general enough to subsume many existing models:

Example 1.1. *Nelson, Vestaggard (2000) discuss the stochastic volatility (SV) models with the following additive noise observation:*

$$Y_{t_i} = h(X_{t_i}) + \delta\zeta_i, \tag{1.2}$$

where h is some known but general sensor function and $\{\delta\zeta_i\}$ is a sequence of i.i.d Gaussian random variables with variance δ.

Example 1.2. *Duffie and Lando (2001) bridge some connections between the structure model and reduced-form model in default risk analysis with the following noisy accounting observation:*

$$\ln Y_{t_i} = \ln X_{t_i} + \delta\zeta_i. \tag{1.3}$$

where $\{\delta\zeta_i\}$ is a sequence of i.i.d Gaussian random variables with variance δ.

Example 1.3. *Duan, Fulop (2007) discuss structure credit risk models with noisy observations:*

$$\ln Y_{t_i} = \ln BS(X_{t_i}, t_i; F, T, \sigma) + \delta\zeta_i, \tag{1.4}$$

where $\{\delta\zeta_i\}$ is also a sequence of i.i.d Gaussian random variables with variance δ, $BS(\cdot)$ is the Black-Scholes (BS) option pricing function parameterized by the face value F, maturity T and volatility σ.

The above works strongly suggest the existence of trading noise in financial market. In particular, the data analysis of Duan, Fulop (2007) demonstrates that the ignorance of the trading noise will induce a severe bias to the estimate of volatility (see "3M" discussed in Duan, Fulop (2007)). The biasing effect of trading on volatility estimation is also empirically tested by Hansen, Lunde (2006) when using intra-day data.

It is remarkable that in above examples, the observation noise terms $N = \{\delta\zeta_i\}$ are all assumed to be a sequence of i.i.d Gaussian random variables. However, the empirical evidence reported by Hansen and Lunde (2006) strongly suggests that the noise in fact be serially correlated. This finding is the motivation to the work of Aït-Sahalia, Mykland and Zhang (2005) in which the estimator proposed is designed to be consistent with the assumption of serially correlated trading noise. Meanwhile, the kernel approach of Barndorff-Nielsen, Hansen, Lunde and Shephard (2007) also implies the dependent noise structure. Indeed, there do exist such situations in which the estimate to trading noise variance δ turns out to zero when the trading noises are assumed to be independent (see "American Express" discussed in Duan, Fulop (2007)). Of course, this doesn't mean there is no trading noise in market. Instead, this is because the trading noises are actually autocorrelated and as a result, it is absorbed into the volatility estimation of the state process. In fact, as discussed in Duan, Fulop (2007): "...the times series structure of the trading noises makes it indistinguishable from the underlying state process dynamics...". To characterize the correlation of noise, Hansen, Lunde (2006) propose some new model where the noises are assumed to be Gaussian random sequence but with stationary covariance and finite dependence. However, this model is actually rooted from the discrete-time and low-frequency setup thus failed to explain many other important statistical features such as: momentum, mean-reversion, as well as discreteness and clustering of price process.

1.2 Contribution and Outline

In this thesis, we propose some novel yet realistic models to describe the price formation in microstructure market. The models we proposed can capture the statistical properties such as mean-reversion, cycle, momentum, as well as the discreteness and clustering of the price. From the trading viewpoint: the mean-reversion refers to the market tendency of getting dragged away from the short term fads and back to economic reality; the cycle refers to the rise and fall of stock prices according to the various time evolutions; the momentum measures the trending of the stock prices and indicates the strength and weakness in the stock prices. Some plain words to the momentum is as follows: "*when the market decides to get greedy or fearful, it stays that way for a while*".

To formulate the model, we specify three types of trading noises in microstructure market: information noise, discrete noise and biasing (clustering) noise. In particular, the information noise corrupts the intrinsic value process X at trading time t_i in the following way:

$$\ln \mathcal{Y}_{t_i} = \ln X_{t_i} + Z_{t_i}^h + \epsilon\zeta_i,$$
$$dZ_t^h = -\alpha_Z Z_t^h dt + dW_t^h.$$

Here, $\mathcal{Y}_{t_i}$ is some intermediate price in the price formation and the (information) noise term N consists of two parts: the independent noise part $\{\epsilon\zeta_i\}$ is a sequence of zero mean Gaussian white noise with standard deviation ϵ thus it coincides with the traditional models; the correlation noise part Z^h is assumed to be some FOU process driven by the FBM W^h with the Hurst index $h \in [\frac{1}{2}, 1)$. It includes the traditional Brownian motion as a special case ($h = \frac{1}{2}$). The information noise term can characterize the correlation, seasonality and momentum in microstructure prices well. The intermediate price $\mathcal{Y}_{t_i}$ can be transferred to the actual market price Y_{t_i} when two other trading noises are added: the discrete and biasing (clustering) noises. In particular, these noises take into account tick levels and price biasing to e.g. whole dollar trades. Roughly speaking, it can be formulated by the equation of the type like:

$$\begin{aligned}
Y_{t_i} = {} & \xi_1 \cdot M \left[\frac{\mathcal{Y}_{t_i} + \frac{M}{2}}{M}\right] + (1-\xi_1)\xi_2 \cdot 5M \left[\frac{\mathcal{Y}_{t_i} + \frac{5M}{2}}{5M}\right] \\
& + (1-\xi_1)(1-\xi_2)\xi_3 \cdot 25M \left[\frac{\mathcal{Y}_{t_i} + \frac{25M}{2}}{25M}\right] \\
& + (1-\xi_1)(1-\xi_2)(1-\xi_3)\xi_4 \cdot 50M \left[\frac{\mathcal{Y}_{t_i} + \frac{50M}{2}}{50M}\right] \\
& + (1-\xi_1)(1-\xi_2)(1-\xi_3)(1-\xi_4) \cdot 100M \left[\frac{\mathcal{Y}_{t_i} + \frac{100M}{2}}{100M}\right],
\end{aligned}$$

where $[x]$ denotes the largest integer no more than x; $M = \frac{1}{100}$ denotes the penny tick level and characterizes the discreteness of the price process; $\xi_1, \xi_2, \xi_3, \xi_4$ are independent Bernoulli random variables which are introduced to describe the clustering of price process.

The above formulation acts as the starting point of this thesis. Based on it, we discuss the Bayes estimation and Bayes model selection problems in microstructure markets. As the mathematical preliminary, we first propose a filtering problem in continuous-time context where the state process can be characterized by the general martingale problem while the observation process is some marked point process. Moreover, we derive the stochastic equations to govern the evolution of the optimal filter and Bayes factor. Next, we derive the robust versions of these equations for numerical implementations. In addition, we introduce the historical process of the FOU process which is driven by the FBM. We then discuss the related martingale problems of historical FOU. It is of importance to filtering theory.

Generally speaking, the contribution of this thesis can be summarized under the following headings:

- Microstructure model: First to bring the long memory (LM) into microstructure; first bring the mean-reverting into microstructure; novel microstructure price model; unlimited tick levels and non-negative prices.
- Stock market: First to make some comparisons of SV models in the presence of trading noises; make the Bayes estimation through SV models.
- New stochastic filtering and Bayes factor equations; robust filtering and Bayes factor equations; novel representations in random measure.
- New particle filtering method to implement the above equations.
- Discuss the historical process of FOU process and its martingale problem; some approximating Markov chains.

Accordingly, more details and thesis outline are given as follows:

Chapter 2 introduces a class of partially-observed models in which the observation is a marked point process while the state is framed in martingale problem. Based on this model, we study the related Bayes estimation problem via nonlinear filtering and the model selection problem via Bayes factor. Accordingly, we derive the stochastic equations for the Bayes filter and Bayes factor. Furthermore, using the "gauge transform" and "path-dependent probability change", we also obtain the robust evolution equations for the Bayes filter and Bayes factor. These robust equations can be updated sequentially with reasonable computation cost and they are practically workable in the presence of modeling errors. Moreover, we propose a new particle filter algorithm to implement these evolution equations approximately and its convergence is also proved mathematically. In addition, we also discuss the historical process of the FOU process. In this study, we propose a sequence of discrete-time

Markov chains which converge weakly to the historical process of FOU. We also derive the corresponding martingale problems satisfied by these Markov chains which approximate the martingale problem to the historical process of FOU. Our result can be used directly in nonlinear historical filtering and its applications in financial econometrics are also discussed later.

Chapter 3 investigates a new class of partially-observed, trading-noise-corrupted microstructure models where the traditional SV models are the value process. The trading noise has three components: multiplicative information noises, rounding (discrete) noise and clustering noise, all chosen to reflect observed artifacts in real price data. The central questions addressed here are: (i) Is SV still discernable in the presence of trading noise microstructure? (ii) If so, which traditional SV models matches the observed price data best? We consider six popular SV models and allow both momentum and cycles in our trading noise. Bayes factor methods are utilized instead of likelihood methods in order to consider volatility models with very different structures. Nonlinear filtering techniques are used to provide a convenient means to compute the Bayes factor on tick-by-tick data while estimating the unknown parameters. With few exceptions, our stock price data exhibits strong evidence of Heston-type volatility.

Chapter 2

The Nonlinear Filtering and Bayesian Model Selection

2.1 Notation and Assumption

Without loss of generality, this thesis focuses on the fixed time period $[0, T]$ for some $T > 0$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space on which $\mathcal{N}$ denotes all its null sets. For any stochastic process S, its augmented natural filtration is defined as

$$\mathcal{F}_t^S \triangleq \sigma\{S_u : 0 \leq u \leq t\} \vee \mathcal{N}.$$

For semi-martingales M and N, $[M, N]$ denotes their cross variation or bracket process; $\langle M, N \rangle$ denotes their predictable quadratic covariation process whenever it is well defined (e.g., when M, N are locally square integrable). For any Polish space K, let $\mathcal{B}(K)$ be its Borel σ-field and $B(K)$ the set of all bounded measurable functions on it; $D_K[0, \infty)$ the space of all right continuous and left limit (RCLL) functions defined on $[0, \infty)$; $D_K[0, T]$ is defined similarly for $T > 0$; $\mathcal{M}_K^c$ the space of all finite counting measures on K endowed with the Prohorov metric. $\mathbb{N}_0$ denotes the set of all nonnegative integers. For any measurable family of pregenerators $\{\mathbf{A}_t\}_{t \geq 0}$ on Polish space K, denote $\mathcal{D}(\mathbf{A}) \subset B(K)$ as its common domain and $\mathcal{R}(\mathbf{A}) \subset B(K)$ as its range. Now suppose the signal process $X \in \mathbb{R}^{n_x}$ and its parameter $\theta \in \mathbb{R}^{n_\theta}$ jointly satisfy the following martingale problem:

Definition 2.1. *(X, θ) is the unique solution of $\mathbb{R}^{n_x + n_\theta}-$valued martingale problem* *for* $\mathbf{A}_t$ *with initial distribution* μ. *That is,* $\mu = \mathbb{P} \circ (X_0, \theta_0)^{-1}$ *and*

$$M_t^f = f(X_t, \theta_t) - f(X_0, \theta_0) - \int_0^t \mathbf{A}_s f(X_s, \theta_s) ds \tag{2.1}$$

is a $\mathcal{F}_t^{X,\,\theta}$*−martingale for each* $f \in \mathcal{D}(\mathbf{A})$. *Moreover, if* $(\widetilde{X},\, \widetilde{\theta})$ *also satisfies* (2.1), *then* $(X,\, \theta)$ *and* $(\widetilde{X},\, \widetilde{\theta})$ *have the same finite dimensional distributions.*

A 1. $\mathcal{D}(\mathbf{A})$ *is closed under multiplication.*

Remark 2.1. *We can discuss the martingale problem in a general Polish space* K *but here we confine ourself to* $K = \mathbb{R}^{n_x + n_\theta}$. *Likewise, we can discuss the martingale problem for* (X, θ) *in a filtration larger than* $\mathcal{F}_t^{X,\,\theta}$.

The observation $\overrightarrow{Y}$ is a marked point process: a double sequence of random variables $(T_n, Z_n,\ n \geq 1)$ where the increasing sequence $\{T_n\}_{n\geq 1} \in [0, T]$ denotes the jump times and $\{Z_n\}_{n\geq 1} \in E$ some attributes of T_n. Here, the measurable space $(E, \mathcal{E})$ is called the mark space. Throughout this thesis, we assume

$$E = \mathbb{N}_0$$

and $\mathcal{E}$ is its discrete $\sigma-$algebra, namely, the $\sigma-$algebra generated by all its subsets. For each $A \in \mathcal{E}$, we can associate the counting measure process $Y_t(A)$, defined as

$$Y_t(A) \triangleq \sum_{n\geq 1} 1_{\{Z_n \in A\}} 1_{\{T_n \leq t\}}. \tag{2.2}$$

In particular, if $A = E$, then we get the underlying total-jump point process:

$$Y(t) \triangleq Y_t(E) = \sum_{n\geq 1} 1_{\{T_n \leq t\}}, \tag{2.3}$$

and for $\forall j \in E$,

$$Y_j(t) \triangleq Y_t(\{j\}) = \sum_{n\geq 1} 1_{\{Z_n = j\}} 1_{\{T_n \leq t\}}. \tag{2.4}$$

This is a generalization of the "multivariate point process" introduced in Brémaud (1981, Page 234) in the sense that we have an infinite collection of marks.

Remark 2.2. *Hereafter, to be consistent with the convention, we use "z" instead* *"j" to denote the element of E.*

Equivalently, we can introduce the random counting measure $Y(dz \times dt)$ on $E \times [0, T]$ by

$$(2.5) \quad Y(\omega, A \times (s, t]) \triangleq Y_t(\omega, A) - Y_s(\omega, A), \qquad \forall \omega \in \Omega,\ A \in \mathcal{E},\ 0 \leq s \leq t \leq T,$$

and for each $t \in [0, T]$, the random counting measure $Y(dz, t)$ on E by

$$(2.6) \qquad Y(\omega, A, t) \triangleq Y_t(\omega, A), \qquad \forall \omega \in \Omega, \quad A \in \mathcal{E}.$$

The random counting measure $Y(dz \times dt)$ and the marked point process $\overrightarrow{Y}$ are equivalent (both being called the observation) as they carry the same statistical information. Moreover, $Y(dz \times dt)$ is a transition kernel from $(\Omega, \mathcal{F})$ into $E \times [0, T]$.

Remark 2.3. *By convention, we interchange the notations between $Y(dz \times dt)$ and $Y(dz, dt)$ provided there is no confusions.*

The natural filtration of $\overrightarrow{Y}$ is defined by

$$(2.7) \qquad \mathcal{F}_t^{\overrightarrow{Y}} \triangleq \sigma(Y_s(A); \quad A \in \mathcal{E}, \quad 0 \leq s \leq t).$$

We introduce the filtration

$$(2.8) \qquad \begin{aligned} \mathcal{F}_t &\triangleq \mathcal{F}_\infty^{X,\theta} \vee \mathcal{F}_t^{\overrightarrow{Y}}, \\ \mathcal{Q}_t &\triangleq \mathcal{F}_t^{X,\theta} \vee \mathcal{F}_t^{\overrightarrow{Y}}, \end{aligned}$$

where for $\mathcal{F}_1, \mathcal{F}_2 \subset \mathcal{F}$, $\mathcal{F}_1 \vee \mathcal{F}_2 = \sigma\{\mathcal{F}_1, \mathcal{F}_2\}$, the $\sigma-$algebra generated by $\mathcal{F}_1, \mathcal{F}_2$.

Now, we make the following assumption:

A 2. *For $z \in E$, $Y_z(t)$ admits a $(\mathbb{P}, \mathcal{F}_t)$-stochastic intensity $\lambda_z(X_t, \theta_t, t)$ where $\lambda_z(x, \theta, t)$ is some nonnegative measurable function of (x, θ, t) and abbreviated as $\lambda_z(t)$.*

Remark 2.4. *Without loss of generality, we assume $\lambda_z(X_t, \theta_t, t)$ is predictable for each $z \in E$. For existence and construction of this predictable version, the reader may refer Brémaud (1981), Page* 31.

Note that $\mathcal{Q}_t \subset \mathcal{F}_t$, so $\lambda_z(X_t, \theta_t, t)$ is also the $(\mathbb{P}, \mathcal{Q}_t)$-stochastic intensity of $Y_z(t)$. Now we can define the measure value process $\lambda_t(dz)$ as

$$\lambda_t(\{z\}) = \lambda_z(t). \tag{2.9}$$

Remark 2.5. *If* A2 *holds true, then set* $A = E$*, we know the total-jump point process* $Y(t)$ *admits a total intensity* $a(X_t, \theta_t, t)$ *satisfying*

$$a(X_t, \theta_t, t) = \lambda_t(\{E\}) = \int_E \lambda_t(dz) = \sum_{z=0}^{\infty} \lambda_z(X_t, \theta_t, t). \tag{2.10}$$

To simplify the notation, we sometimes write $a(t)$ instead of $a(X_t, \theta_t, t)$ when there is no confusion. For further deductions, we invoke the following assumptions:

A 3. *There exist constants* $C_1, C_2, C_3, \varepsilon > 0$ *such that for all* x, θ, t,

$$C_1 \leq a(x, \theta, t) \leq C_2, \quad \sum_{z=0}^{\infty} \frac{1}{(\lambda_z(x, \theta, t))^{\varepsilon}} \leq C_3. \tag{2.11}$$

Based on representation (2.1), (2.2), $(X, \theta, \overrightarrow{Y})$ is framed into a nonlinear filtering model, where (X, θ) is the signal process which is partially observed through the marked point process $\overrightarrow{Y}$ which is of infinite dimensionality. The rest of this chapter proceeds as follows: Section 2 studies the regular filtering problem. We derive the Duncan-Mortensen-Zakai (DMZ) equations which is driven by the observation process. In Section 3, we discuss the robust filtering problem and derive the corresponding robust filtering equation. Section 4, 5 investigate the model selection problem and we obtain the DMZ and robust evolution equations for the Bayes factor. In Section 6, we provide a novel efficient particle filtering algorithm for numerical implementation. Section 7 introduces the historical process of FOU and studies its martingale problems.

2.2 The DMZ Equation and Regular Filtering

The available information of (X, θ) is the observation filtration $\mathcal{F}_t^{\overrightarrow{Y}} \subset \mathcal{F}_t$ and the primary goal of nonlinear filtering is to characterize the Bayesian estimation (posterior distribution)

$$\pi_t(\cdot) = \mathbb{P}[(X_t, \theta_t) \in \cdot \, | \mathcal{F}_t^{\vec{Y}}] \tag{2.12}$$

or equivalently,

$$\pi_t(f) = \mathbb{E}[f(X_t, \theta_t) | \mathcal{F}_t^{\vec{Y}}] \tag{2.13}$$

for $f \in B(\mathbb{R}^{n_x+n_\theta})$. We need to characterize the normalized filter π_t recursively, that is, to find some stochastic filtering equation satisfied by the measure-valued process π_t. This will enable us to sequentially update our Bayes estimates, which is just the posterior distribution. Now suppose for each $z \in E$, $\kappa_z > 0$ is some constant such that

$$\kappa \triangleq \sum_{z=0}^{\infty} \kappa_z = \int_E \kappa_z m(dz) < \infty, \tag{2.14}$$

where $m(dz)$ is the unit counting measure assigned on $(E, \mathcal{E})$, that is

$$m(\{z\}) = 1 \quad \text{for} \quad z \in E.$$

Now consider the continuous-time likelihood function

$$L_t = \exp\left(\int_0^t \int_E \ln \frac{\lambda_z(s)}{\kappa_z} Y(dz, ds) - \int_0^t (a(s, X_s, \theta_s) - \kappa) \, ds \right), \tag{2.15}$$

which by Itô formula satisfies the stochastic integral equation

$$L_t = 1 + \int_0^t \int_E \left(\frac{\lambda_z(s)}{\kappa_z} - 1 \right) L_{s-} \left(Y(dz, ds) - \kappa_z m(dz) ds \right). \tag{2.16}$$

or equivalently the stochastic differential equation

$$dL_t = \int_E \left(\frac{\lambda_z(s)}{\kappa_z} - 1 \right) L_{t-} \left(Y(dz, dt) - \kappa_z m(dz) dt \right). \tag{2.17}$$

Proposition 2.1. *Under* A 2, A 3, L_t^{-1} *is a* $(\mathbb{P}, \mathcal{F}_t)$*-martingale.*

Proof.

$$L_t^{-1} = \exp\left(\int_0^t \int_E \ln \frac{\kappa_z}{\lambda_z(s)} Y(dz, ds) + \int_0^t (a(s, X_s, \theta_s) - \kappa)\, ds\right)$$

which satisfies

$$L_t^{-1} = 1 + \int_0^t \int_E \left(\frac{\kappa_z}{\lambda_z(s)} - 1\right) L_{s-}^{-1} \left(Y(dz, ds) - \lambda_s(dz)ds\right).$$

Thus L_t^{-1} is a $(\mathbb{P}, \mathcal{F}_t)$-local martingale and supermartingale. Moreover,

$$\sum_{z=0}^{\infty} \left(\frac{\kappa_z}{\lambda_z(t)}\right)^{1+\varepsilon} \cdot \frac{\lambda_z(t)}{a(t, X_t, \theta_t)} \leq \frac{\kappa^{1+\varepsilon}}{C_1} C_3, \tag{2.18}$$

$$\int_0^t a(s, X_s, \theta_s)ds \leq C_2 t, \tag{2.19}$$

where C_1, C_2, C_3 are introduced in A 3. Meanwhile, note that $L_0 = 1$ and for each $\delta > 0$, we have

$$\begin{aligned} \mathbb{E}\left(L_0 \exp\left(\delta Y(T)\right)\right) &= \mathbb{E}\left(\mathbb{E}(\exp\left(\delta Y(T)\right)|\mathcal{F}_0)\right) \\ &= \mathbb{E}(\exp((e^\delta - 1) \int_0^T a(s, X_s, \theta_s)ds)) \\ &< \exp((e^\delta - 1)C_2 T) < \infty. \end{aligned} \tag{2.20}$$

Note that here, $Y(T)$ is the ending point of total-jump point process $Y(t)$. Now we can apply the result from Brémaud (1981, T11, Page 242) by setting

$$K_1 = \kappa^{1+\varepsilon} \cdot \frac{C_3}{C_1},$$
$$K_2 = 0,$$
$$B(t) = C_2 t.$$

Then combining (2.18), (2.19), (2.20), we have $\mathbb{E}^{\mathbb{P}} L_T^{-1} = 1$ and this implies L_t^{-1} is a $(\mathbb{P}, \mathcal{F}_t)$-martingale on $[0, T]$. □

Consequently, we can define the reference measure $\mathbb{Q}$ by

$$L_t = \frac{d\mathbb{P}}{d\mathbb{Q}}\Big|_{\mathcal{F}_t} \tag{2.21}$$

Proposition 2.2. *Assume* A2, A3, *then under* $\mathbb{Q}$, $(Y_0(t), Y_1(t), \cdots)$ *are independent Poisson processes with respectively the intensities* $(\kappa_0, \kappa_1, \cdots)$. $(Y_0(t), Y_1(t), \cdots)$ *are also independent of* (X, θ). *Moreover, the probability distribution of* (X, θ) *is unchanged under measure* $\mathbb{Q}$.

Proof. From the property of the intensity of point process, we know,

$$\widetilde{Y}(A, t) \triangleq \int_0^t \int_A Y(dz, ds) - \int_0^t \int_A \lambda_s(dz) ds$$

is a ($\mathbb{P}$, $\mathcal{Q}_t$)-martingale and we can introduce the random measure $\widetilde{Y}(dz, ds)$ accordingly. Based on it, we can apply the integration by parts to get

$$\widetilde{M}_t^g = g(\overrightarrow{Y}_t) - g(\overrightarrow{Y}_0) - \int_0^t \mathbf{B}_s g(\overrightarrow{Y}_s) ds \tag{2.22}$$

is a ($\mathbb{P}$, $\mathcal{Q}_t$)-martingale with

$$\mathbf{B}_t g(\eta) = \int_E \left(g(\eta + \delta_z) - g(\eta)\right) \lambda_t(dz), \tag{2.23}$$

where δ_z is the Dirac measure at z. The common domain $\mathcal{D}(\mathbf{B})$ of $\mathbf{B}_t$ consists of all functions on $\mathcal{M}_E^c$ with the follow form:

$$g(\eta) = \prod_{i=1}^n \eta(g_i) \tag{2.24}$$

with $\eta \in \mathcal{M}_E^c$ and $\{g_i\}_{i=1}^n$ being any continuous, bounded functions on E and $n \in \mathbb{N}$. Combining with the martingale problem (2.1), we know the joint martingale problem for $(X, \theta, \overrightarrow{Y})$ takes the following form:

(2.25)

$$\widetilde{M}_t^{fg} = (fg)(X_t, \theta_t, \overrightarrow{Y}_t) - (fg)(X_0, \theta_0, \overrightarrow{Y}_0) - \int_0^t \left(f\mathbf{B}g + g\mathbf{A}f\right)(X_s, \theta_s, \overrightarrow{Y}_s) ds$$

is a $(\mathbb{P}, \mathcal{Q}_t)$-martingale for $f \in \mathcal{D}(\mathbf{A}), g \in \mathcal{D}(\mathbf{B})$. Here, we omit the subscripts of time t in test functions to simplify the notations. Note that $\lambda_t(\{z\}) = \lambda_z(X_t, \theta_t, t)$, so the martingale problem (2.23), (2.25) both depend on the underlying state process (X, θ). Now, we discuss the joint martingale problem of $(X, \theta, \overrightarrow{Y})$ under $\mathbb{Q}$.

From Proposition 2.1, L_t^{-1} is also a $(\mathbb{P}, \mathcal{Q}_t)$-martingale satisfying:

$$dL_t^{-1} = \int_E \left(\frac{\kappa_z}{\lambda_z(t)} - 1 \right) L_{t-}^{-1} \widetilde{Y}(dz, dt). \tag{2.26}$$

Apply the Girsanov-Meyer theorem (Protter (1990), Page 109), we know

$$\begin{aligned} &\widetilde{Y}(A,t) - \int_0^t \frac{1}{L_s^{-1}} d[L_s^{-1}, \widetilde{Y}(A,s)] \\ &= \int_0^t \int_A Y(dz, ds) - \int_0^t \int_A \lambda_s(dz) ds - \int_0^t \int_A \left(\frac{\kappa_z}{\lambda_z(s)} - 1 \right) \frac{L_{s-}^{-1}}{L_s^{-1}} Y(dz, ds) \\ &= \int_0^t \int_A Y(dz, ds) - \int_0^t \int_A \lambda_s(dz) ds - \int_0^t \int_A \left(1 - \frac{\lambda_z(s)}{\kappa_z} \right) Y(dz, ds) \\ &= \int_0^t \int_A \frac{\lambda_z(s)}{\kappa_z} Y(dz, ds) - \int_0^t \int_A \lambda_s(dz) ds \end{aligned} \tag{2.27}$$

is a $(\mathbb{Q}, \mathcal{Q}_t)$-martingale. Consequently,

$$\int_0^t \int_A Y(dz, ds) - \int_0^t \int_A \kappa_z m(dz) ds \tag{2.28}$$

is a also $(\mathbb{Q}, \mathcal{Q}_t)$-martingale. On the other hand, note that L_t^{-1} is a finite variation process and $f(X_t, \theta_t)$ is RCLL, thus

$$[f(X_t, \theta_t),\ L_t^{-1}] = \sum_{0 \le s \le t} \Delta f(X_s, \theta_s) \cdot \Delta L_s^{-1}. \tag{2.29}$$

However, the assumption A 2 implies $f(X_s, \theta_s)$ and L_t^{-1} have no simultaneous jumps almost surely (see Frey, Runggaldier (2001)), therefore, we have

$$[f(X_t, \theta_t),\ L_t^{-1}] = 0. \tag{2.30}$$

Combining (2.30) and (2.1), we have

$$[M_t^f,\ L_t^{-1}] = 0. \tag{2.31}$$

Note that M_t^f is bounded so from the Girsanov-Meyer Theorem (see Protter (1990), Page 109), it follows M_t^f is still a martingale under $\mathbb{Q}$. This tells us the martingale problem for (X, θ) remains unchanged thus the distribution of (X, θ) remains unchanged too. Now we know the joint martingale problem for $(X, \theta, \overrightarrow{Y})$ under $\mathbb{Q}$ takes the form:

(2.32)
$$\widetilde{M}_t^{fg} = (fg)(X_t, \theta_t, \overrightarrow{Y}_t) - (fg)(X_0, \theta_0, \overrightarrow{Y}_0) - \int_0^t (f\mathbf{B}^* g + g\mathbf{A}f)(X_s, \theta_s, \overrightarrow{Y}_s) ds$$

for $f \in \mathcal{D}(\mathbf{A})$, $g \in \mathcal{D}(\mathbf{B}^*)$ where $\mathcal{D}(\mathbf{B}^*) = \mathcal{D}(\mathbf{B})$ and

$$\mathbf{B}^* g(\eta) = \int_E (g(\eta + \delta_z) - g(\eta)) \kappa(z) m(dz). \tag{2.33}$$

Note that (2.33) does not depend on the underlying process (X, θ). Meanwhile, from the uniqueness of the martingale problem of (X, θ) and the construction of Y, we know the joint martingale problem of (X, θ, Y) is also unique. As a result, (X, θ) is independent of $\overrightarrow{Y}$ under $\mathbb{Q}$. □

Bayes theorem links π_t with the unnormalized filter σ_t by

$$\pi_t(f) = \frac{\sigma_t(f)}{\sigma_t(1)} \tag{2.34}$$

for $f \in B(\mathbb{R}^{n_x + n_\theta})$, where

$$\sigma_t(f) \triangleq \mathbb{E}^{\mathbb{Q}}[f(X_t, \theta_t) L_t | \mathcal{F}_t^{\overrightarrow{Y}}]. \tag{2.35}$$

Theorem 2.1. *Under* A 2, A 3, *the unnormalized filter σ_t is the unique solution of the measure-valued DMZ equation*

(2.36)
$$\sigma_t(f) = \sigma_0(f) + \int_0^t \sigma_s (\mathbf{A}_s f - (a(s) - \kappa) f) ds + \int_0^t \int_E \sigma_{s-} \left(\left(\frac{\lambda_z(s)}{\kappa_z} - 1 \right) f \right) Y(dz, ds)$$

for $f \in \mathcal{D}(\mathbf{A})$.

Proof. From (2.16) and the integration by parts of semi-martingales (Protter (1990), Page 60), we have

$$f(X_t,\theta_t)L_t = f(X_0,\theta_0) + \int_0^t f(X_{s-},\theta_{s-})dL_s + \int_0^t L_{s-}df(X_s,\theta_s) + [f(X,\theta),\ L]_t.$$

We have

$$[f(X_t,\theta_t),\ L_t] = \sum_{0\le s\le t} \Delta f(X_s,\theta_s)\cdot \Delta L_s. \tag{2.37}$$

Once again, A 2 implies $f(X_s,\theta_s)$ and L_t have no simultaneous jumps almost surely thus

$$[f(X_t,\theta_t),\ L_t] = 0. \tag{2.38}$$

Combining (2.38) and (2.1), we have

$$[M_t^f,\ L_t] = 0. \tag{2.39}$$

Therefore, from integration by parts, we have

$$\begin{aligned} f(X_t,\theta_t)L_t &= f(X_0,\theta_0) + \int_0^t L_{s-}dM_s^f + \int_0^t L_s\mathbf{A}_s f(X_s,\theta_s)ds \\ &+ \int_0^t f(X_{s-},\theta_{s-})\left(\int_E \left(\frac{\lambda_z(s)}{\kappa_z}-1\right)L_{s-}(Y(dz,ds)-\kappa_z m(dz)ds)\right) \\ &= f(X_0,\theta_0) + \int_0^t L_{s-}dM_s^f + \int_0^t f(X_{s-},\theta_{s-})L_{s-}\int_E\left(\frac{\lambda_z(s)}{\kappa_z}-1\right)Y(dz,ds) \\ &+ \int_0^t L_s\left(\mathbf{A}_s f(X_s,\theta_s) - f(X_s,\theta_s)\int_E\left(\frac{\lambda_z(s)}{\kappa_z}-1\right)\kappa_z m(dz)\right)ds. \end{aligned}$$

Now, taking expectation under $\mathbb{Q}$ on observation filtration $\mathcal{F}_t^{\overrightarrow{Y}}$ and by linearity, we have

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}}\left(f(X_t,\theta_t)L_t|\mathcal{F}_t^{\overrightarrow{Y}}\right) &= \mathbb{E}^{\mathbb{Q}}\left(f(X_0,\theta_0)|\mathcal{F}_t^{\overrightarrow{Y}}\right) + \mathbb{E}^{\mathbb{Q}}\left(\int_0^t L_{s-}dM_s^f|\mathcal{F}_t^{\overrightarrow{Y}}\right) \\ &+ \mathbb{E}^{\mathbb{Q}}\left(\int_0^t f(X_{s-},\theta_{s-})L_{s-}\int_E\left(\frac{\lambda_z(s)}{\kappa_z}-1\right)Y(dz,ds)|\mathcal{F}_t^{\overrightarrow{Y}}\right) \\ &+ \mathbb{E}^{\mathbb{Q}}\left(\int_0^t L_s\left(\mathbf{A}_s f - f\int_E\left(\frac{\lambda_z(s)}{\kappa_z}-1\right)\kappa_z m(dz)\right)ds|\mathcal{F}_t^{\overrightarrow{Y}}\right). \end{aligned}$$

Now for $0 \le a < b < \infty$, introduce the notation

$$\mathcal{F}^{\overrightarrow{Y}}_{(a,b]} \triangleq \sigma(Y_s(A);\ \ A \in \mathcal{E},\ \ a < s \le b).$$

Then under $\mathbb{Q}$, $\mathcal{F}^{\overrightarrow{Y}}_{(0,t]}$ is independent of $\mathcal{F}_0 = \mathcal{F}^{X,\theta}_{\infty} \vee \mathcal{F}^{\overrightarrow{Y}}_0$ thus

$$(2.40)\quad \mathbb{E}^{\mathbb{Q}}(f(X_0,\theta_0)|\mathcal{F}^{\overrightarrow{Y}}_t) = \mathbb{E}^{\mathbb{Q}}(f(X_0,\theta_0)|\mathcal{F}^{\overrightarrow{Y}}_{(0,t]} \vee \mathcal{F}^{\overrightarrow{Y}}_0) = \mathbb{E}^{\mathbb{Q}}(f(X_0,\theta_0)|\mathcal{F}^{\overrightarrow{Y}}_0) = \sigma_0(f),$$

From Proposition 2.2, we know M^f_t is still a $\mathcal{F}_t$-martingale under measure $\mathbb{Q}$ and $\overrightarrow{Y}$ is compatible with $\mathcal{F}_t$, that is, $\overrightarrow{Y}_t$ is adapted to $\mathcal{F}_t$ and $\mathcal{F}^{\overrightarrow{Y}}_{(t,t+s]}$ is independent of $\mathcal{F}_t$ for each $s > 0$. Moreover, under A 2, A 3,

$$(2.41)\qquad \mathbb{E}^{\mathbb{Q}}\left|\int_0^t L_{s-}dM^f_s\right| < \infty.$$

Therefore we have

$$(2.42)\qquad \mathbb{E}^{\mathbb{Q}}\left(\int_0^t L_{s-}dM^f_s|\mathcal{F}^{\overrightarrow{Y}}_t\right) = 0.$$

To prove (2.42), first consider a sequence of simple processes L_n which converge to L. For each L_n, the stochastic integration $\int_0^t L_n(s-)dM^f_s$ can be written as a summation and we can verify

$$\mathbb{E}^{\mathbb{Q}}\left(\int_0^t L_n(s-)dM^f_s|\mathcal{F}^{\overrightarrow{Y}}_t\right) = 0.$$

and then extend to the general case L_{s-} by approximating arguments. From Fubini's theorem, we have

$$\begin{aligned}
(2.43)\quad & \mathbb{E}^{\mathbb{Q}}\left(\int_0^t L_s\left(\mathbf{A}_s f(X_s,\theta_s) - f(X_s,\theta_s)\int_E\left(\frac{\lambda_z(s)}{\kappa_z}-1\right)\kappa_z m(dz)\right)ds|\mathcal{F}^{\overrightarrow{Y}}_t\right) \\
&= \int_0^t \mathbb{E}^{\mathbb{Q}}\left(L_s\left(\mathbf{A}_s f(X_s,\theta_s) - f(X_s,\theta_s)\int_E(\lambda_z(s)-\kappa_z)\,m(dz)\right)|\mathcal{F}^{\overrightarrow{Y}}_t\right)ds \\
&= \int_0^t \mathbb{E}^{\mathbb{Q}}\left(L_s\left(\mathbf{A}_s f(X_s,\theta_s) - f(X_s,\theta_s)\int_E(\lambda_z(s)-\kappa_z)\,m(dz)\right)|\mathcal{F}^{\overrightarrow{Y}}_s\right)ds \\
&= \int_0^t \sigma_s\left(\mathbf{A}_s f - (a(s)-\kappa)f\right)ds.
\end{aligned}$$

Here we also use the fact that $\mathcal{F}^{\overrightarrow{Y}}_{(s,t]}$ is independent of $\mathcal{F}_s$. Finally,

$$
\begin{aligned}
(2.44)\qquad & \mathbb{E}^{\mathbb{Q}}\left(\int_0^t f(X_{s-},\theta_{s-})L_{s-}\int_E\left(\frac{\lambda_z(s)}{\kappa_z}-1\right)Y(dz,ds))|\mathcal{F}_t^{\overrightarrow{Y}}\right) \\
&= \mathbb{E}^{\mathbb{Q}}\left(\sum_{n=1}^{\infty} 1_{\{T_n\le t\}}\cdot L_{T_n-}\cdot f(X_{T_n-},\theta_{T_n-})\left(\frac{\lambda_{Z_n}(T_n)}{\kappa_{Z_n}}-1\right)|\mathcal{F}_t^{\overrightarrow{Y}}\right) \\
&= \mathbb{E}^{\mathbb{Q}}\left(\lim_{N\longrightarrow\infty}\sum_{n=1}^{N} 1_{\{T_n\le t\}}\cdot L_{T_n-}\cdot f(X_{T_n-},\theta_{T_n-})\left(\frac{\lambda_{Z_n}(T_n)}{\kappa_{Z_n}}-1\right)|\mathcal{F}_t^{\overrightarrow{Y}}\right) \\
&= \lim_{N\longrightarrow\infty}\mathbb{E}^{\mathbb{Q}}\left(\sum_{n=1}^{N} 1_{\{T_n\le t\}}\cdot L_{T_n-}\cdot f(X_{T_n-},\theta_{T_n-})\left(\frac{\lambda_{Z_n}(T_n)}{\kappa_{Z_n}}-1\right)|\mathcal{F}_t^{\overrightarrow{Y}}\right) \\
&= \lim_{N\longrightarrow\infty}\sum_{n=1}^{N}\mathbb{E}^{\mathbb{Q}}\left(1_{\{T_n\le t\}}\cdot L_{T_n-}\cdot f(X_{T_n-},\theta_{T_n-})\left(\frac{\lambda_{Z_n}(T_n)}{\kappa_{Z_n}}-1\right)|\mathcal{F}_t^{\overrightarrow{Y}}\right) \\
&= \lim_{N\longrightarrow\infty}\sum_{n=1}^{N}\mathbb{E}^{\mathbb{Q}}\left(1_{\{T_n\le t\}}\cdot L_{T_n-}\cdot f(X_{T_n-},\theta_{T_n-})\left(\frac{\lambda_{Z_n}(T_n)}{\kappa_{Z_n}}-1\right)|\mathcal{F}_{T_n-}^{\overrightarrow{Y}}\right) \\
&= \lim_{N\longrightarrow\infty}\sum_{n=1}^{N}\sigma_{T_n-}\left(\left(\frac{\lambda_{Z_n}(T_n)}{\kappa_{Z_n}}-1\right)f\right) \\
&= \int_0^t\int_E \sigma_{s-}\left(\left(\frac{\lambda_z(s)}{\kappa_z}-1\right)f\right)Y(dz,ds).
\end{aligned}
$$

Here, the first and last equalities used the summation representation of the stochastic integral to random counting measure; the second equality is because for any fixed time t, there are only finite jump times T_n and the limit is in the sense of almost surely; the third equality is from the Dominated Convergence Theorem (DCT) of conditional expectation (A 3 ensures the existence of such dominated control variable); the fourth equality is just the linearity of conditional expectation; the fifth equality is because the independence of $\sigma\{\mathcal{F}_u^{\overrightarrow{Y}}, u\ge T_n\}$ and $\mathcal{F}_{T_n-}$; the sixth equality is from the definition of unnormalized filter.

Combining the above terms together, we obtain the DMZ equation (2.36). As to the solution uniqueness of (2.36), first note that the uniqueness of the martingale problem (2.1) implies the Markov property of (X,θ), therefore the semigroup of the pregenerator $\mathbf{A}_t$ exists and we denoted by $\mathbf{S}$. Recall T_n is the n^{th} jump time and we can introduce $T_0 = 0$ for convention. From the (2.36), we know

$$\sigma_t(f) = \sigma_{T_n}(f) + \int_{T_n}^{t} \sigma_s\left(\mathbf{A}_s f - (a(s) - \kappa) f\right) ds,$$

for $t \in [T_n, T_{n+1})$ and $f \in \mathcal{D}(\mathbf{A})$. Now, we can define,

$$\chi(t, u\ f) = \sigma_u(\mathbf{S}_{t-u} f) + \int_u^t \sigma_s(\mathbf{S}_{t-s} f - (a(s) - \kappa) f) ds \tag{2.45}$$

for all $f \in \mathcal{D}(\mathbf{A})$ and $u \leq t \in [T_n, T_{n+1})$. Then, from the definition of semigroup, we know

$$\frac{d\chi(t, u\ f)}{du} = \frac{d\sigma_u(\mathbf{S}_{t-u} f)}{du} + \frac{d}{du} \int_u^t \sigma_s(\mathbf{S}_{t-s} f - (a(s) - \kappa) f) ds = 0. \tag{2.46}$$

The remaining procedures are same to that of Kouritzin and Zeng (2005), Appendix A. □

The DMZ filtering equation (2.36) involves some stochastic integration so the unnormalized filter σ_t is not the robust filter and it is not easy to be implemented in real time especially when the observations are of rapid changes. In the following, using the gauge transform and path-dependent probability change, we will show how to derive the robust filter through an evolution equation parameterized by the observation path. Empirical results show that the robust filter does indeed perform favorably when applied to real data problem. Clark (1978) introduces the robust filter and some other important works on the robust filter include Davis (1980, 1981), where a semi-group approach to robust filter is proposed when the signal is Hunt process with additive white noise observation, Pardoux (1979) and Heunis (1990), where a stochastic partial differential equation approach are given when the signal is multi-dimensional diffusion. However, all of these works are discussed within the classical additive white noise observation structure and thus they are not readily applicable here.

2.3 The Evolution Equation and Robust Filtering

In the remainder of this section, we let $y(\cdot) = \{y_t,\ 0 \le t \le T\}$ denote an arbitrary but fixed observation trajectory, in other words, $y(\cdot) = Y(\cdot,\ \omega)$ for some $\omega \in \Omega$.

Definition 2.2. *For $f \in B(\mathbb{R}^{n_x+n_\theta})$, the gauge transform ν_t is defined as*

$$\nu_t(f) \triangleq \mathbb{E}^{\mathbb{Q}}\left[f(X_t,\theta_t)L_t \exp\left(-\int_E \ln\frac{\lambda_z(t)}{\kappa_z}Y(dz,t)\right)\Big|\mathcal{F}_t^{\overline{Y}}\right]. \tag{2.47}$$

It is equivalent to σ_t in the following sense,

$$\nu_t(f) = \sigma_t\left(f\exp\left(-\int_E \ln\frac{\lambda_z(t)}{\kappa_z}Y(dz,t)\right)\right),$$
$$\sigma_t(f) = \nu_t\left(f\exp\left(\int_E \ln\frac{\lambda_z(t)}{\kappa_z}Y(dz,t)\right)\right).$$

To simplify the notation, we introduce the Lie bracket for operator $\mathbf{A}_t$.

Definition 2.3. *For $f_1, f_2 \in \mathcal{D}(\mathbf{A})$, the Lie bracket for $\mathbf{A}_t$ is defined as*

$$[f_1,f_2]^t \triangleq (\mathbf{A}_t f_1 f_2 - f_1\mathbf{A}_t f_2 - f_2\mathbf{A}_t f_1). \tag{2.48}$$

The following definition is taken from K. L. Chung (1982).

Definition 2.4. *A measurable stochastic process X satisfying*

$$\lim_{n\longrightarrow\infty} X_{T_n} = X_T \qquad a.s.,$$

for any increasing sequence of stopping times $\{T_n\}_{n=1}^{\infty}$ with the limit T is said to be quasi-left-continuous.

Definition 2.5. *For any filtration $\{\mathcal{H}_t\}$, an adapted stochastic process X is said to be a $(\mathbb{P},\mathcal{H}_t)-$special semi-martingale if it admits a decomposition*

$$X = X_0 + M + A,$$

where X_0 is a $\mathcal{H}_0-$measurable random variable, M is a $(\mathbb{P},\mathcal{H}_t)-$local martingale while A is a $\mathcal{H}_t-$predictable process with finite variation over each finite interval $[0,t]$.

Lemma 2.1. *Under* A 1, *for* $f_1, f_2 \in \mathcal{D}(\mathbf{A})$, $\langle M^{f_1}, M^{f_2}\rangle$ *is well defined and*

$$\langle M^{f_1}, M^{f_2}\rangle_t = \int_0^t [f_1, f_2]^s(X_s, \theta_s)ds. \tag{2.49}$$

Moreover, for $f \in \mathcal{D}(\mathbf{A})$, M_t^f *is quasi-left-continuous.*

Proof. From (2.1), for $f \in \mathcal{D}(\mathbf{A})$, M_t^f is a martingale with $M_0^f \equiv 0$. M^f is locally bounded because f is bounded and for fixed t, $\int_0^t \mathbf{A}_s f(X_s, \theta_s)ds$ is also bounded. Therefore, from Dellacherie and Meyer (1978, Page 227), we know the angle bracket $\langle M^{f_1}, M^{f_2}\rangle$ is well defined. Next, from A 1, $f^2 \in \mathcal{D}(\mathbf{A})$ and

$$f^2(X_t, \theta_t) = f^2(X_0, \theta_0) + \int_0^t \mathbf{A}_s f^2(X_s, \theta_s)ds + M_t^{f^2} \tag{2.50}$$

for some martingale $M_t^{f^2}$. However, applying Itô's formula and note that (2.1),

$$f^2(X_t, \theta_t) = f^2(X_0, \theta_0) + \int_0^t 2f\mathbf{A}_s f(X_s, \theta_s)ds + \int_0^t 2f(X_{s-}, \theta_{s-})dM_s^f + [M^f, M^f]_t.$$

Therefore,

$$\int_0^t (2f\mathbf{A}_s f - \mathbf{A}_s f^2)(X_s, \theta_s)ds + [M^f, M^f]_t = M_t^{f^2} - \int_0^t 2f(X_{s-}, \theta_{s-})dM_s^f.$$

However,

$$K_t \triangleq [M^f, M^f]_t - \langle M^f, M^f\rangle_t$$

is a local martingale and we have

$$\int_0^t (2f\mathbf{A}_s f - \mathbf{A}_s f^2)(X_s, \theta_s)ds + \langle M^f, M^f\rangle_t = M_t^{f^2} - \int_0^t 2f(X_{s-}, \theta_{s-})dM_s^f - K_t.$$

It follows that both sides are decompositions of special semi-martingales. In particular, the left hand side is a RCLL, adapted predictable finite variation process null at zero while the right hand side is a local martingale starting from zero. Then, from the the uniqueness of decomposition to special semi-martingale (see Jacod and Shiryaev (1987), Page 32), both sides have to be constants and note that they are both starting from 0, therefore we have

$$\langle M^f, M^f \rangle_t = \int_0^t \left(\mathbf{A}_s f^2(X_s, \theta_s) - 2f\mathbf{A}_s f(X_s, \theta_s) \right) ds = \int_0^t [f, f]^s(X_s, \theta_s) ds.$$

Following from the polarization, for $f_1, f_2 \in \mathcal{D}(\mathbf{A})$,

$$\langle M^{f_1}, M^{f_2} \rangle_t = \int_0^t [f_1, f_2]^s(X_s, \theta_s) ds. \tag{2.51}$$

Moreover, M_t^f is quasi-left-continuous because $\langle M^f, M^f \rangle_t$ is absolutely continuous with respect to the Lebesgue measure for $f \in \mathcal{D}(\mathbf{A})$(see Jacod and Shiryaev (1987). Theorem 4.2). □

To derive the robust filter, we generalize the martingale problem (2.1) to the time-space version by first defining a new domain as follows:

Definition 2.6. *Define $\mathcal{D}(\mathbf{A}^*) \subset B(\mathbb{R}^{n_x+n_\theta} \times [0, \infty))$ as the set of all functions:*

$$f^*(x, \theta, t) : \mathbb{R}^{n_x+n_\theta} \times [0, \infty) \longrightarrow \mathbb{R}$$

which have bounded partial derivative $\frac{\partial f^*}{\partial t}$ and such that:

$$\forall t \in [0, \infty), \quad f_t \triangleq f^*(\cdot, \cdot, t) \in \mathcal{D}(\mathbf{A}), \tag{2.52}$$

and

$$M_t^{f^*} \triangleq f^*(X_t, \theta_t, t) - f^*(X_0, \theta_0, 0) - \int_0^t \mathbf{A}_s^* f^*(X_s, \theta_s, s) ds \tag{2.53}$$

is $\{\mathcal{F}_t^{X,\theta}\}$−martingale where

$$\mathbf{A}_t^* f^*(x, \theta, t) = \frac{\partial f^*}{\partial t}(x, \theta, t) + \mathbf{A}_t(f_t)(x, \theta). \tag{2.54}$$

This new domain is rich enough to include many time-dependent test functions. For example, if we denote $\mathcal{G}$ is the set of all absolutely continuous functions $g : [0, \infty) \longrightarrow \mathbb{R}$ with bounded derivative $g' : [0, \infty) \longrightarrow \mathbb{R}$, then on $\mathcal{D}(\mathbf{A}) \otimes \mathcal{G}$ (recall $\mathcal{D}(\mathbf{A})$ is the domain of martingale problem (2.1)), we can define the operator $\mathbf{A}_t^*$ on functions of the form $f^* = g \otimes f$ where $g \in \mathcal{G}$ and $f \in \mathcal{D}(\mathbf{A})$ by

$$\mathbf{A}_t^* f^*(x, \theta, t) = g'(t) f(x, \theta) + g(t) \mathbf{A}_t f(x, \theta) \tag{2.55}$$

for $t \in [0, T]$, $(x, \theta) \in \mathbb{R}^{n_x + n_\theta}$. In fact, apply the Itô formula, it is easy to verify that

$$M_t^{f^*} = f^*(X_t, \theta_t, t) - f^*(X_0, \theta_0, 0) - \int_0^t \mathbf{A}_s^* f^*(X_s, \theta_s, s) ds \tag{2.56}$$

is $\{\mathcal{F}_t^{X,\,\theta}\}-$martingale for $f^* \in \mathcal{D}(\mathbf{A}) \otimes \mathcal{G}$.

Remark 2.6. *Hereafter, to ease the notation, we still write* $\mathbf{A}, \mathcal{D}(\mathbf{A})$ *instead of* $\mathbf{A}^*, \mathcal{D}(\mathbf{A}^*)$ *unless other specified.*

Now we assume

A 4. $\forall z \in E$, *there exists a constant* D_z *such that* $0 < D_z \leq \lambda_z$.

A 5. $\forall z \in E$, $\lambda_z \in \mathcal{D}(\mathbf{A})$.

A 6. $\forall z \in E$, $\{M^{\lambda_z}\}_{t \geq 0}$ *is continuous martingale.*

The following theorem gives the robust filtering equation.

Theorem 2.2. *Assume* A 1 − 6 *hold true, then* $\nu_t(f)$ *satisfies the following evolution equation*

$$\nu_t(f) = \nu_0(f) + \int_0^t \nu_s(\widetilde{\mathbf{A}}_s^Y f) ds, \tag{2.57}$$

where $\mathcal{D}(\widetilde{\mathbf{A}}^Y) = \mathcal{D}(\mathbf{A})$ *and*

$$\begin{aligned} \widetilde{\mathbf{A}}_t^Y f = & \left(\mathbf{A}_t + \kappa - a(t) + \int_{E \times E} \frac{[\lambda_z, \lambda_\zeta]^t}{2\lambda_z \lambda_\zeta} Y(dz, t) Y(d\zeta, t) \right) f \\ & - \int_E \left(\frac{[\lambda_z, f]^t + 2f \mathbf{A}_t(\lambda_z)}{\lambda_z} + \frac{f}{\lambda_z} \cdot \frac{\partial \lambda_z}{\partial t} - \frac{f \mathbf{A}_t(\lambda_z^2)}{2\lambda_z^2} \right) Y(dz, t). \end{aligned} \tag{2.58}$$

Unlike the DMZ filtering equation (2.36), the evolution equation (2.57) does not involve any stochastic integration. In contrast, its randomness is only expressed through the parameterized observation path thus it is the robust equation we are seeking and ν_t becomes the robust nonlinear filter. One advantage of our robust equation is that there has only one time scale "dt" here while in (2.36), there have two time scales: "dt" and "dY_t" which will complicate the numerical computation. Meanwhile, it is easier to establish the convergence result when approximating the robust filter equation. This is especially crucial to the particle filtering method introduced in the sequel. To prove Theorem 2.2, we need some preliminary results.

Proposition 2.3. *For each $g \in C_0^2(\mathbb{R})$, the space of twice continuously differentiable functions on $\mathbb{R}$ with compact support, and $f(x, \theta, t) : \mathbb{R}^{n_x+n_\theta} \times [0,T] \longrightarrow \mathbb{R}$ such that $f(\cdot,\cdot,t) \in \mathcal{D}(\mathbf{A})$ for each $t \in [0,T]$ and M_t^f is a continuous martingale, we have $g \circ f \in \mathcal{D}(\mathbf{A})$ and*

$$\begin{aligned} \mathbf{A}_t(g(f)) &= g'(f) \cdot \left(\frac{\partial f}{\partial t} + \mathbf{A}_t f\right) + \frac{1}{2} g''(f) \cdot [f,f]^t, \\ M_t^{g(f)} &= \int_0^t g'(f)(X_s, \theta_s) dM_s^f. \end{aligned} \tag{2.59}$$

Proof. From (2.1) and Itô formula, one has that

$$dg(f(X_t,\theta_t)) = g'(f)(\frac{\partial f}{\partial t} + \mathbf{A}_t f)(X_t,\theta_t)dt + g'(f)(X_t,\theta_t)dM_t^f + \frac{1}{2}g''(f)d[M^f, M^f]_t.$$

Note that M_t^f is a continuous martingale. Then, from Lemma 2.1,

$$[M^f, M^f]_t = \langle M^f, M^f \rangle_t = \int_0^t [f,f]^s ds.$$

Thus,

$$\begin{aligned} dg(f(X_t,\theta_t)) = g'(f) &\left(\mathbf{A}_t + \frac{\partial f}{\partial t}\right) f(X_t,\theta_t)dt + \frac{1}{2}g''(f)[f,f]^t(X_t,\theta_t)dt \\ &+ g'(f)(X_t,\theta_t)dM_t^f. \end{aligned}$$

Introduce $M_t^{g(f)} = \int_0^t g'(f)(X_s,\theta_s)dM_s^f$. Because $f \in B(\mathbb{R}^{n_x+n_\theta})$ is bounded and g' is continuous, so we have $g'(f)$ is a bounded function. As a result, the continuous local martingale $M_t^{g(f)}$ is of bounded cross variation thus it is martingale. Therefore, $g \circ f \in \mathcal{D}(\mathbf{A})$ and we have (2.59). ☐

Proposition 2.4. *Assume* A $1-6$ *hold true, then for* $z \in E$, $\ln \lambda_z \in \mathcal{D}(\mathbf{A})$ *and*

$$M_t^{\ln \lambda_z} = \int_0^t \lambda_z^{-1}(X_s, \theta_s) dM_s^{\lambda_z}, \tag{2.60}$$

$$\mathbf{A}_t(\ln \lambda_z) = \frac{\mathbf{A}_t \lambda_z}{\lambda_z} + \frac{1}{\lambda_z} \cdot \frac{\partial \lambda_z}{\partial t} - \frac{[\lambda_z, \lambda_z]^t}{2\lambda_z^2}. \tag{2.61}$$

Moreover, for $f \in \mathcal{D}(\mathbf{A})$, $z, \zeta \in E$,

$$[\ln \lambda_z, \ln \lambda_\zeta]^t = \frac{[\lambda_z, \lambda_\zeta]^t}{\lambda_z \lambda_\zeta}, \tag{2.62}$$

$$[\ln \lambda_z, f]^t = \frac{[\lambda_z, f]^t}{\lambda_z}. \tag{2.63}$$

Proof. In Proposition 2.3, letting $g = \ln(x)$ and $f = \lambda_z$, we have

$$g \circ f = \ln \lambda_z \in \mathcal{D}(\mathbf{A}) \tag{2.64}$$

and

$$\begin{aligned} \mathbf{A}_t \ln(\lambda_z) &= \ln'(\lambda_z) \cdot \left(\frac{\partial \lambda_z}{\partial t} + \mathbf{A}_t \lambda_z \right) + \frac{1}{2} \ln''(\lambda_z) \cdot [\lambda_z, \lambda_z]^t \\ &= \frac{\mathbf{A}_t \lambda_z}{\lambda_z} + \frac{1}{\lambda_z} \cdot \frac{\partial \lambda_z}{\partial t} - \frac{[\lambda_z, \lambda_z]^t}{2\lambda_z^2}. \end{aligned} \tag{2.65}$$

Now we prove $(2.62), (2.63)$. First, for $f, g \in \mathcal{D}(\mathbf{A})$ satisfying

$$\mathbb{E} \int_0^t \frac{1}{f^2(X_s, \theta_s)} ds < \infty, \ \mathbb{E} \int_0^t \frac{1}{g^2(X_s, \theta_s)} ds < \infty, \tag{2.66}$$

we have

$$\begin{aligned} d \ln f(X_t, \theta_t) &= \left(\frac{1}{f} \mathbf{A}_t f + \frac{1}{f} \frac{\partial f}{\partial t} - \frac{1}{2f^2}[f, f]^t \right) (X_t, \theta_t) dt + \frac{1}{f} dM_t^f, \\ d \ln g(X_t, \theta_t) &= \left(\frac{1}{g} \mathbf{A}_t g + \frac{1}{g} \frac{\partial g}{\partial t} - \frac{1}{2g^2}[g, g]^t \right) (X_t, \theta_t) dt + \frac{1}{g} dM_t^g. \end{aligned}$$

By integration by parts, we get

(2.67)
$$d(\ln f \cdot \ln g) = \left(\ln f\left(\frac{\mathbf{A}_t g}{g} + \frac{1}{g}\frac{\partial g}{\partial t} - \frac{[g,g]^t}{2g^2}\right) + \ln g\left(\frac{\mathbf{A}_t f}{f} + \frac{1}{f}\frac{\partial f}{\partial t} - \frac{[f,f]^t}{2f^2}\right)\right) dt$$
$$+ \frac{[f,g]^t}{fg} + \left(\frac{1}{f} dM_t^f + \frac{1}{g} dM_t^g\right).$$

From (2.66) and boundness of the Lie bracket process, we have

(2.68)
$$\mathbb{E}\int_0^t \frac{[f,f]^s}{f^2}(X_s,\theta_s)ds < \infty, \ \mathbb{E}\int_0^t \frac{[g,g]^s}{g^2}(X_s,\theta_s)ds < \infty$$

for $t \in [0,T]$, it follows the last term in (2.67) is a martingale, thus we know

(2.69)
$$\ln f \cdot \ln g \in \mathcal{D}(A),$$

and

(2.70)
$$\mathbf{A}_t(\ln f \cdot \ln g) = \frac{[f,g]^t}{fg} + \ln f\left(\frac{\mathbf{A}_t g}{g} - \frac{[g,g]^t}{2g^2} + \frac{1}{g}\cdot\frac{\partial g}{\partial t}\right)$$
$$+ \ln g\left(\frac{\mathbf{A}_t f}{f} - \frac{[f,f]^t}{2f^2} + \frac{1}{f}\cdot\frac{\partial f}{\partial t}\right).$$

In (2.70), let $f = \lambda_z, g = \lambda_\zeta$, then we have

(2.71)
$$[\ln\lambda_z, \ln\lambda_\zeta]^t = \mathbf{A}_t(\ln\lambda_z \ln\lambda_\zeta) - \ln\lambda_z \cdot \mathbf{A}_t(\ln\lambda_\zeta) - \ln\lambda_\zeta \cdot \mathbf{A}_t(\ln\lambda_z)$$
$$= \frac{[\lambda_z,\lambda_\zeta]^t}{\lambda_z\lambda_\zeta}.$$

Similarly, we have

$$d(\ln\lambda_z \cdot f) = \left(\ln\lambda_z \cdot \mathbf{A}_t f + f\left(\frac{1}{\lambda_z}\mathbf{A}_t(\lambda_z) + \frac{1}{\lambda_z}\frac{\partial\lambda_z}{\partial t} - \frac{[\lambda_z,\lambda_z]^t}{2\lambda_z^2}\right) + \frac{[\lambda_z,f]^t}{\lambda_z}\right) dt$$
$$+ \left(\ln\lambda_z dM_t^f + \frac{f}{\lambda_z} dM_t^{\lambda_z}\right).$$

From A 4, we know

(2.72)
$$\int_0^t (\ln\lambda_z)^2 [f,f]^s(X_s,\theta_s)ds < \infty, \ \int_0^t \frac{f^2[\lambda_z,\lambda_z]^s}{\lambda_z^2}(X_s,\theta_s)ds < \infty,$$

thus it follows

$$\int_0^t \ln \lambda_z dM_s^f + \int_0^t \frac{f}{\lambda_z} dM_s^{\lambda_z}$$

is a martingale. Therefore, $\ln \lambda_z \cdot f \in \mathcal{D}(A)$ and

$$\mathbf{A}_t(\ln \lambda_z \cdot f) = \left(\ln \lambda_z \cdot \mathbf{A}_t f + f \left(\frac{1}{\lambda_z} \cdot \frac{\partial \lambda_z}{\partial t} + \frac{1}{\lambda_z} \mathbf{A}_t(\lambda_z) - \frac{[\lambda_z, \lambda_z]^t}{2\lambda_z^2} \right) + \frac{[\lambda_z, f]^t}{\lambda_z} \right).$$

It follows

$$[\ln \lambda_z, f]^t = \mathbf{A}_t(\ln \lambda_z f) - \ln \lambda_z \mathbf{A}_t f - f \mathbf{A}_t \ln \lambda_z = \frac{[\lambda_z, f]^t}{\lambda_z}.$$

□

Proposition 2.5. *Assume* A 1 − 6 *hold true, then for* $z \in E$

$$\text{(2.73)} \quad \ln \lambda_z(X_t, \theta_t, t) Y_z(t) = \int_0^t \ln \lambda_z(X_s, \theta_s, s) dY_z(s) + \int_0^t Y_z(s-) \lambda_z^{-1}(X_s, \theta_s) dM_s^{\lambda_z} + \int_0^t Y_z(s) \mathbf{A}_s(\ln \lambda_z)(X_s, \theta_s, s) ds.$$

Here, $\mathbf{A}_t(\ln \lambda_z) = \frac{1}{\lambda_z} \cdot \left(\frac{\partial \lambda_z}{\partial t} + \mathbf{A}_t \lambda_z \right) - \frac{[\lambda_z, \lambda_z]^t}{2\lambda_z^2}$.

Proof. Note that the quadratic variation process does not change under equivalent measures. Meanwhile, from Proposition 2.2, $(Y_0(t), Y_1(t), \cdots)$ are independent Poisson processes with intensity $(\kappa_0, \kappa_1, \cdots)$ and they are independent of (X, θ) under $\mathbb{Q}$. Therefore, we have for $\forall z \in E$,

$$[\ln \lambda_z(X_t, \theta_t, t), \;\; Y_z(t)] = 0,$$

and by integrations by parts,

$$\begin{aligned} \text{(2.74)} \quad \ln \lambda_z(X_t, \theta_t, t) Y_z(t) &= \int_0^t \ln \lambda_z(X_s, \theta_s, s) dY_z(s) + \int_0^t Y_z(s-) d \ln \lambda_z(X_s, \theta_s, s). \\ &= \int_0^t \ln \lambda_z(X_s, \theta_s, s) dY_z(s) + \int_0^t Y_z(s-) \lambda_z^{-1}(X_s, \theta_s) dM_s^{\lambda_z} \\ &+ \int_0^t Y_z(s) \mathbf{A}_s(\ln \lambda_z)(X_s, \theta_s, s) ds. \end{aligned}$$

Hence the result. □

Proposition 2.6. *Under* A1 − 6, *for* $f \in \mathcal{D}(\mathbf{A})$,

$$\nu_t(f) = \mathbb{E}^{\mathbb{Q}}[f(X_t, \theta_t) \cdot \Xi_t^Y \cdot \mathbf{D}_t^Y] \tag{2.75}$$

where

(2.76)
$$\Xi_t^Y \triangleq \exp\Big(\int_0^t -\{\int_E \mathbf{A}_s(\ln \lambda_z)Y(dz) + (a(s) - \kappa) - \int_{E\times E} \frac{[\lambda_z, \lambda_\zeta]^s}{2\lambda_z\lambda_\zeta} Y(dz, d\zeta)\}ds,$$

(2.77)
$$\mathbf{D}_t^Y \triangleq \exp\left(-\int_0^t \int_E Y_z(s-)\frac{1}{\lambda_z} dM_s^{\lambda_z} m(dz) - \int_0^t \int_{E\times E} \frac{[\lambda_z, \lambda_\zeta]^s}{2\lambda_z\lambda_\zeta} Y(dz, d\zeta)ds\right).$$

Proof. From Proposition 2.5, we have

$$L_t = \exp\left(\left(\int_E (\ln \lambda_z - \ln \kappa_z)Y(dz) - \int_0^t \left(\int_E \mathbf{A}(\ln \lambda_z)Y(dz) + [a(s) - \kappa]\right) ds\right) \cdot \right.$$
$$\left. \exp\left(-\int_0^t \int_E Y_z(s-)\lambda_z^{-1} dM_s^{\lambda_z} m(dz)\right).\right.$$

Thus, the result follows from the independence of (X, θ) and $\overrightarrow{Y}$ under $\mathbb{Q}$. □

Lemma 2.2. *Under* A 1 − 6, $\mathbf{D}_t^Y$ *is a continuous martingale.*

Proof. From A 6, we know the stochastic integration

$$\int_0^t \int_E Y_z(s-)\frac{1}{\lambda_z} dM_s^{\lambda_z} m(dz)$$

is continuous (see Protter (1990), Page 53) and from the representation (2.77),

$$\mathbf{D}_t^Y = -\int_0^t \int_E \mathbf{D}_s^Y Y_z(s) \frac{1}{\lambda_z} dM_s^{\lambda_z} m(dz), \tag{2.78}$$

that $\mathbf{D}_t^Y$ is a continuous local martingale. Meanwhile, we have

$$\mathbb{E} \exp\left(\int_0^t \int_{E\times E} \frac{[\lambda_z, \lambda_\zeta]^s}{2\lambda_z\lambda_\zeta}(X_s, \theta_s)Y(dz, s)Y(d\zeta, s)ds\right) < \infty, \qquad \forall t \in [0, T].$$

Then, from the Novikov criteria, $\mathbf{D}_t^Y$ is a martingale. This completes the proof. □

After the first probability change $\mathbb{Q}$, the distribution of (X, θ) remains unchanged. In contrast, with the second path-dependent probability change, the distribution of X is altered and the new distribution can be characterized by some observation-path-dependent martingale problem. The gauge transform ν_t will then take the form of Feynman-Kac multiplicative functional. Now, we are ready to introduce the observation-path-dependent probability change. For given $y(\cdot)$, introduce the path-dependent probability $\widehat{\mathbb{Q}}^Y$ by

$$\left.\frac{d\widehat{\mathbb{Q}}^Y}{d\mathbb{Q}}\right|_{\mathcal{F}_t} = \mathbf{D}_t^Y, \tag{2.79}$$

under which the gauge transform can be characterized in form of Feynman-Kac multiplicative functional:

$$\nu_t(f) = \widehat{\mathbb{E}}^Y[f(X_t, \theta_t)\Xi_t^Y], \tag{2.80}$$

where $\widehat{\mathbb{E}}^Y(\cdot)$ denotes the expectation under $\widehat{\mathbb{Q}}^Y$.

Lemma 2.3. *Assume* A $1-6$, *then under* $\widehat{\mathbb{Q}}^Y$, (X, θ) *is a solution of the* $\widehat{\mathbf{A}}^Y$ *martingale problem*

$$df(X_t, \theta_t) = \widehat{\mathbf{A}}_t^Y f(X_t, \theta_t)dt + d\widehat{M}_t^f, \tag{2.81}$$

where $\mathcal{D}(\widehat{\mathbf{A}}^Y) = \mathcal{D}(\mathbf{A})$ *and for* $f \in \mathcal{D}(\widehat{\mathbf{A}}^Y)$,

$$\widehat{\mathbf{A}}_t^Y f \triangleq \mathbf{A}_t f - \int_E \frac{[\lambda_z, f]^t}{\lambda_z} Y(dz, t) \tag{2.82}$$

and

$$d\widehat{M}_t^f \triangleq dM_t^f + \int_E \frac{[\lambda_z, f]^t}{\lambda_z} Y(dz, t)dt. \tag{2.83}$$

Proof. For $f \in \mathcal{D}(\mathbf{A})$, we have

$$\begin{aligned}
df(X_t, \theta_t) &= \mathbf{A}_t f(X_t, \theta_t)dt + dM_t^f \\
&= \mathbf{A}_t f(X_t, \theta_t)dt - \int_E \frac{[\lambda_z, f]^t}{\lambda_z} Y(dz, t) + dM_t^f + \int_E \frac{[\lambda_z, f]^t}{\lambda_z} Y(dz, t) \\
&= \widehat{\mathbf{A}}_t^Y f(X_t, \theta_t)dt + d\widehat{M}_t^f.
\end{aligned}$$

Here, $\widehat{\mathbf{A}}_t^Y$ and $\widehat{M}_t^f$ are defined as in (2.82), (2.83). Now we only need to show $\widehat{M}_t^f$ is a martingale under $\widehat{\mathbb{Q}}^Y$. Due to the continuity of $\mathbf{D}_t^Y$, we have

$$\Delta[M^f, \mathbf{D}^Y]_t = 0,$$

so

$$[M^f, \mathbf{D}^Y]_t = \langle M^f, \mathbf{D}^Y \rangle_t = -\int_0^t \int_E \mathbf{D}_s^Y [f, \ln \lambda_z]^s Y(dz, s) ds.$$

From (2.63), we have

$$\begin{aligned} d\widehat{M}_t^f &= dM_t^f + \int_E \frac{[\lambda_z, f]^t}{\lambda_z} Y(dz, t) dt \\ &= dM_t^f + \int_E [\ln \lambda_z, f]^t (X_t, \theta_t) Y(dz, t) dt \\ &= dM_t^f - \frac{1}{\mathbf{D}_t^Y} d[M^f, \mathbf{D}^Y]_t. \end{aligned} \tag{2.84}$$

Thus from the Girsanov-Meyer theorem, $\widehat{M}_t^f$ is a local martingale under $\widehat{\mathbb{Q}}^Y$. Moreover, we know $\widehat{M}_t^f$ is locally bounded. Any local martingale that is locally bounded is a martingale. (see Protter (1990), Page 35). Hence the result follows directly. □

Proof of Theorem 2.2

Proof. From integration by parts,

$$\begin{aligned} d(f(X_t, \theta_t)\Xi_t^Y) = \Xi_t^Y d\widehat{M}_t^f + \widehat{\mathbf{A}}_t^Y f \Xi_t^Y dt + f \Xi_t^Y \{ -\int_E (\mathbf{A}_t \ln \lambda_z) Y(dz) \\ - (a(t) - \kappa) + \int_{E \times E} \frac{[\lambda_z, \lambda_\zeta]^t}{2\lambda_z \lambda_\zeta} (X_t, \theta_t, t) Y(dz, t) Y(d\zeta, t) \} dt. \end{aligned}$$

Note that $a(x, \theta, t)$ is bounded because of A 3 and it follows λ_z is also a bounded function. Meanwhile, for any $t \in [0, T]$, there are at most finite many jumps (non explosion) occurring during $[0, t]$ so $Y_z(t)$ is also bounded for any $z \in E$. Combining these points together, we know Ξ_t^Y is locally bounded. In Lemma 2.3, we already prove $\widehat{M}_t^f$ is a locally bounded martingale under $\widehat{\mathbb{Q}}^Y$, therefore (see Protter (1990), Page 66, Corollary 3), we have

$$\widehat{\mathbb{E}}^Y [\widehat{M}^f, \widehat{M}^f]_t < \infty$$

for any $t \in [0,T]$. Combining these points together, we have

$$\widehat{\mathbb{E}}^Y[\int_0^t \Xi_s^Y d\widehat{M}_s^f, \int_0^t \Xi_s^Y d\widehat{M}_s^f]_t = \widehat{\mathbb{E}}^Y \int_0^t (\Xi_s^Y)^2 d[\widehat{M}^f, \widehat{M}^f]_s < \infty. \tag{2.85}$$

Therefore, (see also Protter (1990), Page 66, Corollary 3), we have

$$\int_0^t \Xi_s^Y d\widehat{M}_s^f$$

is a $\widehat{\mathbb{Q}}^Y$-martingale. Then take expectation under $\widehat{\mathbb{Q}}^Y$,

$$\begin{aligned} \nu_t(f) = \nu_0(f) \quad &+ \quad \int_0^t \nu_s(\widehat{\mathbf{A}}_s^Y f)ds + \int_0^t \nu_s(f\{-\int_E (\mathbf{A}_s \ln \lambda_z)Y(dz,s) - (a(s) - \kappa) \\ &+ \quad \int_{E\times E} \frac{[\lambda_z, \lambda_\zeta]^s}{2\lambda_z\lambda_\zeta} Y(dz,s)Y(d\zeta,s)\})ds. \end{aligned}$$

Recalling that

$$\begin{aligned} \widehat{\mathbf{A}}_t^Y f &= \mathbf{A}_t f - \int_E \frac{[\lambda_z, f]^t}{\lambda_z} Y(dz,t), \\ \mathbf{A}_t(\ln \lambda_z) &= \frac{\mathbf{A}_t(\lambda_z)}{\lambda_z} + \frac{1}{\lambda_z} \cdot \frac{\partial \lambda_z}{\partial t} - \frac{[\lambda_z, \lambda_z]^t}{2\lambda_z^2}. \end{aligned}$$

Therefore we get

$$\nu_t(f) = \nu_0(f) + \int_0^t \nu_s(\widetilde{\mathbf{A}}_s^Y f)ds,$$

where

$$\begin{aligned} \widetilde{\mathbf{A}}_t^Y f = &\left(\mathbf{A}_t + \kappa - a(t) + \int_{E\times E} \frac{[\lambda_z, \lambda_\zeta]^s}{2\lambda_z\lambda_\zeta} Y(dz,s)Y(d\zeta,s)\right) f \\ &- \int_E \left(\frac{[\lambda_z, f]^t + 2f\mathbf{A}_t(\lambda_z)}{\lambda_z} + \frac{f}{\lambda_z} \cdot \frac{\partial \lambda_z}{\partial t} - \frac{f\mathbf{A}_t(\lambda_z^2)}{2\lambda_z^2}\right) Y(dz,t). \end{aligned}$$

This completes the proof. □

2.4 Bayes Model Selection

The main objective of this section is to use Bayes factor to investigate the model selection problem. To use the Bayes factor method, we need only be able to transform all observation models of interests into the same canonical process via Girsanov measure change. The signal models can be singular to one another. Kouritzin and Zeng (2005) discuss the Bayesian model selection problem to determine which of a class of financial models best represents given financial data such as stock price. However, their work does not use the robust equations and their non-robust equation do not strictly apply to our more general models.

The available information concerning the state is the observation process $\overrightarrow{Y}$. The Bayes factor determines which class of models best fits such observed datum by doing pairwise comparisons. Unlike Kouritzin and Zeng (2005), our method is based on the robust filter. The underlying robust filtering equation has only one time scale "dt" so only common calculus formulas are required to derive the dynamics of Bayes factor. Suppose there are two models:

$$M^{(k)} \triangleq (X^{(k)}, \theta^{(k)}, \overrightarrow{Y}^{(k)})$$

with total intensity function

$$a^{(k)} = a^{(k)}(X^{(k)}, \theta^{(k)}, t) = \int_E \lambda^{(k)}(dz, t)$$

where the random measure $\lambda^{(k)}(dz, t)$ satisfying

$$\lambda^{(k)}(\{z\}, t) = \lambda_z^{(k)}(X^{(k)}, \theta^{(k)}, t).$$

The generators of the martingale problem are respectively $\mathbf{A}^{(k)}$ for $k = 1, 2$. The joint likelihood of $(X^{(k)}, \theta^{(k)}, \overrightarrow{Y}^{(k)})$ at time t is denoted by $L_t^{(k)}$ which satisfies

$$L_t^{(k)} = 1 + \int_0^t \int_E \left(\frac{\lambda_z^{(k)}}{\kappa_z^{(k)}} - 1 \right) L_{s-}^{(k)} \left(Y^{(k)}(dz, ds) - \kappa_z^{(k)} m(dz) ds \right).$$

Here, we assume

$$\kappa_z^{(1)} = \kappa_z^{(2)}$$

for each $z \in E$. The normalized filter $\pi_t^{(k)}, \ k = 1, 2$ satisfies

$$\pi^{(k)}(f_k, t) = \frac{\sigma^{(k)}(f_k, t)}{\sigma^{(k)}(1, t)}$$

where for $k = 1, 2$, the unnormalized filter $\sigma_t^{(k)}$ is defined as

$$\sigma^{(k)}(f_k, t) \triangleq \mathbb{E}^{\mathbb{Q}^{(k)}}[f_k(X_t^{(k)}, \theta_t^{(k)})L_t^{(k)} | \mathcal{F}_t^{\vec{Y}^{(k)}}]$$

and $\sigma^{(k)}(1, t)$ is the integrated (or marginal) likelihood of $\vec{Y}^{(k)}$ for $k = 1, 2$.

Definition 2.7. *The filter ratio processes are defined as*

$$q_1(f_1, t) = \frac{\sigma^{(1)}(f_1, t)}{\sigma^{(2)}(1, t)} \quad \text{and} \quad q_2(f_2, t) = \frac{\sigma^{(2)}(f_2, t)}{\sigma^{(1)}(1, t)}. \tag{2.86}$$

As a sequel, the Bayes factor is defined as

Definition 2.8.

$$B_{12}(t) = q_1(1, t); \quad B_{21}(t) = q_2(1, t). \tag{2.87}$$

Once the definition of Bayes factor is given, the next step is how to interpret it. Here, we refer the work of Kass and Raftery (1995), Kouritzin and Zeng (2005) and demonstrate it through the following table.

B_{12}	Evidence against Model 2
$1 - 3$	Not worth more than a bare mention
$3 - 12$	Positive
$12 - 150$	Strong
> 150	Decisive

Now, we turn to the issue of how to calculate the Bayes factor. As discussed in Kouritzin and Zeng (2005), there exist two alternatives to compute the Bayes factor. The first one is calculate the integrated likelihood $\sigma^{(k)}(1, t), \ k = 1, 2$ respectively and then take the ratio to get the Bayes factor

$$B_{12}(t) = \frac{\sigma^{(1)}(1, t)}{\sigma^{(2)}(1, t)}, \tag{2.88}$$

$$B_{21}(t) = \frac{\sigma^{(2)}(1,t)}{\sigma^{(1)}(1,t)}. \tag{2.89}$$

where $\sigma^{(1)}$ and $\sigma^{(2)}$ are computed by the unnormalized filtering equation. However, this approach is not always computational efficient or numerically stable. It is quite possible that both $\sigma^{(1)}(1,t)$ and $\sigma^{(2)}(1,t)$ get very large or very small as t increases. For these reasons, we focus on the second approach which consists of two steps: firstly, we derive the evolution equation which describes the dynamics of Bayes factor; secondly, we develop some particle filter algorithm to implement this equation directly. The following equation for filter ratio process generalizes that from Kouritzin and Zeng (2005):

Theorem 2.3. *Suppose there are two models $M^{(i)}$, $i = 1, 2$ satisfies assumptions* A2 − 3, *then*

$$\begin{aligned} q_t^{(i)}(f) = q_0^{(i)}(f) &+ \int_0^t q_s^{(i)}(\mathbf{A}_s^{(i)} f - a_s^{(i)} f) + \frac{q_s^{(i)}(f) q_s^{(3-i)}(a^{(3-i)})}{q_s^{(3-i)}(1)} ds \\ &+ \int_0^t \int_E \left(\frac{q_{s-}^{(i)}(f\lambda_z^{(i)}(s)) q_{s-}^{(3-i)}(1)}{q_{s-}^{(3-i)}(\lambda_z^{(3-i)}(s))} - q_{s-}^{(i)}(f) \right) Y(dz, ds). \end{aligned} \tag{2.90}$$

In particular, suppose the models $M^{(1)} = M^{(2)}$, then we get the evolution equation of normalized filter $\pi_t(f)$:

$$\begin{aligned} \pi_t(f) = \pi_0(f) &+ \int_0^t \pi_s(\mathbf{A} f - a(s) f) + \pi_s(f)\pi_s(a(s)) ds \\ &+ \int_0^t \int_E \left(\frac{\pi_{s-}(f\lambda_z)}{\pi_{s-}(\lambda_z)} - \pi_{s-}(f) \right) Y(dz, ds). \end{aligned} \tag{2.91}$$

Proof. Use Theorem 2.1, we can get the evolution equation of $\sigma_t^{(i)}(f)$ and $\sigma_t^{(3-i)}(1)$ respectively and then apply the Itô formula to $\frac{\sigma_t^{(i)}(f)}{\sigma_t^{(3-i)}(1)}$. Applying (2.90) to the equality $\pi_t(f) = \frac{\sigma_t^{(1)}(f)}{\sigma_t^{(1)}(1)}$, we get (2.91). □

2.5 Robust Bayes Model Selection

Recall the gauge transform

$$\nu^{(k)}(f_k,t) \triangleq \mathbb{E}^{\mathbb{Q}^{(k)}}\left[f_k(X_t^{(k)},\theta_t^{(k)})L_t^{(k)} \exp\left(-\int_E \ln \frac{\lambda_z^{(k)}}{\kappa_z^{(k)}}(X_t^{(k)},\theta_t^{(k)})Y(dz)\right) |\mathcal{F}_t^{\overrightarrow{Y}^{(k)}}\right].$$

Now for $k = 1, 2$, we can introduce the notations

$$g^{(k)} \triangleq \exp\left(-\int_E \ln \frac{\lambda_z^{(k)}}{\kappa_z^{(k)}} Y(dz,t)\right). \tag{2.92}$$

From the definition of the filter ration processes, one has

$$\begin{aligned} q_1(f_1,t) &= \frac{\sigma^{(1)}(f_1,t)}{\sigma^{(2)}(1,t)} = \frac{\nu^{(1)}(f_1 g^{(1)},t)}{\nu^{(2)}(g^{(2)},t)} = \widetilde{q}_1(f_1 g^{(1)},t), \\ q_2(f_2,t) &= \frac{\sigma^{(2)}(f_2,t)}{\sigma^{(1)}(1,t)} = \frac{\nu^{(2)}(f_2 g^{(2)},t)}{\nu^{(1)}(g^{(1)},t)} = \widetilde{q}_2(f_2 g^{(2)},t), \end{aligned} \tag{2.93}$$

where

$$\widetilde{q}_1(f_1,t) \triangleq \frac{\nu^{(1)}(f_1,t)}{\nu^{(2)}(g^{(2)},t)} \text{ and } \widetilde{q}_2(f_2,t) \triangleq \frac{\nu^{(2)}(f_2,t)}{\nu^{(1)}(g^{(1)},t)}. \tag{2.94}$$

The main result of this section is the following theorem:

Theorem 2.4. *Suppose there are two models $M^{(i)}$, $i = 1, 2$ satisfies assumptions* A1 − 6, *then*

(2.95)

$$\begin{aligned} d\widetilde{q}_i(f_i,t) &= \widetilde{q}_i\left(\widehat{\mathbf{A}}_t^{Y,\,(i)} f_i - a^{(i)} f_i,\ t\right) dt \\ &+ \widetilde{q}_i\left(f_i\left(-\int_E \mathbf{A}_t^{(i)} \ln \lambda_z^{(i)} Y(dz) + \int_{E\times E} \frac{[\lambda_z,\lambda_\zeta]^{(i),\,t}}{2\lambda_z^{(i)}\lambda_\zeta^{(i)}} Y(dz)Y(d\zeta)\right)\right) dt \\ &- \frac{\widetilde{q}_i(f_i,t)}{\widetilde{q}_{3-i}(g^{(3-i)},t)} \cdot \widetilde{q}_{3-i}\left(\widehat{\mathbf{A}}_t^{Y,\,(3-i)} g^{(3-i)} - a^{(3-i)} g^{(3-i)},\ t\right) dt - \frac{\widetilde{q}_{3-i}(f_i,t)}{\widetilde{q}_{3-i}(g^{(3-i)},t)} \\ &\cdot \widetilde{q}_{3-i}\left(g^{(3-i)}\left(-\int_E \mathbf{A}_t^{(3-i)} \ln \lambda_z^{(3-i)} Y(dz) + \int_{E\times E} \frac{[\lambda_z,\lambda_\zeta]^{(3-i),\,t}}{2\lambda_z^{(3-i)}\lambda_\zeta^{(3-i)}} Y(dz,d\zeta)\right)\right) dt. \end{aligned}$$

Proof. Note that

$$\widetilde{q}_1(f_1,t) = \frac{\nu^{(1)}(f_1,t)}{\nu^{(2)}(g^{(2)},t)} \quad \text{and} \quad \widetilde{q}_2(f_2,t) = \frac{\nu^{(2)}(f_2,t)}{\nu^{(1)}(g^{(1)},t)}.$$

Moreover, from Theorem 2.2, we have

$$d\nu^{(1)}(f_1,t) = \nu^{(1)}\left(\widehat{\mathbf{A}}_t^{y,\,(1)} f_1 - (a^{(1)}(s) - \kappa^{(1)})f_1, t\right) dt$$
$$+ \nu^{(1)}\left(f_1\left(-\int_E \mathbf{A}_t^{(1)} \ln \lambda_z^{(1)} Y(dz) + \int_{E\times E} \frac{[\lambda_z, \lambda_\zeta]^{(1),\,t}}{2\lambda_z\lambda_\zeta} Y(dz)Y(d\zeta)\right)\right)$$

and

$$d\nu^{(2)}(f_2,t) = \nu^{(2)}\left(\widehat{\mathbf{A}}_t^{y,\,(2)} f_2 - (a^{(2)}(s) - \kappa^{(2)})f_2, t\right) dt$$
$$+ \nu^{(2)}\left(f_2\left(-\int_E \mathbf{A}_t^{(2)} \ln \lambda_z^{(2)} Y(dz) + \int_{E\times E} \frac{[\lambda_z, \lambda_\zeta]^{(2),\,t}}{2\lambda_z\lambda_\zeta} Y(dz)Y(d\zeta)\right)\right)$$

By the product rule, it follows

$$d\widetilde{q}_1(f_1,t) = \frac{\nu^{(2)}(g^{(2)},t)d\nu^{(1)}(f_1,t) - \nu^{(1)}(f_1,t)d\nu^{(2)}(g^{(2)},t)}{\left(\nu^{(2)}(g^{(2)},t)\right)^2}$$
$$= \frac{d\nu^{(1)}(f_1,t)}{\nu^{(2)}(g^{(2)},t)} - \frac{\frac{\nu^{(1)}(f_1,\,t)d\nu^{(2)}(g^{(2)},\,t)}{\nu^{(2)}(g^{(2)},\,t)\nu^{(1)}(g^{(1)},\,t)}}{\frac{\nu^{(2)}(g^{(2)},\,t)}{\nu^{(1)}(g^{(1)},\,t)}}.$$

Therefore

$$d\widetilde{q}_1(f_1,t) = \frac{d\nu^{(1)}(f_1,t)}{\nu^{(2)}(g^{(2)},t)} - \frac{\frac{\nu^{(1)}(f_1,\,t)}{\nu^{(2)}(g^{(2)},\,t)} \cdot \frac{d\nu^{(2)}(g^{(2)},\,t)}{\nu^{(1)}(g^{(1)},\,t)}}{\frac{\nu^{(2)}(g^{(2)},\,t)}{\nu^{(1)}(g^{(1)},\,t)}}$$
$$= \frac{\nu^{(1)}\left(\widehat{\mathbf{A}}_t^{y,(1)} f_1 - (a^{(1)}(s) - \kappa^{(1)})f_1,\, t\right)}{\nu^{(2)}(g^{(2)},\, t)} dt$$
$$+ \frac{\nu^{(1)}\left(f_1(-\int_E \mathbf{A}_t^{(1)} \ln \lambda_z^{(1)} Y(dz) + \int_{E\times E} \frac{[\lambda_z,\lambda_\zeta]^{(1),\,t}}{2\lambda_z\lambda_\zeta} Y(dz)Y(d\zeta))\right)}{\nu^{(2)}(g^{(2)},t)} dt$$
$$- \frac{\frac{\nu^{(1)}(f_1,\,t)}{\nu^{(2)}(g^{(2)},\,t)}}{\frac{\nu^{(2)}(g^{(2)},\,t)}{\nu^{(1)}(g^{(1)},\,t)}} \cdot \frac{d\nu^{(2)}(g^{(2)},t)}{\nu^{(1)}(g^{(1)},t)}.$$

Thus

$$d\widetilde{q}_1(f_1,t) = \widetilde{q}_1\left(\widehat{\mathbf{A}}_t^{y,\,(1)} f_1 - (a^{(1)}(s) - \kappa^{(1)}) f_1,\ t\right) dt$$
$$+ \widetilde{q}_1\left(f_1\left(-\int_E \mathbf{A}_t^{(1)} \ln \lambda_z Y(dz,t) + \int_{E\times E} \frac{[\lambda_z, \lambda_\zeta]^{(1),\,t}}{2\lambda_z\lambda_\zeta}\right),\ \ t\right) dt$$
$$- \frac{\widetilde{q}_1(f_1,t)}{\widetilde{q}_2(g^{(2)},t)} \cdot \frac{\nu^{(2)}\left(\widehat{\mathbf{A}}_t^{y,\,(2)} g^{(2)} - (a^{(2)}(s) - \kappa^{(2)}) g^{(2)},\ t\right)}{\nu^{(1)}(g^{(1)},\ \ t)} dt$$
$$- \frac{\widetilde{q}_1(f_1,t)}{\widetilde{q}_2(g^{(2)},t)} \cdot \frac{\nu^{(2)}\left(g^{(2)}(-\int_E \mathbf{A}_t^{(2)} \ln \lambda_z^{(2)} Y(dz) + \int_{E\times E} \frac{[\lambda_z,\lambda_\zeta]^{(2),\,t}}{2\lambda_z\lambda_\zeta})\right)}{\nu^{(1)}(g^{(1)})} dt.$$

It follows that

$$d\widetilde{q}_1(f_1,t) = \widetilde{q}_1\left(\widehat{\mathbf{A}}_t^{y,\,(1)} f_1 - (a^{(1)}(s) - \kappa^{(1)}) f_1,\ \ t\right) dt$$
$$+ \widetilde{q}_1\left(f_1\left(-\int_E \mathbf{A}_t^{(1)} \ln \lambda_z^{(1)} Y(dz) + \int_{E\times E} \frac{[\lambda_z, \lambda_\zeta]^{(1),\,t}}{2\lambda_z\lambda_\zeta} Y(dz)Y(d\zeta)\right)\right) dt$$
$$- \frac{\widetilde{q}_1(f_1,t)}{\widetilde{q}_2(g^{(2)},t)} \cdot \widetilde{q}_2\left(\widehat{\mathbf{A}}_t^{y,\,(2)} g^{(2)} - (a^{(2)}(s) - \kappa^{(2)}) g^{(2)},\ \ t\right) dt$$
$$- \frac{\widetilde{q}_2(f_1,t)}{\widetilde{q}_2(g^{(2)},\ \ t)} \cdot \widetilde{q}_2\left(g^{(2)}\left(-\int_E \mathbf{A}_t^{(2)} \ln \lambda_z^{(2)} Y(dz) + \int_{E\times E} \frac{[\lambda_z, \lambda_\zeta]^{(2)\ ,t}}{2\lambda_z\lambda_\zeta}\right)\right) dt.$$

Due to $\kappa^{(1)} = \kappa^{(2)}$, then

$$\widetilde{q}_1(f_1,\ \ t) \cdot \widetilde{q}_2(\kappa^{(2)} g^{(2)},\ \ t) = \widetilde{q}_1(\kappa^{(1)} f_1,\ \ t) \cdot \widetilde{q}_2(g^{(2)},\ \ t), \tag{2.96}$$

we have that

$$d\widetilde{q}_1(f_1,t) = \widetilde{q}_1\left(\widehat{\mathbf{A}}_t^{y,\,(1)} f_1 - a^{(1)} f_1,\ t\right) dt$$
$$+ \widetilde{q}_1\left(f_1\left(-\int_E \mathbf{A}_t^{(1)} \ln \lambda_z^{(1)} Y(dz,t) + \int_{E\times E} \frac{[\lambda_z, \lambda_\zeta]^{(1),\ t}}{2\lambda_z\lambda_\zeta} Y(dz)Y(d\zeta)\right)\right) dt$$
$$- \frac{\widetilde{q}_1(f_1,t)}{\widetilde{q}_2(g^{(2)},t)} \cdot \widetilde{q}_2\left(\widehat{\mathbf{A}}_t^{y,\,(2)} g^{(2)} - a^{(2)} g^{(2)},\ t\right) dt$$
$$- \frac{\widetilde{q}_2(f_1)}{\widetilde{q}_2(g^{(2)})} \cdot \widetilde{q}_2\left(g^{(2)}\left(-\int_E \mathbf{A}_t^{(2)} \ln \lambda_z^{(2)} Y(dz) + \int_{E\times E} \frac{[\lambda_z, \lambda_\zeta]^{(2),\,t}}{2\lambda_z\lambda_\zeta} Y(dz, d\zeta)\right)\right) dt.$$

Similarly, the following result holds true

$$
\begin{aligned}
&d\widetilde{q}_2(f_2,t) = \widetilde{q}_2\left(\widehat{\mathbf{A}}_t^{y,\,(2)} f_2 - a^{(2)} f_2,\ t\right) dt \\
&+ \widetilde{q}_2\left(f_2\left(-\int_E \mathbf{A}_t^{(2)} \ln \lambda_z^{(2)} Y(dz) + \int_{E\times E} \frac{[\lambda_z,\lambda_\zeta]^{(2),\,t}}{2\lambda_z\lambda_\zeta} Y(dz)Y(d\zeta)\right)\right) dt \\
&- \frac{\widetilde{q}_2(f_2,t)}{\widetilde{q}_1(g^{(1)},t)} \cdot \widetilde{q}_1\left(\widehat{\mathbf{A}}_t^{y,\,(1)} g^{(1)} - a^{(1)} g^{(1)},\ t\right) dt \\
&- \frac{\widetilde{q}_1(f_2)}{\widetilde{q}_1(g^{(1)})} \cdot \widetilde{q}_1\left(g^{(1)}\left(-\int_E \mathbf{A}_t^{(1)} \ln \lambda_z^{(1)} Y(dz) + \int_{E\times E} \frac{[\lambda_z,\lambda_\zeta]^{(1),\,t}}{2\lambda_z\lambda_\zeta} Y(dz,d\zeta)\right)\right) dt.
\end{aligned}
$$

This completes the proof. □

2.6 Particle Filtering to Bayesian Estimation and Model Selection

Theoretically, Theorem 2.2 and 2.4 solve the Bayes estimation and model selection problem in that they already give the corresponding evolution equations. However, in numerical applications, we need some efficient and recursive approximation algorithm to implement this evolution equation.

To avoid the "curse of dimensionality," we propose some numerical algorithm based on the particle filtering instead of the Markov Chain approximation utilized in Zeng (2003) to implement the robust evolution equation (2.57) and (2.95). The following algorithm to be proposed is a new particle filter that can be thought of as a generalization of Del Moral, Noyer and Salut (1995), Del Moral, Jacod and Protter (2001). Similar works discussing the particle filtering methods include Desai, Viens and Lele (2003), Viens, Tindel (2004).

2.6.1 Particle Filtering to Bayesian Estimation

The key of particle filtering is to construct some particle system whose empirical distribution will converge weakly to the robust filter (or gauge transform) ν_t (which is specified in Theorem 2.2) as the number of particles tends to infinity. As discussed before, the robust filter has the advantage that it is continuously dependent on the underlying observation Y. Thus, we need only to consider the simple N-equalized time partitions

$$\{\tau_0 = 0, \tau_1 = \frac{T}{N}, \cdots \tau_i = \frac{iT}{N}, \cdots \tau_N = T\}$$

to make particle approximation converge regardless the trading times $\{t_1, t_2, \cdots\}$.

Now, let $\{\varphi_N(t);\ t \geq 0\}$ be a sequence of measure-valued processes which represent the empirical distributions of the particle system to be constructed. In general, the particle filtering algorithm can be divided into three consecutive phases:

- *Initialization,*
- *Evolution (prediction),*
- *Re-sampling (updating).*

Initialization

- At $\tau_0 = 0$, we draw m_N independent particles $\{P^k\}_{k=1}^{m_N}$ with the equal weights $\frac{1}{m_N}$ from the joint prior distribution $\pi_0(\cdot)$ of $(X_0; \theta) \in \mathbb{R}^{n_x+n_\theta}$. The states of these particles at time t are denoted as $\{P_t^k\}_{k=1}^{m_N}$. Here, m_N is some positive integer that satisfies

  $$\lim_{N \longrightarrow \infty} m_N = \infty.$$

  The empirical (occupation) measure of this particle system at $\tau_0 = 0$ is

  $$\varphi_N(0) \triangleq \frac{1}{m_N} \sum_{k=1}^{m_N} \delta_{P_0^k}(\cdot),$$

  which satisfies

  $$\lim_{N \longrightarrow \infty} (\varphi_N(0), f) = \pi_0(f) \quad \forall f \in B(\mathbb{R}^{n_x+n_\theta}).$$

  Here, $\delta_x(\cdot)$ is the Dirac measure at x.

  Remark 2.7. *Note that $L_0 = 1$ and $Y_j(0) = 0$ for all $j = 0, 1, \cdots$ thus*

  $$\pi_0(f) = \sigma_0(f) = \nu_0(f) \quad \forall f \in B(\mathbb{R}^{n_x+n_\theta}).$$

Remark 2.8. *As there has no special information, it is convenient to assign the uninformative prior, that is, the uniform distributions to* $(X_0;\theta)$ *where the range* $[\alpha_{X_0},\beta_{X_0}]\subset\mathbb{R}^{n_x}$ *and* $[\alpha_\theta,\beta_\theta]\subset\mathbb{R}^{n_\theta}$ *are determined from empirical study.*

Evolution

- During the interval $[\tau_{i-1},\tau_i),\ \ i=1,2,\cdots,N$, all particles move independently according to the same law of $(X;\theta)$. In the case θ is time-invariant, it suffices to consider only the dynamics of X during these intervals. From Lemma 2.3, the particles evolve according to the following path-dependent martingale problem

$$df(X_t^k,\theta_t^k)=\widehat{\mathbf{A}}_t^Y f(X_t^k,\theta_t^k)dt+d\widehat{M}_t^{f,k},$$

where $\mathcal{D}(\widehat{\mathbf{A}}^Y)=\mathcal{D}(\mathbf{A})$ and for $f\in\mathcal{D}(\widehat{\mathbf{A}}^Y)$,

$$\widehat{\mathbf{A}}_t^Y f\triangleq\mathbf{A}_t f-\int_E\frac{[\lambda_z,f]^t}{\lambda_z}Y(dz,t)$$

and

$$d\widehat{M}_t^{f,k}\triangleq dM_t^{f,k}+\int_E\frac{[\lambda_z,f]^t}{\lambda_z}(X_t^k,\theta_t^k)Y(dz,t)dt.$$

is a martingale.

Testing Particle Weight

- For $k=1,2,\cdots,m_N$, each particle is given a weight $\omega_i^k(\tau_i)$ at the ending point τ_i based on the likelihood of the observation depending on its trajectory realized on $[\tau_{i-1},\tau_i)$:

$$\omega_i^k(\tau_i)\triangleq\exp\Big(\int_{\tau_{i-1}}^{\tau_i}\int_E\big[-\mathbf{A}_s(\ln\lambda_z)(X_s^k,\theta_s^k,s)Y(dz,s)-(a(s)-\kappa)\big]+\int_{\tau_{i-1}}^{\tau_i}\int_{E\times E}\frac{[\lambda_z,\lambda_\zeta]^s}{2\lambda_z\lambda_\zeta}(X_s^k,\theta_s^k,s)Y(dz)Y(d\zeta)\Big).\tag{2.97}$$

Note that the parameter vector θ is time variant.

Remark 2.9. *In practice we can replace* E *by* $E^n\triangleq\{1,2,\cdots n\}$ *for some large* n.

Then, we have

$$\omega_i^k(\tau_i) = \exp(\int_{\tau_{i-1}}^{\tau_i} \{-\sum_{j=1}^{n} [Y_j(s)\mathbf{A}(\ln\lambda_j)(X_s^k, \theta_s^k, s)] - \big(a(X_s^k, \theta_s^k, s) - n\big) + \sum_{i,j=1}^{n} \frac{Y_iY_j}{2}(s)[\ln\lambda_i, \ln\lambda_j](X_s^k, \theta_s^k, s)\}ds). \tag{2.98}$$

Moreover, for numerical computation, a discrete version of the weight function is

$$\begin{aligned}\omega_i^k(\tau_i) = \exp(&-\sum_{j=1}^{n}\Big[Y_j(\tau_{i-1})\mathbf{A}(\ln\lambda_j)(X_{\tau_{i-1}}^k, \theta_{\tau_{i-1}}^k, \tau_{i-1})\Big]\cdot(\tau_i - \tau_{i-1}) \\ &- \Big(a(X_{\tau_{i-1}}^k, \theta_{\tau_{i-1}}^k, \tau_{i-1}) - n\Big)\cdot(\tau_i - \tau_{i-1}) \\ &+ \sum_{i,j=1}^{n} \frac{Y_iY_j}{2}(\tau_{i-1})[\ln\lambda_i, \ln\lambda_j](X_{\tau_{i-1}}^k, \theta_{\tau_{i-1}}^k, \tau_{i-1})\cdot(\tau_i - \tau_{i-1})).\end{aligned} \tag{2.99}$$

Re-sampling

- For $k = 1, 2, \cdots, m_N$, $i = 1, 2, \cdots, N$, we already discussed how to give each particle a weight $\omega_i^k(\tau_i)$ at τ_i. These weights are stored along with the states of particles before re-sampling. The average weight at τ_i is

$$\omega_i(\tau_i) \triangleq \frac{1}{m_N}\sum_{k=1}^{m_N}\omega_i^k(\tau_i) \tag{2.100}$$

plays an important role in the re-sampling procedure.

Remark 2.10. *To simplify the notation, we rewrite* $\omega_i(\tau_i) = \omega_i$; $\omega_i^k(\tau_i) = \omega_i^k$ *and ignore the partition symbol* τ_i *whenever there has no confusion.*

We are ready to demonstrate the re-sampling procedure. Roughly speaking, if a particle has a weight $\omega_i^k = r\omega_i + z$, where $r \in \{0, 1, 2, \cdots\}$ and $z \in (0, \omega_i)$ before the re-sampling, then there will be r or $r + 1$ particles at this state after the re-sampling with a probability selected in order to leave the system unbiased. To illustrate the re-sampling explicitly, it is helpful to describe its pseudo code in detail.

Pseudo Code of Re-sampling

- In what follows, for $k = 1, \cdots, m_N$, denote P_t^k the state of the k^{th} particle at time t. In particular, $P_{\tau_i -}^k$ the state of the k^{th} particle prior to the re-sampling (but after the evolution) and $P_{\tau_i}^k$ the position when the re-sampling is done. Here, the re-sampling time is set to be τ_i. We reorder the particles $\{P_{\tau_i -}^k\}_{k=1}^{m_N}$ according to their weights ω_i^k (beginning from the highest weight) and calculate ω_i as in (2.100).
- $m = 0$; for $l = 1$ to m_N :
- While $\omega_i^k \geq \omega_i$: then $\omega_i^k = \omega_i^k - \omega_i$; $m = m + 1$; $P_{\tau_i}^m = P_{\tau_i -}^l$.
- Evaluate the remaining weight: $W = (m_N - m)\omega_i$.
- Re-sample the rest particles for evolution: for $l = m + 1$ to m_N, $\omega_i = 0$; r is sampled uniformly from $[0, 1]$.
- For $k = 1$ to m_N, if $r \in [\frac{w_i}{W}, \frac{w_i + w_i^k}{W})$, then $P_{\tau_i}^l = P_{\tau_i -}^k, \omega_i = \omega_i + \omega_i^k$.

An useful example is as follows. Suppose prior to resampling, there are 5 particles:

Particle	Weight
$P_{\tau_k -}^1$	1
$P_{\tau_k -}^2$	0.5
$P_{\tau_k -}^3$	0.25
$P_{\tau_k -}^4$	0.125
$P_{\tau_k -}^5$	0.125

Then $\omega_k = 0.4$ and after the first part of the resampling we would have

Particle	Weight	Site
$P_{\tau_k}^1$	0.4	$P_{t_k -}^1$
$P_{\tau_k}^2$	0.4	$P_{t_k -}^1$
$P_{\tau_k}^3$	0.4	$P_{t_k -}^2$

and there would be two more independent particles to be randomly positioned at the sites according to:

Site	Probability
$P^1_{\tau_k-}$	$\frac{1-0.8}{2-1.2} = \frac{1}{4}$
$P^2_{\tau_k-}$	$\frac{0.5-0.4}{0.8} = \frac{1}{8}$
$P^3_{\tau_k-}$	$\frac{0.25}{0.8} = 0.3125$
$P^4_{\tau_k-}$	$\frac{0.125}{0.8} = 0.15625$
$P^5_{\tau_k-}$	0.15625

with each also given the weight 0.4. Note that the expected weight of each particle site before and after resampling is the same. That is, $1 = 2 \times 0.4 + 2 \times \frac{1}{4} \times 0.4$. Also, the weight of 0.4 is the same for all particles after resampling and as such is irrelevant. Therefore, the weight can be set to one and thrown out.

Bayesian Estimation

Note that

$$\pi_t(f) = \frac{\nu_t\Big(f \exp(\int_E \ln(\frac{\lambda_z}{\kappa_z})Y(t, dz))\Big)}{\nu_t\Big(\exp(\int_E \ln(\frac{\lambda_z}{\kappa_z})Y(t, dz))\Big)}.$$

Therefore, the particle system approximating the normalized filter $\pi(\cdot)$ is

$$\pi_{N,\,t}(f) = \frac{\sum_{k=1}^{m_N} f(P_t^k) \exp(\int_E \ln(\frac{\lambda_z}{\kappa_z})(P_t^k, t))Y(t, dz))}{\sum_{k=1}^{m_N} \exp(\int_E \ln(\frac{\lambda_z}{\kappa_z})(P_t^k, t))Y(t, dz))}$$

for all $f \in B(\mathbb{R}^{n_\theta + n_x})$.

2.6.2 Particle Filtering to Model Selection

The evolution equation of the Bayes factor also admits no explicit form solution, thus we also apply the particle filtering method to solve it numerically. The procedures are similar to that of the Bayes estimation problem and we introduce the particle pair as $\{P_t^{(k_1)}, P_t^{(k_2)}\}_{k=1}^{m_N}$. It is notable that now the weight takes the form as follows.

Evaluate the Weights

- During the interval $[\tau_{i-1}, \tau_i)$, $i = 1, 2, \cdots, N$, the particles $\{P_t^{(k_1)}, P_t^{(k_2)}\}_{k=1}^{m_N}$ move independently according to the law of model $M^{(k_1)}, M^{(k_2)}$ given by Lemma 2.3. Intuitively, the particle system explores the state space following the law of the state process.

- For $i = 1, 2, \cdots, N$, the particles $\{P_t^{(k_1)}\}_{k=1}^{m_N}$ are respectively given a weight $\{\omega_i^{(k_1), k}\}_{k=1}^{m_N}$ at time τ_i based on its trajectory realized on $[\tau_{i-1}, \tau_i)$:

$$\omega_i^{(k_1), k} = \Upsilon_i^{(k_1)}(\tau_i) \cdot \Lambda_i^{(k_1), k}(\tau_i) \cdot \exp\Big(\int_{\tau_{i-1}}^{\tau_i} \Big\{\int_E \big[-\mathbf{A}(\ln \lambda_z^{(k_1)})Y(dz)\big]$$

$$- a(X_s^{(k_1), k}, \theta_s^{(k_1), k}, s) + \int_{E \times E} \frac{[\lambda_z^{(k_1)}, \lambda_\zeta^{(k_1)}]}{2\lambda_z \lambda_\zeta} Y(dz) Y(d\zeta)\Big\} ds\Big),$$

where

$$\Upsilon_i^{(k_1)}(t) = \exp\left(-\int_{\tau_{i-1}}^{t} \frac{\sum_{k=1}^{m_N}\left(\widehat{\mathbf{A}}^{y, (k_2)} g^{(k_2)} - a^{(k_2)} g^{(k_2)}\right)}{\sum_{k=1}^{m_N} g^{(k_2)}} ds\right),$$

$$\Lambda_i^{(k_1)}(t) = \left(-\int_{\tau_{i-1}}^{t} \frac{\sum_{k=1}^{m_N}\left(g^{(k_2)} \Gamma^{(k_2)}\right)}{\sum_{k=1}^{m_N} g^{(k_2)}} ds\right).$$

Here,

$$\Gamma^{(k_1)} \triangleq -\int_E \mathbf{A}^{(k_2)} \ln \lambda_z^{(k_2)} Y(dz) + \int_{E \times E} \frac{[\lambda_z{}^{(k_2)}, \lambda_\zeta^{(k_2)}]}{2\lambda_z \lambda_\zeta} Y(dz) Y(d\zeta).$$

Note that both $\Upsilon_i^{(k_1)}(t)$ and $\Lambda_i^{(k_1)}(t)$ does not depend on k thus can be omitted in re-sampling step and thus the weight coincides with that given in Section 2.6.1.

- For $i = 1, 2, \cdots, N$, the particles $\{P_t^{(k_2)}\}_{k=1}^{m_N}$ are given a weight $\{\omega_i^{(k_2), k}\}_{k=1}^{m_N}$ similarly. The remaining procedures of particle filtering are just an analog of that given in Section 2.6.1 and we make the re-sampling step to successively incorporate the new information into the Bayes factor. Here we ignore the deduction details to save space.

2.7 Historical Process and Long Memory Process

There is an increasing amount of literature which provide strong evidence of the long-memory dependence and self-similarity exhibited in numerous fields (see Mandelbrot (1971)). The continuous-time fractional Brownian motion (FBM) proposed by Mandelbrot and Van Ness (1968) can be employed as the building block of stochastic models to capture these statistical properties. FBM is one of the simplest stochastic process exhibiting the long-memory dependence and self-similarity. It seems very desirable to extend the classical stochastic systems driven by the Brownian motion to analogues in which the driving process is FBM. For example, the stochastic processes satisfying the linear stochastic differential equation driven by FBM. Among them, one of the most important is the fractional Ornstein-Uhlenbeck (FOU) process.

However, it is well known that the FBM and FOU are both non-Markovian except its special Brownian motion and Ornstein-Uhlenbeck (O-U) process cases. Due to this reason, the traditional stochastic filtering theory fails to work directly because it is no longer possible to characterize FBM or FOU through some type of martingale problem. Instead, inspired by Dawson and Perkins (1991), Dynkin (1991), as well as Kouritzin, Long and Sun (2003), we switch to the historical process of FBM or FOU process. This is because for any stochastic process S, we can introduce its historical process S^H which is definitely Markovian and it is possible to investigate the related martingale problem.

In the sequel, we construct a sequence of discrete-parameter Markov chains which converge point-wise to the historical process of FOU. Then we derive the martingale problems satisfied by these Markov chains and these martingale problems can be used to approximately characterize the martingale problem of the historical FOU which we are interested. It is worthwhile to remark here that our results can be generalized to any stochastic process provided its finite-dimensional distribution is given. In particular, to any Gaussian process where the finite-dimensional distributions are determined by its covariance and mean functions.

2.7.1 Fractional Ornstein-Uhlenbeck Process

Definition 2.9. *The fractional Brownian motion (FBM) $B_t^H, t \in [0, \infty)$ of Hurst index $H \in (0,1)$, is a centered Gaussian process starting from zero with covariance* *function* R_H :

$$R_H(t,s) = \mathbb{E}[B_t^H B_s^H] \triangleq \frac{1}{2}(t^{2H} + s^{2H} - |t-s|^{2H}) \tag{2.101}$$

for any $0 \leq s, t < \infty$.

Proposition 2.7. *The FBM with Hurst index* $H \in (0,1)$ *is of stationary increments and self-similarity in the sense*

$$\frac{1}{a^H} B_{at}^H \stackrel{\circ}{=} B_t^H$$

for $a > 0$ *and* $t \geq 0$, *where* "$\stackrel{\circ}{=}$" *means equal in distributions.*

Proof. For any $0 \leq s, t < \infty$, $B_t^H - B_s^H$ is a central Gaussian random variable. Moreover, from (2.101), it follows immediately that

$$\begin{aligned}
\mathbb{E}[(B_t^H - B_s^H)^2] &= \mathbb{E}[(B_t^H)^2 + (B_s^H)^H - 2B_t^H B_s^H] \\
&= \mathbb{E}(B_t^H)^2 + \mathbb{E}(B_s^H)^2 - 2\mathbb{E}(B_t^H B_s^H) \\
&= t^{2H} + s^{2H} - (t^{2H} + s^{2H} - |t-s|^{2H}) \\
&= |t-s|^{2H}.
\end{aligned}$$

Therefore the FBM has stationary increments. On the other hand, $B_{\alpha t}^H = B_{\alpha t}^H - B_0^H$ is a Gaussian random variable with mean zero and variance

$$\mathbb{E}[(B_{\alpha t}^H)^2] = \mathbb{E}[(B_t^H)^2 + (B_s^H)^2 - 2B_t^H B_s^H] = (\alpha t)^{2H} = (\alpha^H)^2 t^{2H} = (\alpha^H)^2 \mathbb{E}[(B_t^H)^2].$$

Hence the FBM is self-similar. □

Proposition 2.8. *The FBM with Hurst index* $H \in (0,1)$ *admits a version which is* $\beta-$*Hölder continuous almost surely for* $\beta < H$.

Proof. By the self-similarity and stationarity of the increments we have, for $\alpha > 0, t > s$,

$$\mathbb{E}|B_t^H - B_s^H|^\alpha = \mathbb{E}|B_{t-s}^H|^\alpha = (t-s)^{\alpha H}\mathbb{E}|B_1^H|^\alpha = (t-s)^\alpha C_\alpha,$$

where the constant C_α is the $\alpha^{th}-$absolute moment of standard normal random variable. Then, the result follows immediately directly from the Kolmogorov-Čentsov theorem. □

In what follows we shall always use such Hölder continuous version of the fractional Brownian motion.

Proposition 2.9. *The FBM with Hurst index $H \in (0,1)$ is a Markov process if and only of $H = \frac{1}{2}$.*

Proof. From Proposition (11.7) of Kallenberg (1997), we know a Gaussian process is Markovian if and only if its covariance function R_H satisfies

$$R_H(s,u) = \frac{R_H(s,t)R_H(t,u)}{R_H(t,t)}.$$

However, this is true if and only if $H = \frac{1}{2}$ by noting 2.101. □

Remark 2.11. *If $H > \frac{1}{2}$, then B_t^H is a long memory process in the sense that*

$$\sum_{i=1}^{\infty} \mathbb{E}[B_1^H(B_{n+1}^H - B_n^H)] = \frac{1}{2}\sum_{i=1}^{\infty} \left\{(n+1)^{2H} + (n-1)^{2H} - 2n^{2H}\right\} = \infty$$

for all $n \geq 1$. The LM suggests the correlations between long lags observations are not negligible.

Definition 2.10. *The fractional Ornstein-Uhlenbeck (FOU) Process X^H is defined as*

$$\begin{aligned} dX_t^H &= \alpha(\mu - X_t^H)dt + \sigma dB_t^H, \\ X_0^H &= x_0 \end{aligned} \tag{2.102}$$

for $\alpha, \sigma > 0$, μ is the asymptotic mean and B_t^H is a FBM with Hurst index $H \in (\frac{1}{2}, 1)$.

Here, the fractional stochastic integration is introduced in Duncan, Hu and Pasik-Duncan (2000) using the wick product (unlike the pathwise integration introduced in Lin (1995)). The FOU process can capture both the LM and mean reverting properties.

Proposition 2.10. *The FOU process defined in* (2.102) *can be represented explicitly as*

$$X_t^H = \mu + (x_0 - \mu)e^{-\alpha t} + \sigma e^{-\alpha t} \int_0^t e^{\alpha s} dB_s^H \tag{2.103}$$

Proof. Let $Y_t^H = X_t^H \cdot e^{\alpha t}$, then apply the fractional Itô formula (see Duncan, Hu and Pasik-Duncan (2000)), we have

$$\begin{aligned} dY_t^H &= \alpha \cdot e^{\alpha t} X_t^H dt + e^{\alpha t} dX_t^H \\ &= e^{\alpha t} \alpha \mu dt + \sigma e^{\alpha t} dB_t^H. \end{aligned} \tag{2.104}$$

Then it follows that

$$\begin{aligned} Y_t^H &= x_0 + \alpha\mu \int_0^t e^{\alpha s} ds + \sigma \int_0^t e^{\alpha s} dB_s^H \\ &= x_0 + \mu(e^{\alpha t} - 1) + \sigma \int_0^t e^{\alpha s} dB_s^H \end{aligned}$$

$$X_t^H = e^{-\alpha t} Y_t^H = \mu + (x_0 - \mu)e^{-\alpha t} + \sigma e^{-\alpha t} \int_0^t e^{\alpha s} dB_s^H.$$

Hence the result. □

Proposition 2.11. *The FOU process (2.102) is a continuous Gaussian process with the mean function*

$$m(t) \triangleq \mathbb{E}(X_t^H) = \mu + (x_0 - \mu)e^{-\alpha t}, \tag{2.105}$$

and for $0 \leq s, t < \infty$, *covariance function*

$$Cov(X_t^H, X_s^H) = \sigma^2 e^{-\alpha(t+s)} \int_0^t \int_0^s e^{\alpha u} e^{\alpha v} \varphi(u, v) du dv, \tag{2.106}$$

where the kernel $\varphi : [0, \infty) \times [0, \infty) \longrightarrow \mathbb{R}$ *be given by*

$$\varphi(s, t) \triangleq H(2H - 1)|s - t|^{2H-2}. \tag{2.107}$$

In particular, the FOU has the variance function

$$\text{(2.108)} \qquad Var(X_t^H) = \sigma^2 e^{-2\alpha t} \int_0^t \int_0^t e^{\alpha u} e^{\alpha v} \varphi(u,v) du dv.$$

Proof. From the construction of fractional stochastic integration of Duncan, Hu and Pasik-Duncan (2000), we know the fractional stochastic integration is a Gaussian process with continuous sample path when the integrand is deterministic. Thus from (2.103), we know for fixed t, X_t^H is Gaussian random variable. The expectation of the fractional stochastic integration is zero mean (see Hu, Øksendal and Sulem (2003)) and for for $0 \leq s, t < \infty$, the covariance function is

$$\begin{aligned} Cov(X_t^H, X_s^H) &= \mathbb{E}\left(X_t^H - \mathbb{E}X_t^H\right)\left(X_s^H - \mathbb{E}X_s^H\right) \\ &= \sigma^2 e^{-\alpha(t+s)} \mathbb{E}\left(\int_0^t e^{\alpha u} dW_u^H \cdot \int_0^s e^{\alpha v} dW_v^H\right) \\ &= \sigma^2 e^{-\alpha(t+s)} \int_0^t \int_0^s e^{\alpha u} e^{\alpha v} \varphi(u,v) du dv. \end{aligned}$$

Here, the kernel $\varphi : [0, \infty) \times [0, \infty) \longrightarrow \mathbb{R}$ is given by

$$\varphi(s,t) \triangleq H(2H-1)|s-t|^{2H-2}.$$

Here, (2.106) is from Duncan, Hu and Pasik- Duncan (2000). It is the generalized form of the fractional Itô isometry in Hu, Øksendal and Sulem (2003) which gives the variance function

$$\text{(2.109)} \qquad Var(X_t^H) = \sigma^2 e^{-2\alpha t} \int_0^t \int_0^t e^{\alpha u} e^{\alpha v} \varphi(u,v) du dv.$$

Hence the result. □

O-U Process

An useful special case is when $H = \frac{1}{2}$ and X^H becomes the O-U process. For notation consistency with later chapters, we use Z instead X to denote the O-U process.

Proposition 2.12. *The solution to the O-U equation*

$$dZ_t = -\alpha_Z Z_t dt + \alpha_Z dW_t, \quad Z_0 = 0$$

is a central Gaussian process with covariance function

$$\rho(s,t) \triangleq Cov(Z_s, Z_t) = \frac{\alpha_Z}{2}\left(e^{-\alpha_Z(t-s)} - e^{-\alpha_Z(t+s)}\right).$$

Proof. Let $Y_t = Z_t \cdot e^{\alpha_Z t}$, then from the Itô formula, we have

$$dY_t = \alpha_Z \cdot e^{\alpha_Z t} Z_t dt + e^{\alpha_Z t} dZ_t = \alpha_Z \cdot e^{\alpha_Z t} dW_t.$$

Thus

$$Z_t = \alpha_Z \cdot e^{-\alpha_Z t} Y_t = \alpha_Z \cdot e^{-\alpha_Z t} \int_0^t e^{\alpha_Z s} dW_s.$$

This is a Gaussian process with mean function

$$\mathbb{E}[Z_t] = 0$$

and

$$Var(Z_t) = \frac{\alpha_Z}{2}\left(1 - e^{-2\alpha_Z t}\right).$$

Moreover, for $0 \le s < t < \infty$, the covariance function is

$$\rho(s,t) = \frac{\alpha_Z}{2}\left(e^{-\alpha_Z(t-s)} - e^{-\alpha_Z(t+s)}\right).$$

□

2.7.2 Historical Process of FOU

Although we can derive the explicit form of FOU process, however, except $H = \frac{1}{2}$, the FBM and FOU is neither semi-martingale nor Markov process. This makes the stochastic analysis to them becomes more infeasible. Following Dynkin (1991), Kouritzin, Long and Sun (2003), for any stochastic process S, we can introduce its historical process $\widehat{S}$ whose state at time t is the path of S over time interval $[0, t]$. The principle property of historical process is that, for any $0 \le t < \infty$, the σ-algebra $\sigma(\widehat{S}_t)$ coincides with $\mathcal{F}_t^{\widehat{S}}$. From now on, we confine our attentions to the historical process $\widehat{X}$ of the FOU X^H. Its historical process can be introduced through its historical variant which is defined path-by-path, that is, for $\omega \in \Omega$, the realization of $X^H(\cdot,\omega)$, the historical variant of $\widehat{X}$ is defined at time t as

$$\widehat{X}_t(\tau) = \begin{cases} X_\tau^H, & \text{for } 0 \le \tau \le t, \\ X_t^H, & \text{for } t < \tau < \infty. \end{cases} \tag{2.110}$$

It follows directly that any historical process is a Markov process because its current status includes all information of the whole past trajectory. In general case, for stochastic process S with state space K and right continuous left limit (càdlàg) sample paths, the state space of its historical process is $D_K[0,\infty)$, the space of all right continuous left limit (càdlàg) functions. In case of the FOU process, the state space of its historical process is $C_{\mathbb{R}}[0,\infty)$, the space of all real-valued continuous functions on $[0,\infty)$. Hereafter, for notation simplicity, we write X instead X^H to denote the FOU process with Hurst index H. In the same way, we write $\widehat{X}$ instead $\widehat{X}^H$ to denote the historical process of FOU.

Here, we aim to derive the martingale problem satisfied by the historical process of FOU process. Keep this in mind, we first introduce a sequence of discrete-parameter Markov chain which point-wise converges to the historical process of FOU process. Using the transition function of the Markov chain, we can characterize its evolution through some discrete-time martingale problem.

General Partition

For $n \in \mathbb{N}$, we introduce the partition I_n of $[0,\infty)$, an ordered subset

$$I_n = \{t_0^n, t_1^n, \cdots\} \subset [0,\infty) \tag{2.111}$$

satisfying

$$0 = t_0^n < t_1^n < \cdots < t_k^n < \cdots < \infty,$$

and

$$\lim_{k\longrightarrow\infty} t_k^n = \infty.$$

The mesh of I_n is defined as

$$|I_n| \triangleq \max\{|t_{j+1}^n - t_j^n|, j = 0, 1, \cdots\}.$$

Note that $\{I_n, n \in \mathbb{N}\}$ is a sequence of partitions of $[0, \infty)$ and for each n, we let $t_j^n, j = 0, 1, \cdots$ denote the members of I_n. To ease the burden of notation, we omit the superscript n in t_j^n below whenever the meaning of I_n is clear from the context.

Remark 2.12. *A particular and simple case is the following equalized partition*

$$I_n = \{0, \frac{T}{n}, \cdots, \frac{iT}{n}, \cdots\} \subset [0, \infty). \tag{2.112}$$

Now define the discrete-parameter chain $X^{I_n} = \{X^{I_n}(j), j = 0, 1, \cdots\}$ where

$$X^{I_n}(j) = X_{t_j} \quad \text{for } j = 0, 1, \cdots \tag{2.113}$$

We know X^{I_n} is not Markov because it is the discrete-parameter sampling of FOU. Alternatively, we can introduce the *historical chain* $\widehat{X}^{I_n} = \{\widehat{X}_k^{I_n}, k = 0, 1, \cdots\}$ where $\widehat{X}_k^{I_n} \in \mathbb{R}^\infty$ is defined as

$$\widehat{X}_k^{I_n}(j) = \begin{cases} X_{t_j}, & \text{for } 0 \le j \le k, \\ X_{t_k}, & \text{for } j > k. \end{cases} \tag{2.114}$$

That is,

$$\widehat{X}_k^{I_n} = \{X_{t_0}, X_{t_1}, \cdots, X_{t_k}, X_{t_k}, \cdots\}. \tag{2.115}$$

Now we can embed X^{I_n} into some right continuous-time process $X_t^{I_n}$ with càdlàg sample paths through

$$X_t^{I_n} = \sum_{i=0}^{\infty} X_{t_i} 1_{[t_i, t_{i+1})}(t) = \begin{cases} X_0, & \text{if } t \in [0, t_1), \\ X_{t_1}, & \text{if } t \in [t_1, t_2), \\ \cdots \\ X_{t_i}, & \text{if } t \in [t_i, t_{i+1}), \\ \cdots \end{cases} \tag{2.116}$$

and introduce its historical process $\widehat{X}_t^{I_n}$ by

$$(2.117)\qquad \widehat{X}_t^{I_n}(\tau) = \sum_{i=0}^{\infty} X_{t_i} 1_{[t_i, t_{i+1})}(t \wedge \tau) = \begin{cases} X_0, & \text{if } t \wedge \tau \in [0, t_1), \\ X_{t_1}, & \text{if } t \wedge \tau \in [t_1, t_2), \\ \cdots & \\ X_{t_i}, & \text{if } t \wedge \tau \in [t_i, t_{i+1}), \\ \cdots & \end{cases}$$

Because the sample paths of the FOU process X are almost surely continuous on $[0, \infty)$, thus we have path-wise convergence, that is,

$$(2.118)\qquad \widehat{X}_t^{I_n} \Longrightarrow \widehat{X}_t$$

in $C_{\mathbb{R}}[0, \infty)$ when $\lim_{n \longrightarrow \infty} |I_n| = 0$. Therefore, the martingale problem of the historical process $\widehat{X}_t$ should be approximately determined by the martingale problems of $\{\widehat{X}_t^{I_n}\}_{n \in \mathbb{N}}$. Meanwhile, from its definition, it follows that $\widehat{X}_t^{I_n}$ is equivalent to $\widehat{X}_j^{I_n}$ in the sense that they carry the same information. Here, j satisfies $t \in [t_j, t_{j+1})$. Therefore, we limit our attention to the martingale problem of $\widehat{X}^{I_n}$ in the following. It is obvious that $\widehat{X}^{I_n}$ is a discrete-parameter Markov chain because its historical $\sigma-$algebra

$$\mathcal{F}_k^{\widehat{X}^{I_n}} \triangleq \sigma\{\widehat{X}_1^{I_n}, \widehat{X}_2^{I_n}, \cdots, \widehat{X}_k^{I_n}\}$$

is generated by $\widehat{X}_k^{I_n}$. To derive its martingale problem, we need specify the one-step transition function of $\widehat{X}^{I_n}$ from k to $k+1$. For $k \in \mathbb{N}$, we can introduce

$$(2.119)\qquad \mathbb{R}_k^{\infty} = \{x \in \mathbb{R}^{\infty} \text{ such that } x_{k+j} = x_k \text{ for } \forall j \in \mathbb{N}\},$$

and the $k-$dimensional projection Π_k for $x = (x_1, x_2, \cdots) \in \mathbb{R}^{\infty}$:

$$(2.120)\qquad \Pi_k\, x = (x_1, \cdots, x_k),$$

for $\Gamma \subset \mathbb{R}^{\infty}$:

$$(2.121)\qquad \Pi_k \Gamma = \{\Pi_k\, x, x \in \Gamma).$$

Moreover, for $x \in \mathbb{R}^\infty, \Gamma \subset \mathbb{R}^\infty$, we can define the $k-$section set of Γ at x:

$$\Gamma_x^{k,k+1} \triangleq \{y : \quad (\Pi_k x, y) \in \Pi_{k+1}\left(\Gamma \cap \mathbb{R}_{k+1}^\infty\right)\}. \tag{2.122}$$

Remark 2.13. *If* $\Gamma \cap \mathbb{R}_{k+1}^\infty = \emptyset$, *then* $\Gamma_x^{k,k+1} = \emptyset$.

Now, for $x \in \mathbb{R}^\infty, \Gamma \in \mathcal{B}(\mathbb{R}^\infty)$,

$$\begin{aligned} \mu(k,k+1,x,\Gamma) &\triangleq \mathbb{P}(\widehat{X}_k^{I_n} \in \mathbb{R}_k^\infty,\ \ \widehat{X}_{k+1}^{I_n} \in \Gamma | \widehat{X}_k^{I_n} = x) \\ &= \mathbb{P}(\widehat{X}_k^{I_n} \in \mathbb{R}_k^\infty,\ \ \Pi_{k+1}\widehat{X}_{k+1}^{I_n} \in \Pi_{k+1}(\Gamma \cap \mathbb{R}_{k+1}^\infty) | \widehat{X}_k^{I_n} = x) \\ &= \mathbb{P}(\widehat{X}_k^{I_n} \in \mathbb{R}_k^\infty,\ \ \widehat{X}_{k+1}^{I_n}(k+1) \in \Gamma_x^{k,k+1} | \widehat{X}_k^{I_n} = x) \\ &= \mathbb{P}(\widehat{X}_k^{I_n} \in \mathbb{R}_k^\infty,\ \ X^{I_n}(k+1) \in \Gamma_x^{k,k+1} | \widehat{X}_k^{I_n} = x) \\ &= \int_{\Gamma_x^{k,k+1}} 1_{\{x \in \mathbb{R}_k^\infty\}} p(y|\Pi_k\, x) dy. \end{aligned} \tag{2.123}$$

Here, $p(y|\cdot)$ is the conditional probability density function of $X^{I_n}(k+1) = y$ given $\Pi_k \widehat{X}_k^{I_n}$. Note that $(\Pi_k \widehat{X}_k^{I_n}, X^{I_n}(k+1))$ is a Gaussian random vector with mean $\mu = 0$, thus we have:

$$\begin{aligned} p(y|\Pi_k\, x) &= p(y|x_1, x_2, \cdots, x_k) \\ &= \frac{|D_k|^{\frac{1}{2}}}{\sqrt{2\pi}|D_{k+1}|^{\frac{1}{2}}} e^{-\frac{1}{2}[(\Pi_k x, y) D_{k+1}^{-1} (\Pi_k x, y)' - \Pi_k x D_k^{-1} \Pi_k x']}, \end{aligned} \tag{2.124}$$

where for each $k \in \mathbb{N}_0$, D_k is determined by the covariance function of FOU process which is given in (2.106). It is easy to check that (2.123) is a one-step transition function to the Markov chain $\{\widehat{X}_k^{I_n}\}$. With the transition function in hand, we can show the martingale problem satisfied by $\widehat{X}_k^{I_n}$. For $k \in \mathbb{N}$, define the time-invariant operator $\mathbf{A}_k^{I_n} : B(\mathbb{R}^\infty) \longrightarrow B(\mathbb{R}^\infty)$ by

$$\mathbf{A}_k^{I_n} f(x) = \int f(y) \mu(k, k+1, x, dy) - f(x). \tag{2.125}$$

Therefore, from Ethier and Kurtz (1986), the discrete-time martingale problem for $\widehat{X}^{I_n}$ can be described as follows:

$$f(\widehat{X}_k^{I_n}) = f(\widehat{X}_0^{I_n}) + \sum_{i=0}^{k} \mathbf{A}_i^{I_n} f(\widehat{X}_i^{I_n}) + M_k^{f,I_n}. \tag{2.126}$$

Consequently, the martingale problem to the historical process of FOU process can be described as follows:

$$f^*(\widehat{X}_t) = f^*(\widehat{X}_0) + \int_0^t \mathbf{A}_s f^*(\widehat{X}_s)ds + M_t^{f^*}, \tag{2.127}$$

where the test function $f^* \in B(D_R[0,\infty))$ takes the following form: there exists $0 < t_1 < t_2 < \cdots < t_k < \cdots < \infty$, $f \in B(\mathbb{R}^\infty)$ such that

$$\begin{aligned} f^*(\widehat{X}_t) &= f \circ (\pi_{t_1}, \pi_{t_2}, \cdots, \pi_{t_k,\cdots})(\widehat{X}_t) \\ &= f(\pi_{t_1}\widehat{X}_t, \pi_{t_2}\widehat{X}_t, \cdots, \pi_{t_k}\widehat{X}_t, \cdots) \\ &= f(X_{t\wedge t_1}, X_{t\wedge t_2}, \cdots X_{t\wedge t_k}, \cdots). \end{aligned} \tag{2.128}$$

Here, for $s \in [0,\infty)$, π_s (note that it is not the unnormalized filter defined in (2.12)) is the projection operation on $B(D_R[0,\infty))$ at s such that

$$\pi_s \widehat{X}_t = X_{s\wedge t} = \begin{cases} X_s, & \text{for } s \le t, \\ X_t, & \text{for } s > t. \end{cases} \tag{2.129}$$

Consequently, the generator $\mathbf{A}_s$ is defined as: for all f^* which is of the form in (2.128):

$$\mathbf{A}_t f^* = \sum_{k=1}^{\infty} \frac{1}{t_k - t_{k-1}} \mathbf{A}_k^{I^n} f^* \cdot 1_{[t_{k-1},t_k)}(t), \tag{2.130}$$

where

$$I^n = \{0 = t_0 < t_1 < \cdots < t_k < \cdots < \infty\}. \tag{2.131}$$

Chapter 3

Microstructural Stock Market

3.1 Introduction

In principle, the financial market can be modeled by the micro- or macro-structure approaches. The macrostructure (macro-movement) refers to daily, weekly or monthly closing price which are of the low frequency. In contrast, the microstructure (micro-movement) refers to high-frequency transaction (trade-by-trade) prices. The availability of such high frequency data provides researchers an opportunity to study the market at any scale. There is a large amount of literature discussing microstructure financial markets (see Black (1986), Chan and Lakonishok (1993), Hasbrouck (1996, 1999), Engle and Russell (1998), Engle (2000), Bandi and Russell (2006) etc.). Unlike the macrostructure market, the trading noises in microstructure market are not negligible thus the intrinsic value of the asset can not be observed directly. In this paper, we introduce a class of partially-observed microstructure models where the asset price can be formulated as distorted, corrupted counting measure observations of a macrostructure value-process model. Previously, Zeng (2003) studied the Duncan-Mortensen-Zakai (DMZ) equation while Kouritzin and Zeng (2005) derive their Bayes factor equation for a motivating yet more-limited microstructure model.

In this chapter, we discuss the model selection problem in microstructure market. The motivation is to evaluate which of the competing stochastic volatility models best fits the observed transaction data. Our Bayes factor method, provides an effective means to conduct our statistical comparisons since it provides real time empirical evidence as to which model best fits the market data while allowing the value-process model components to be mathematically singular to one another. When applying the Bayes factor method, we need only transform all observation model components into the same canonical process via Girsanov measure change. The interested readers may refer to Kass and Raftery (1995) for a comprehensive survey of Bayes factor and model selection. Kouritzin and Zeng (2005) discuss the Bayesian model selection problem to determine which of a class of financial models best represents given financial data such as stock price. However, their work is based on a far cruder model with obvious limitations and the DMZ filtering equation instead of the robust filtering equation. One advantage of our robust equation is that there is only one time scale "dt" here while in the DMZ filtering equation, there have two time scales: "dt" and "dY_t" which complicate the numerical computation.

Our use of robust filter is both practical and novel. To the best of our knowledge, neither the linear nor nonlinear robust filter has ever been applied to the research of microstructure financial market before. We fill this gap and also first characterize the Bayes factor of model selection in terms of the robust nonlinear filter. The rest of this paper proceeds as follows: Section 2 lays out the partially-observed microstructure model to explain the existing price bias in intrinsic value. Section 3 discusses the Bayes estimation problem and presents the evolution equations that characterize the Bayes filter. We also provide a novel efficient particle filtering algorithm to implement these equations. Some numerical results are reported. Section 4 investigates the model selection problem using Bayes factor method to test which SV model best fits the observed price data.

3.2 The Partially-Observed Microstructure Model

3.2.1 Construction of the State Process

Our goal is to evaluate and compare SV models within the microstructure framework. Throughout this paper, we assume the financial state $X \in \mathbb{R}^{n_x}$ with its parameter $\theta \in \mathbb{R}^{n_\theta}$ jointly satisfy the martingale problem (2.1).

Remark 3.1. *The martingale problem technique proposed by Stroock and Varadhan*

(1979) provides a general formulation of the Markov processes. See Ethier and Kurtz (1986) for more details of the general martingale problem and the associated operator approach.

The formulation (2.1) covers most of the interesting financial quantities, e.g., stock price, interest rate, exchange rate or commodity price. Although we focus on the stock price throughout this paper, the methods proposed here can just as easily be applied to other classes of financial assets. In our paper, the state process X consists of two components: the equilibrium price (intrinsic value) S and the possible stochastic volatility V. Both are latent variables. The most commonly seen example of (2.1) in finance is the *geometric Brownian motion* (GBM) arising from the classical Black-Scholes (BS) option pricing formula.

3.2.2 Stochastic Volatility Models

Example 3.1. (GBM model)*(see Black and Scholes (1973), Merton (1973))*

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t, \tag{3.1}$$

where $n_x = 1, n_\theta = 2$, $X_t = S_t$, $\theta = (\mu, \sigma)$ and W_t is standard Brownian motion. The generator $\mathbf{A}^{(1)}$ is

$$\mathbf{A}^{(1)} f(s,\theta) = \frac{1}{2}\sigma^2 s^2 \frac{d^2 f}{ds^2}(s,\theta) + \mu s \frac{df}{ds}(s,\theta). \tag{3.2}$$

To account for the well-known *volatility smile* (see Jackwerth, J. and Rubinstein, M. (1996) for a detailed survey) observed from the market option prices, the GBM model is generalized to stochastic volatility (SV) models, where the volatility σ itself also follows some stochastic process. In the sequel, S and V will be used to denote the stock value and its volatility respectively. Then, some of the popular SV models include:

Example 3.2. (Hull &White model)*(see Hull & White, 1987)*

$$\frac{dS_t}{S_t} = \mu dt + V_t^{\frac{1}{2}} dW_t, \tag{3.3}$$

$$\frac{dV_t}{V_t} = \nu dt + \kappa dB_t, \tag{3.4}$$

where $n_x = 2, n_\theta = 3$, $X_t = (S_t, V_t)$, $\theta = (\mu, \nu, \kappa)$ and W_t and B_t are independent standard Brownian motions. The generator $\mathbf{A}^{(2)}$ is

$$\begin{aligned} \mathbf{A}^{(2)} f(s,v,\theta) = &\frac{1}{2} v s^2 \frac{\partial^2 f}{\partial s^2}(s,v,\theta) + \mu s \frac{\partial f}{\partial s}(s,v,\theta) + \frac{1}{2}\kappa^2 v^2 \frac{\partial^2 f}{\partial v^2}(s,v,\theta) \\ &+ \nu v \frac{\partial f}{\partial v}(s,v,\theta). \end{aligned} \tag{3.5}$$

Example 3.3. (Logrithmetic Ornstein-Uhlenbeck model) *(see Scott, 1987.)*

$$\frac{dS_t}{S_t} = \mu dt + V_t dW_t, \tag{3.6}$$

$$dV_t = V_t \left(\frac{1}{2}\nu^2 - \varrho(\ln V_t - \varpi) \right) dt + \kappa V_t dB_t, \tag{3.7}$$

where $n_x = 2, n_\theta = 5$, $X_t = (S_t, V_t)$, $\theta = (\mu, \nu, \varrho, \varpi, \kappa)$ and W_t and B_t are independent standard Brownian motions. The generator $\mathbf{A}^{(3)}$ of this model is

$$\begin{aligned} \mathbf{A}^{(3)} f(s,v,\theta) = &\frac{1}{2} v^2 s^2 \frac{\partial^2 f}{\partial s^2}(s,v,\theta) + \mu s \frac{\partial f}{\partial s}(s,v,\theta) + \frac{1}{2}\kappa^2 v^2 \frac{\partial^2 f}{\partial v^2}(s,v,\theta) \\ &+ v \left(\frac{1}{2}\nu^2 - \varrho(\ln v - \varpi) \right) \frac{\partial f}{\partial v}(s,v,\theta). \end{aligned} \tag{3.8}$$

Example 3.4. (GARCH (1,1) diffusion model) *(see Nelson, 1990.)*

$$\frac{dS_t}{S_t} = \mu dt + V_t^{\frac{1}{2}} dW_t, \tag{3.9}$$

$$dV_t = (\nu - \varrho V_t) dt + \kappa V_t dB_t, \tag{3.10}$$

where $n_x = 2, n_\theta = 4$, $X_t = (S_t, V_t)$, $\theta = (\mu, \nu, \varrho, \kappa)$ and W_t, B_t are independent standard Brownian motions. The generator $\mathbf{A}^{(4)}$ of this model is

$$\begin{aligned} \mathbf{A}^{(4)} f(s,v,\theta) = &\frac{1}{2} v s^2 \frac{\partial^2 f}{\partial s^2}(s,v,\theta) + \mu s \frac{\partial f}{\partial s}(s,v,\theta) + \frac{1}{2}\kappa^2 v^2 \frac{\partial^2 f}{\partial v^2}(s,v,\theta) \\ &+ (\nu - \varrho v) \frac{\partial f}{\partial v}(s,v,\theta). \end{aligned} \tag{3.11}$$

Example 3.5. (Heston model)*(see Heston, 1993.)*

$$\frac{dS_t}{S_t} = \mu dt + V_t^{\frac{1}{2}} dW_t, \tag{3.12}$$

$$dV_t = (\nu - \varrho V_t)dt + \kappa V_t^{\frac{1}{2}} dB_t, \tag{3.13}$$

where $n_x = 2, n_\theta = 4$, $X_t = (S_t, V_t)$, $\theta = (\mu, \nu, \varrho, \kappa)$ *and* W_t, B_t *are independent standard Brownian motions. The generator* $\mathbf{A}^{(5)}$ *of this model is*

$$\begin{aligned}\mathbf{A}^{(5)} f(s,v,\theta) = \frac{1}{2} v s^2 \frac{\partial^2 f}{\partial s^2}(s,v,\theta) + \mu s \frac{\partial f}{\partial s}(s,v,\theta) + \frac{1}{2}\kappa^2 v \frac{\partial^2 f}{\partial v^2}(s,v,\theta) \\ + (\nu - \varrho v)\frac{\partial f}{\partial v}(s,v,\theta).\end{aligned} \tag{3.14}$$

Example 3.1 − 3.5 can be summarized by the following *"generalized diffusion stochastic volatility "* (GDSV) model (see Nielsen and Vestergaard, 2000):

$$dS_t = b_1(S_t, V_t, \theta)dt + c_1(S_t, V_t, \theta)dW_t, \tag{3.15}$$

$$dV_t = b_2(S_t, V_t, \theta)dt + c_2(S_t, V_t, \theta)dB_t, \tag{3.16}$$

$$X_0 = (S_0, V_0).$$

where $(X_0, \theta_0) \in \mathbb{R}^{n_\theta + 2}$, $\mu = \mathbb{P} \circ (X_0, \theta_0)^{-1}$, W_t and B_t are independent standard Brownian motions. The coefficients b_1, b_2 and c_1, c_2 satisfy the regularity conditions so the existence and uniqueness of the strong solution is known for all these examples. The generator $\mathbf{A}$ in this general case is time-invariant and satisfies

$$\begin{aligned}\mathbf{A} f(s,v,\theta) = \frac{1}{2} c_1^2(s,v,\theta) \frac{\partial^2 f}{\partial s^2}(s,v,\theta) + b_1(s,v,\theta)\frac{\partial f}{\partial s}(s,v,\theta) \\ + \frac{1}{2} c_2^2(s,v,\theta)\frac{\partial^2 f}{\partial v^2}(s,v,\theta) + b_2(s,v,\theta)\frac{\partial f}{\partial v}(s,v,\theta),\end{aligned} \tag{3.17}$$

where $\mathcal{D}(\mathbf{A})$ is the set of all bounded second-order continuously differentiable functions on $\mathbb{R}^{n_\theta+2}$. For comprehensive comparison, we also consider the following *jumping stochastic volatility geometric Brownian motion* (JSV-GBM) model.

Example 3.6. (JSV-GBM model)*(see Kouritzin and Zeng, 2005.)*

$$\frac{dS_t}{S_t} = \mu dt + V_t dW_t, \tag{3.18}$$

$$dV_t = (J_{N_{t-}+1} - V_{t-})dN_t, \tag{3.19}$$

where W_t is a standard Brownian motion, N_t is a Poisson process with intensity λ; $\{J_i\}$ is a sequence of independent random variables independent of W_t, N_t and uniformly distributed on a range $[\alpha_J, \beta_J]$. Here, W_t and N_t are also independent and in this case $n_x = 2, n_\theta = 4$, $X_t = (S_t, V_t)$, $\theta = (\mu, \lambda, \alpha_J, \beta_J)$. The generator $\mathbf{A}^{(6)}$ of this model is

$$\begin{aligned} \mathbf{A}^{(6)} f(s,v,\theta) = &\frac{1}{2} v^2 s^2 \frac{\partial^2 f}{\partial s^2}(s,v,\theta) + \mu s \frac{\partial f}{\partial s}(s,v,\theta) \\ &+ \lambda \int_{\alpha_J}^{\beta_J} (f(s,z,\theta) - f(s,v,\theta)) \frac{1}{\beta_J - \alpha_J} dz. \end{aligned} \tag{3.20}$$

In summary, we have

SV Model	Example	State X	State Parameter θ	Generator
GBM	1	S	(μ, σ)	$\mathbf{A}^{(1)}$
Hull & White	2	(S, V)	(μ, ν, κ)	$\mathbf{A}^{(2)}$
Log O-U	3	(S, V)	$(\mu, \nu, \varrho, \varpi, \kappa)$	$\mathbf{A}^{(3)}$
GARCH Diffusion	4	(S, V)	$(\mu, \nu, \varrho, \kappa)$	$\mathbf{A}^{(4)}$
Heston	5	(S, V)	$(\mu, \nu, \varrho, \kappa)$	$\mathbf{A}^{(5)}$
JSV-GBM	6	(S, V)	$(\mu, \lambda, \alpha_J, \beta_J)$	$\mathbf{A}^{(6)}$

To derive the full benefits, we give the following remarks:

Remark 3.2. *In Example 3.3, applying Itô formula to $V_t^* = \ln V_t$, we get the following formulation*

$$\frac{dS_t}{S_t} = \mu dt + e^{V_t^*} dW_t, \tag{3.21}$$

$$dV_t^* = (a + bV_t^*)dt + cdB_t, \tag{3.22}$$

where $\theta = (\mu, a, b, c)$ satisfies $a = \frac{1}{2}\nu^2 + \varrho\varpi - \frac{1}{2}\kappa^2$; $b = -\varrho$; $c = \kappa$. It follows that V^* is a mean-reverting O-U process. The generator becomes

$$\begin{aligned} \mathbf{A}^* f(s,v^*,\theta) = &\frac{1}{2} e^{2v^*} s^2 \frac{\partial^2 f}{\partial s^2}(s,v^*,\theta) + \mu s \frac{\partial f}{\partial s}(s,v^*,\theta) + \frac{1}{2} c^2 \frac{\partial^2 f}{\partial^2 v^*}(s,v^*,\theta) \\ &+ (a + bv^*) \frac{\partial f}{\partial v^*}(s,v^*,\theta). \end{aligned} \tag{3.23}$$

Remark 3.3. *The GARCH diffusion model in Example* 3.4 *is the continuous-time limit of many classical GARCH-type discrete-time processes as discussed in Nelson (1990), Drost and Werker (1996).*

Remark 3.4. *The volatility drift parameters* (ν, ϱ, ϖ) *in Example* 3.3, (ν, ϱ) *in Example* 3.4, 3.5 *are introduced to capture the mean-reverting nature of the volatility process whereas Example* 3.1, 3.2, 3.6 *are not mean-reverting.*

Remark 3.5. *The Example* 6 *is closely related to the asset price models with Markov modulated volatilities discussed in Elliott, Malcolm and Tsoi (2003).*

The above SV models account for the random behaviors of implied and historical variances of the asset return rate from numerous empirical studies. Meanwhile, the SV models are also crucial for the derivative pricing and hedging in financial markets. A naturally raised but important issue is the parameter estimation of these SV models and model specification. However, estimating the above SV models poses substantial difficulties. One difficulty is that their transition probability functions are hard to get in closed form thus it is not easy to implement the maximum likelihood estimation (MLE) except a few special cases (see Aït-Sahalia, Kimmel (2007)). Even when we can get the closed-form of transition function but it usually turns out to be irregular. Hence, in this paper, we choose the Bayesian filtering approach due the following properties: (1) The Bayes estimate is the least mean square error estimates and does not require the availability or regularity of the likelihood function. (2) The Bayes estimate can be computed recursively. (3) The Bayesian hypothesis testing procedures can be conducted through the Bayes factor that is the ratio of conditional likelihoods and easily computable. Most importantly, it does not require the signals X in our examples $3.1-3.6$ to be equivalent in the sense of Girsanov measure change, which they are met.

3.2.3 Construction of the Observation Process

Now we turn to the construction of the transaction price $\overrightarrow{Y}$ in microstructure market. It is the observation process in our problem. For illustration purpose, we present Figure 3.1 by plotting the transaction prices of Microsoft (MSFT) in March, 1994. The data are extracted from the database of Trade and Quote (TAQ) distributed by New York Stock Exchange (NYSE).There are three prominent features of Y im-

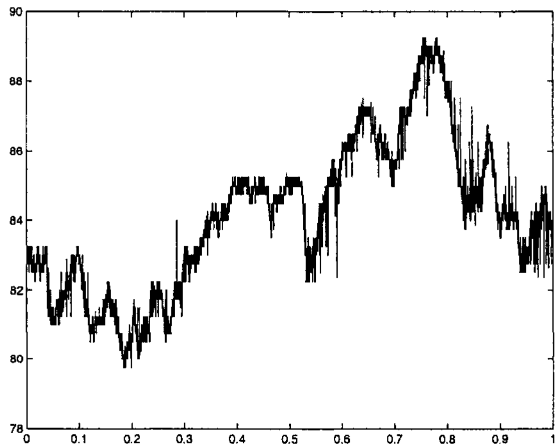


Figure 3.1: The transaction prices of MSFT in March 1994.

plied in Figure 3.1: (1) As well documented in Black (1986), Chan and Lakonishok (1993), Engle (2000), there exist significant trading noises within the market's microstructure which make the transaction price Y deviate from the equilibrium price S. (2) Compared to the state process X which takes value in continuous space, the observation process Y resides in a discrete space

$$(y_0 = 0, y_1 = \frac{1}{M}, \cdots y_j = \frac{j}{M}, \cdots)$$

given by the multiples of the minimum price variation $\frac{1}{M}$ for some positive integer M. We called such minimal variation in financial market the "tick" [1]. Herein, we take $M = 100$ unless otherwise stated. (3) Unlike the macrostructural price, price changes occur only at irregularly spaced trading times (see Engle (2000)): $(T_1, T_2, \cdots)$. Its total intensity $a(X_t, \theta_t, t)$, representing the trading activity, is a time-varying measurable function of x and θ.

Incorporating Trading Noises

Now we show how to incorporate trading noises into Y. Firstly, it is apparent from Figure 1 that prices are restricted on the grid of ticks and that there are a significant number of outliers representing trading noise. Secondly, it is clear that there can be

[1]The tick size in 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.

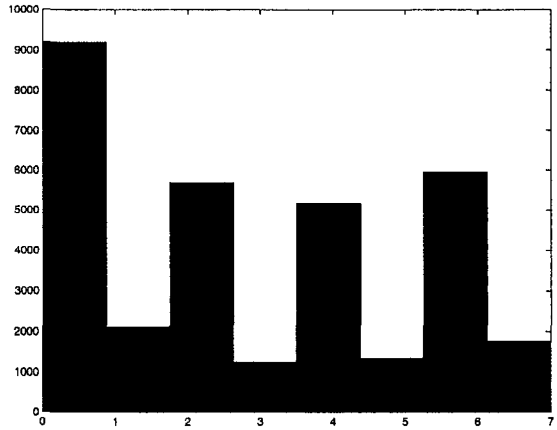


Figure 3.2: The strong clustering of MSFT stock prices in March 1994 where $M = 8$.

price biasing to certain "more whole" prices. For example, a price of \$25 may be more common than $\$24\frac{7}{8}$ or $\$25\frac{1}{8}$. This effect can be seen from Figure 3.2, the histogram of transaction prices of MSFT in March 1994, where the transactions whose tick fraction part is 0 is significantly more than other ticks. Note that here we use older data (March 1994) in these figures for expository reasons. In the sequel, we will also consider more recent data. Our microstructure model is constructed according to the following lines: at each trading time T_i, the price Y_{t_i} is

$$Y_{T_i} = F(X_{T_i}, T_i) \tag{3.24}$$

where $y = F(x, t)$ is a random transform with some transition probability function $p(y|x, t)$. The formulation (3.24) is similar to that of Hasbrouck (1996): X is the intrinsic and permanent component while F acts as the transitory component due to the existence of the trading noises. In general, there are three trading noises in microstructure market: discrete, clustering and non-clustering. As a result, we move from the value X to price T in three steps.

- To incorporate the non-clustering noise. For the i^{th}–transaction with trading time T_i, we define the intermediate variable $\mathcal{Y}_{T_i}$ as:

$$\ln \mathcal{Y}_{T_i} = \ln X_{T_i} + Z^h_{T_i} + \epsilon\zeta_i, \tag{3.25}$$

$$dZ^h_t = -\alpha_Z Z^h_t dt + dW^h_t, \quad Z^h_0 = z_0, \tag{3.26}$$

where the observation noise consists of two components: Z^h and ζ. Here, Z^h is the correlation component of the non-clustering noise while $\zeta = \{\zeta_i\}_{i=1}^{\infty}$ is a sequence of white noise with variance $\epsilon > 0$. The correlated noise component Z^h satisfies the FOU velocity process; its initial value z_0 is independent of W^h. Here, W^h is a fractional Brownian motion with Hurst index $h \in (0,1)$ that is independent of the state process X; $\alpha_Z > 0$ is the mean-revering parameter of the FOU process. Here, the non-clustering noise is introduced to represent all trading noises due to unequal information etc. that can not be explained by the discrete and clustering noises. If $h = \frac{1}{2}$, the driven process W^h reduces to the classical Brownian motion. Our non-clustering noise is more reasonable than that of Zeng (2003) in that: (1) We can preclude the possibility of negative price; (2) The long memory structure in our noise term can capture the empirical feature of momentum observed in transaction prices. Recall the momentum measures the rate of rise or fall in stock prices and it indicates the trending of the stock prices. (3) The mean-reverting structure of the FOU process captures the cycles property of prices. (see Black (1986)).

For the noise component Z_t^h, we can introduce its historical process $\widehat{Z}_t^h$ as in (2.110) and its martingale problem exists and computer workable approximations can be generated. This makes it possible for us to simulate the evolution of the noise term Z and it plays an important role in the particle filtering to come. To ease the notation, hereafter we fix i and write

$$t_i = T_i, x = X_{t_i}, y = Y_{t_i}, z = Z_{t_i}^h, \hat{z} = \widehat{Z}_{t_i}^h$$

in the following arguments whenever no confusion occurs. For tick price level $y_j = \frac{j}{M}, j = 1, 2, \cdots$, consider the interval $[y_j - \frac{1}{2M}, y_j + \frac{1}{2M})$. Suppose $\mathcal{Y}_{t_i}$ falls into the j^{th}–interval, then if there has no clustering (biasing) noise, the trading price Y_{t_i} is just y_j. Note that

$$\epsilon\zeta_i = \ln \mathcal{Y}_{t_i} - \ln X_{t_i} - Z_{t_i}^h.$$

Therefore, if

$$y_j - \frac{1}{2M} \leq \mathcal{Y}_{t_i} \leq y_j + \frac{1}{2M},$$

then it is equivalent to

$$\ln\left(y_j - \frac{1}{2M}\right) - \ln X_{t_i} - Z^h_{t_i} \le \epsilon\zeta_i \le \ln\left(y_j + \frac{1}{2M}\right) - \ln X_{t_i} - Z^h_{t_i},$$

or expressed through the historical process of Z^h as

$$\ln\left(y_j - \frac{1}{2M}\right) - \ln X_{t_i} - \pi_{t_i}\widehat{Z}^h_{t_i} \le \epsilon\zeta_i \le \ln\left(y_j + \frac{1}{2M}\right) - \ln X_{t_i} - \pi_{t_i}\widehat{Z}^h_{t_i}.$$

Recall here, for $t \in [0,T]$, π_t is the projection at t. So the probability of trading at y_j given X_{t_i}, $Z^h_{t_i}$ would be

$$R(y_j|X_{t_i},\theta,Z^h_{t_i},t_i) \triangleq \int_{\ln\left(\frac{y_j-\frac{1}{2M}}{X_{t_i}\cdot e^{Z^h_{t_i}}}\right)}^{\ln\left(\frac{y_j+\frac{1}{2M}}{X_{t_i}\cdot e^{Z^h_{t_i}}}\right)} \frac{1}{\sqrt{2\pi}\epsilon} e^{-\frac{u^2}{2\epsilon^2}}\,du, \tag{3.27}$$

or equivalently

$$R(y_j|X_{t_i},\theta,\widehat{Z}^h_{t_i},t_i) = \int_{\ln\left(\frac{y_j-\frac{1}{2M}}{X_{t_i}e^{\pi_{t_i}\widehat{Z}^h_{t_i}}}\right)}^{\ln\left(\frac{y_j+\frac{1}{2M}}{X_{t_i}e^{\pi_{t_i}\widehat{Z}^h_{t_i}}}\right)} \frac{1}{\sqrt{2\pi}\epsilon} e^{-\frac{u^2}{2\epsilon^2}}\,du. \tag{3.28}$$

It follows

$$\begin{aligned} R(y_j|x,\theta,z,t) &\triangleq \int_{\ln\left(\frac{y_j-\frac{1}{2M}}{x\cdot e^{z}}\right)}^{\ln\left(\frac{y_j+\frac{1}{2M}}{x\cdot e^{z}}\right)} \frac{1}{\sqrt{2\pi}\epsilon} e^{-\frac{u^2}{2\epsilon^2}}\,du, \\ R(y_j|x,\theta,\hat{z},t) &\triangleq \int_{\ln\left(\frac{y_j-\frac{1}{2M}}{x\cdot\pi_t e^{\hat{z}}}\right)}^{\ln\left(\frac{y_j+\frac{1}{2M}}{x\cdot e^{\pi_t\hat{z}}}\right)} \frac{1}{\sqrt{2\pi}\epsilon} e^{-\frac{u^2}{2\epsilon^2}}\,du \end{aligned} \tag{3.29}$$

Here, the noise term $\widehat{Z}$ can be incorporated into the state vector. The advantage of this formulation is that it is possible for us to estimate $\widehat{Z}$ thus Z jointly with other state components using the filtering method.

- Finally, incorporating the clustering noise to capture the clustering phenomenon.

Note that $M = 100$, it is convenient to introduce the following notations:

(3.30) $D_1 = \{\text{The integers in } (0, 100] \text{ that are not multiples of } 5\}$,
$D_2 = \{\text{The integers in } (0, 100] \text{ that are multiples of 5 but not of } 25\}$,
$D_3 = \{25, 75\}, \ \ D_4 = \{50\}, \ \ D_5 = \{100\}$.

If the fractional part of the price y is in D_1, then it will stay in the same level with probability $1 - \alpha$ and moves to the closest multiple of 5 cents, that is, the closest tick level in $D_2 \cup D_3 \cup D_4 \cup D_5$ with probability α; if the fractional part of the price y is in D_2, then it will stay in the same level with probability $1 - \beta$ and moves to the closest tick level in $D_3 \cup D_4 \cup D_5$ with probability β; if the fractional part of the price y is in D_3, then it will stay in the same level with probability $1 - \gamma_1 - \gamma_2$ and moves to the closest tick level in D_4 with probability γ_1, or the closest tick level in D_5, with probability γ_2. In summary, the transition probability function is obtained iteratively by

Case 1. *If the fractional part of y_j belong to D_1,*

$$p(y_j|x, \theta, \hat{z}, t) = R(y_j|x, \theta, \hat{z}, t)(1 - \alpha). \tag{3.31}$$

Case 2. *If the fractional part of y_j belong to D_2,*

$$p(y_j|x, \theta, \hat{z}, t) = R^*(y_j|x, \theta, \hat{z}, t)(1 - \beta), \tag{3.32}$$

where

(3.33)
$$R^*(y_j|x,\theta,\hat{z},t) \triangleq R(y_j|x,\theta,\hat{z},t) + \alpha\left(R(y_{j-1}|x,\theta,\hat{z},t) + R(y_{j-2}|x,\theta,\hat{z},t)\right) + \alpha\left(R(y_{j+1}|x,\theta,\hat{z},t) + R(y_{j+2}|x,\theta,\hat{z},t)\right)$$

Case 3. *If the fractional part of y_j belong to D_3,*

$$p(y_j|x, \theta, \hat{z}, t) = R^{**}(y_j|x, \theta, \hat{z}, t)(1 - \gamma_1 - \gamma_2), \tag{3.34}$$

where

(3.35)
$$R^{**}(y_j|x,\theta,\hat{z},t) \triangleq R^*(y_j|x,\theta,\hat{z},t) + \beta(R^*(y_{j-5}|x,\theta,\hat{z},t) + R^*(y_{j-10}|x,\theta,\hat{z},t))$$

$$+\beta(R^*(y_{j+5}|x,\theta,\hat{z},t)+R^*(y_{j+10}|x,\theta,\hat{z},t)).$$

Case 4. *If the fractional part of y_j belong to D_4,*

(3.36)
$$p(y_j|x,\theta,\hat{z},t)=R^{**}(y_j|x,\theta,\hat{z},t)+\gamma_1(R^{**}(y_{j-25}|x,\theta,\hat{z},t)+R^{**}(y_{j-25}|x,\theta,\hat{z},t)).$$

Case 5. *If the fractional part of y_j belong to D_5,*

(3.37)
$$p(y_j|x,\theta,\hat{z},t)=R^{**}(y_j|x,\theta,\hat{z},t)+\gamma_2(R^{**}(y_{j-25}|x,\theta,\hat{z},t)+R^{**}(y_{j+25}|x,\theta,\hat{z},t)).$$

Moreover,

Case 6. *For $j=0$,*

(3.38)
$$\begin{aligned}p(y_j|x,\theta,\hat{z},t)&=R(y_0|x,\theta,\hat{z},t)+\alpha(R(y_1|x,\theta,\hat{z},t)+R(y_2|x,\theta,\hat{z},t))\\&\quad+\beta(R^*(y_5|x,\theta,\hat{z},t)+R^*(y_{10}|x,\theta,\hat{z},t))+\gamma_2R^{**}(y_{25}|x,\theta,\hat{z},t).\end{aligned}$$

Our clustering setup is more reasonable than that of Zeng (2003) in that we capture the feature that the likelihood of clustering on one specific tick level is continuously changed.

Remark 3.6. *The parameters in the above microstructure model can be classified into two categories: the nonclustering noise parameter (α_Z, h, ϵ) and the clustering noise parameters $(\alpha,\beta,\gamma_1,\gamma_2)$.*

The empirical studies suggest that the tick size $\frac{1}{M}$ plays an important role in microstructure market analysis. For example, Huang, Stoll (2001) found the tick size is closely related to many key microstructure characteristics such as the price clustering, market depth, bid-ask spread, trading volume, etc. For these reasons, we recall: the NYSE converted all 3525 listed securities to decimal pricing in January 29, 2001. The NASDAQ began its decimal test phase with 14 securities in March 12, 2001, followed by another 197 securities in March 26, 2001. All remaining NASDAQ securities are converted to decimal trading in April 9, 2001. Such tick size adjustments (from $M=8$ or 16 to $M=100$), enable the bid-ask spread as small as one cent and provide a fairer market for the individual investors. Price clustering plays a central role in microstructure market and is necessary to examine this characteristic in both pre- and post-decimal pricing for comprehensive understanding.

Before the tick adjustments, the price clustering occurred between the eighthes. Our study shows NYSE stocks exhibit the clustering on even-eighth. This is consistent with the study of Huang, Stoll (2001), Chung, Van Ness (2004), Chung, Kim, Kitsabunnarat (2005). Moreover, the degree of price clustering in NYSE is relatively weaker than that of NASDAQ. Figure 3.3 shows the typical moderate clustering observed from the two-month tick data of General Electronic (GE), February-March, 1995. Its fraction of even eights is approximately 53%. This is consistent with Barclay (1997) where he examined 472 stocks from NASDAQ before and after their listing in NYSE or American Stock Exchange (AMEX): before the listing, the average fraction of even-eights is 78% while after listing, it drops to about 56%. To examine

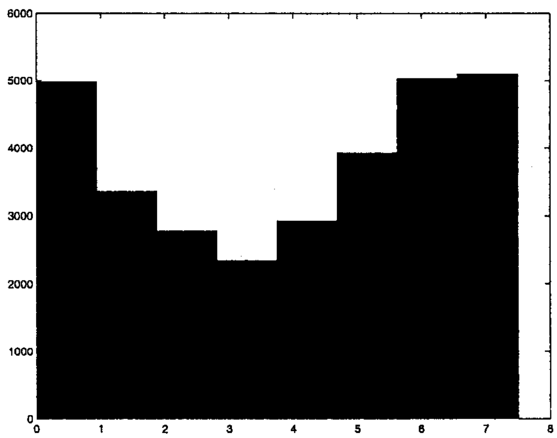


Figure 3.3: The moderate tick clustering of GE in Feb.-Mar. 1995 where $M = 8$.

the price clustering after the tick size adjustment ($M = 100$), we sample 16 NYSE listed stocks for one month, Mar. 2005, and examine their tick price behaviors with the associated microstructure characteristics. These 16 stocks are selected based on the criteria of cross-section, market capitalization and liquidity. They are tabulated with their symbols as follows.

NYSE Stock	Ticker Symbol	NYSE Stock	Ticker Symbol
American Express	AXP	Home Depot	HD
AT&T	T	Honeywell	HON
Boeing	BA	IBM	IBM
Citigroup	C	Pepsi	PEP
Exxon	XOM	Pfizer	PFE
General Motors	GM	Walmart	WMT
Goldman Sachs	GS	Walt Disney	DIS
Goodyear	GT	Xerox	XRX

Based on the transaction price data of these 16 listed stocks, we examine the clustering with the tick as 1 cent and our study shows there exist moderate clustering at the multiples of 5 cents as shown in Figure 3.4 plotting in terms of pennies and Figure 3.5 in terms of mode 25. Our clustering model (3.31) – (3.38) can well characterize this phenomenon. Based on the 16 listed stock prices, the clustering parameters can

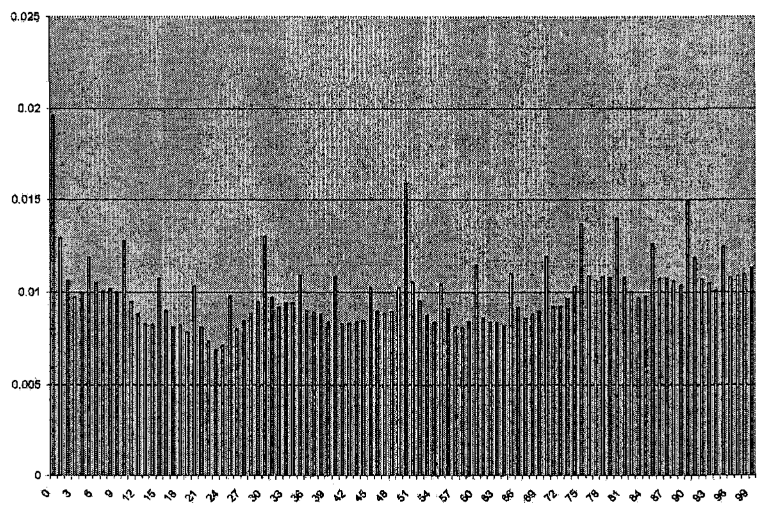


Figure 3.4: The Clustering on Pennies for 16 NYSE stocks on March 2005.

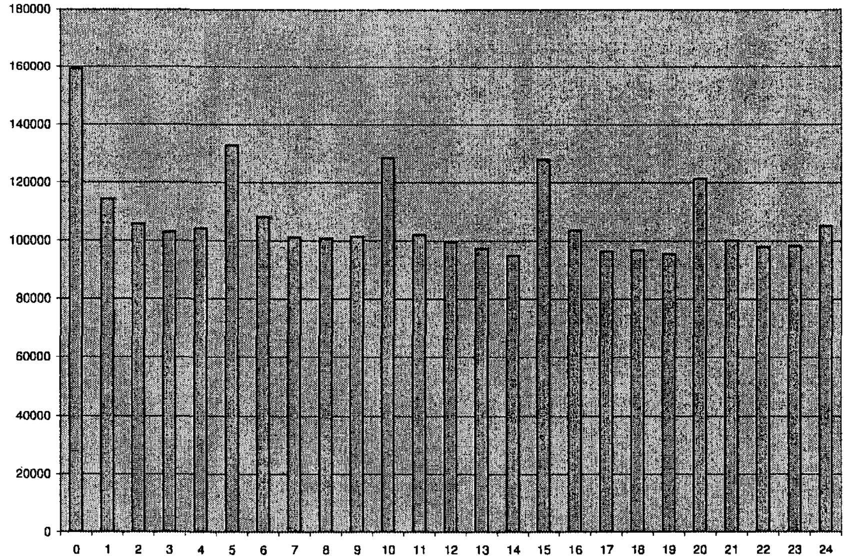


Figure 3.5: The Clustering in Nickels for 16 NYSE stocks on March 2005.

be estimated using the relative frequency analysis as follows.

Clustering Parameters	Estimate
α	0.060475
β	0.046883
γ_1	0.03883
γ_2	0.16525

3.3 Nonlinear Filtering and Bayes Estimation

Another approach of constructing the price process is to formulate it as a marked point process $\overrightarrow{Y}$: a double sequence of random variables $\overrightarrow{Y} = (t_i, Y_{t_i},\ i \geq 1)$ where $t_i \in [0, T]$ denotes the trading time of $i^{th}-$trade and Y_{t_i} the corresponding trading price. As discussed in Chapter 2, the mark space is $(E, \mathcal{E})$ where $E = \mathbb{N}_0$ and $\mathcal{E}$ is its discrete $\sigma-$algebra. Here, $j \in E$ corresponds to the $j^{th}-$tick level $\frac{j}{M}$. For each $A \in \mathcal{E}$, we can associate the counting process $Y_t(A)$

$$Y_t(A) \triangleq \sum_{i \geq 1} 1_{\{Y_{t_i} \in A\}} 1_{\{t_i \leq t\}} \tag{3.39}$$

to denote the accumulated trades that occurred in tick level set A. In particular, for $\forall j \in E$,

$$Y_j(t) \triangleq Y_t(\{j\}) = \sum_{i\geq 1} 1_{\{Y_{t_i}=j\}} 1_{\{t_i \leq t\}}$$

denotes the total trades at $j^{th}-$tick level $\frac{j}{M}$. As discussed in Chapter 2, we can also define the random counting measure $Y(dz \times dt)$ on $[0,T] \times \mathcal{E}$ by

$$Y(\omega, A \times [0,t]) \triangleq Y_t(\omega, A), \qquad \forall \omega \in \Omega, \quad A \in \mathcal{E}, \quad t \in [0,T].$$

Alternatively, we have

$$\lambda_z(X_t, \theta_t, t) = a(X_t, \theta_t, t) \cdot p(y_z | X_t, \theta_t, t), \tag{3.40}$$

where $X = (S, V)$ and $p(y_z|x, \theta, t)$ is the transition probability function from the state x to y_z at time t, conditional on the parameter θ. The equivalence of these two constructions can be established heuristically as follows:

$$\lim_{h\longrightarrow 0} \frac{\mathbb{P}\left(Y_z(t+h) - Y_z(t) > 0 | \mathcal{F}_t\right)}{h} = a(X_t, \theta, t) \cdot p(y_z|X_t, \theta_t, t) = \lambda_z(X_t, \theta_t, t).$$

Thus this structure of intensities makes these two constructions statistically equivalent. In the following, we focus on the nonlinear filtering approach and demonstrate how to derive the robust filter from it:

$$\pi_t(\cdot) = \mathbb{P}[(X_t, \theta_t) \in \cdot \, | \mathcal{F}_t^Y]$$

or equivalently,

$$\pi_t(f) = \mathbb{E}[f(X_t, \theta_t) | \mathcal{F}_t^Y]$$

for $f \in B(\mathbb{R}^{n_x + n_\theta})$. As discussed in Chapter 2, we have:

Theorem 3.1. *Under* A 2 − 3, *the unnormalized filter* σ_t *is the unique measure-valued solution of the stochastic filtering equation*

$$\sigma_t(f) = \sigma_0(f) + \int_0^t \sigma_s \left(\mathbf{A}f - (a(s) - \kappa) f\right) ds + \int_0^t \int_E \sigma_{s-} \left(\left(\frac{\lambda_z(s-)}{\kappa_z} - 1 \right) f \right) Y(dz, ds),$$

for $t > 0$ *and* $f \in \mathcal{D}(\mathbf{A})$.

To utilize the robust filter, we can verify the regularity conditions A 1, A 5, A 6 for Examples 3.1 − 3.6. A 1 is obvious because the domain $\mathcal{D}(\mathbf{A}^{(k)}),\ \ k = 1, \cdots, 6$ in these examples is just the set of all twice continuously-differentiable functions which are closed under multiplication. Recall here $\mathbf{A}^{(k)}$ is the pregenerator of the martingale problem for Examples 3.1 − 3.6. Moreover, the total intensity $a(S_t, V_t, t)$ is a smooth deterministic function of t. So, for $\forall z \in E$ and $t \in [0, T]$ fixed, the tick intensity function $\lambda_z \triangleq \lambda_z(\cdot, \cdot, t)$ is a twice continuously differentiable function of $x = (s, v)$. It follows $\lambda_z \in \mathcal{D}(\mathbf{A}^{(k)})$. Meanwhile, $\{M^{\lambda_z}\}_{t \geq 0}$ is continuous martingale in Example 3.1 − 3.6. The following theorem gives the evolution equation of the robust filter.

Theorem 3.2. *Assume* **A** 1−6 *hold true. Then* $\nu_t(f)$ *satisfies the evolution equation*

$$\nu_t(f) = \nu_0(f) + \int_0^t \nu_s(\widetilde{\mathbf{A}}_s^Y f) ds, \tag{3.41}$$

where $\mathcal{D}(\widetilde{\mathbf{A}}^Y) = \mathcal{D}(\mathbf{A})$ *and*

$$\begin{aligned} \widetilde{\mathbf{A}}_t^Y f = &\left(\mathbf{A}_t + \kappa - a(t) + \int_{E \times E} \frac{[\lambda_z, \lambda_\zeta]^t}{2\lambda_z \lambda_\zeta} Y(dz, t) Y(d\zeta, t) \right) f \\ &- \int_E \left(\frac{[\lambda_z, f]^t + 2f\mathbf{A}_t(\lambda_z)}{\lambda_z} - \frac{f\mathbf{A}_t(\lambda_z^2)}{2\lambda_z^2} \right) Y(dz, t). \end{aligned} \tag{3.42}$$

The above evolution equation does not involve any stochastic integration. In contrast, its randomness is just characterized through the parameterized observation path thus it is just the robust equation we are seeking and ν_t becomes the robust nonlinear filter. Theoretically, Theorem 3.2 solves the Bayes estimation problem in that it gives the evolution of the posterior distribution of (X, θ) conditional on the observation Y. However, in real applications, we need some efficient and recursive approximation to implement this evolution equation. To start, recall the path-dependent probability $\widehat{\mathbb{Q}}^Y$ introduced by (2.79):

$$\frac{d\widehat{\mathbb{Q}}^Y}{d\mathbb{Q}}\Big|_{\mathcal{F}_t} = \mathbf{D}_t^Y,$$

$$\mathbf{D}_t^Y \triangleq \exp\left(-\int_E \int_0^t Y_z(s-) dM_s^{\ln \lambda_z} m(dz) - \int_0^t \int_{E \times E} \frac{[\lambda_z, \lambda_\zeta]^s}{2\lambda_z \lambda_\zeta} Y(dz, s) Y(d\zeta, s) ds \right).$$

Proposition 3.1. *Suppose* X *satisfies the GDSV model (see* (3.15), (3.16)*) and* $\lambda_z = \lambda_z(s, v, \theta)$ *given by* (3.40)*, then under* $\widehat{\mathbb{Q}}^Y$*,* X *is governed by the path-dependent* *diffusion equation*

$$dS_t = b_1^Y(S_t, V_t, \theta)dt + c_1(S_t, V_t, \theta)dW_t, \tag{3.43}$$

$$dV_t = b_2^Y(S_t, V_t, \theta)dt + c_2(S_t, V_t, \theta)dB_t. \tag{3.44}$$

Here, W_t, B_t are independent Brownian motions and

$$b_1^Y(S_t, V_t, \theta) = b_1(S_t, V_t, \theta) - c_1^2(S_t, V_t, \theta) \cdot \int_E \left(\frac{1}{\lambda_z}\frac{\partial \lambda_z}{\partial s}\right)(S_t, V_t, \theta)Y(dz, t), \tag{3.45}$$

$$b_2^Y(S_t, V_t, \theta) = b_2(S_t, V_t, \theta) - c_2^2(S_t, V_t, \theta) \cdot \int_E \left(\frac{1}{\lambda_z}\frac{\partial \lambda_z}{\partial v}\right)(S_t, V_t, \theta)Y(dz, t). \tag{3.46}$$

Proof. It suffices to show that under $\widehat{\mathbb{Q}}^Y$, (S, V, θ) is a solution of the $\overline{\mathbf{A}}^Y$ martingale problem with generator

$$\begin{aligned}
\overline{\mathbf{A}}^Y f(s, v, \theta) = {} & \frac{1}{2}c_1^2(s, v, \theta)\frac{\partial^2 f}{\partial s^2}(s, v, \theta) + \frac{1}{2}c_2^2(s, v, \theta)\frac{\partial^2 f}{\partial v^2}(s, v, \theta) \\
& + \left(b_1(s, v, \theta) - c_1^2(s, v, \theta) \cdot \int_E \left(\frac{1}{\lambda_z}\frac{\partial \lambda_z}{\partial s}\right)(s, v, \theta)Y(dz, s)\right)\frac{\partial f}{\partial s}(s, v, \theta) \\
& + \left(b_2(s, v, \theta) - c_2^2(s, v, \theta) \cdot \int_E \left(\frac{1}{\lambda_z}\frac{\partial \lambda_z}{\partial s}\right)(s, v, \theta)Y(dz, s)\right)\frac{\partial f}{\partial v}(s, v, \theta).
\end{aligned}$$

From Lemma 2.3, under $\widehat{\mathbb{Q}}^Y$, (S, V, θ) is a solution of the $\widehat{\mathbf{A}}^Y$ martingale problem with

$$\widehat{\mathbf{A}}_t^Y f \triangleq \mathbf{A}_t f - \int_E [\ln \lambda_z, f]^t Y(dz, t).$$

Define

$$\mathbf{I} \triangleq \mathbf{A}(\ln \lambda_z \cdot f).$$

It follows

$$\begin{aligned}
\mathbf{I} &= b_1\frac{\partial(\ln \lambda_z \cdot f)}{\partial s} + b_2\frac{\partial(\ln \lambda_z \cdot f)}{\partial v} + \frac{1}{2}c_1^2\frac{\partial^2(\ln \lambda_z \cdot f)}{\partial s^2} + \frac{1}{2}c_2^2\frac{\partial^2(\ln \lambda_z \cdot f)}{\partial v^2} \\
&= b_1 \ln \lambda_z \cdot \frac{\partial f}{\partial s} + b_1 f \cdot \frac{1}{\lambda_z} \cdot \frac{\partial \lambda_z}{\partial s} + b_2 \ln \lambda_z \cdot \frac{\partial f}{\partial v} + b_2 f \cdot \frac{1}{\lambda_z} \cdot \frac{\partial \lambda_z}{\partial v} \\
&+ \frac{1}{2}c_1^2 \ln \lambda_z \cdot \frac{\partial^2 f}{\partial s^2} + \frac{1}{2}c_1^2\frac{\partial f}{\partial s} \cdot \frac{1}{\lambda_z} \cdot \frac{\partial \lambda_z}{\partial s} + \frac{1}{2}c_1^2\frac{\partial f}{\partial s} \cdot \frac{1}{\lambda_z} \cdot \frac{\partial \lambda_z}{\partial s} \\
&+ \frac{1}{2}c_1^2 f\frac{1}{\lambda_z} \cdot \frac{\partial^2 \lambda_z}{\partial x^2} - \frac{1}{2}c_1^2 f\frac{1}{\lambda_z^2} \cdot \left(\frac{\partial \lambda_z}{\partial s}\right)^2
\end{aligned}$$

$$
\begin{aligned}
&+ \frac{1}{2}c_2^2 \ln\lambda_z \cdot \frac{\partial^2 f}{\partial v^2} + \frac{1}{2}c_2^2 \frac{\partial f}{\partial v} \cdot \frac{1}{\lambda_z} \cdot \frac{\partial \lambda_z}{\partial v} + \frac{1}{2}c_2^2 \frac{\partial f}{\partial v} \cdot \frac{1}{\lambda_z} \cdot \frac{\partial \lambda_z}{\partial v} \\
&+ \frac{1}{2}c_2^2 v^2 f \frac{1}{\lambda_z} \cdot \frac{\partial^2 \lambda_z}{\partial v^2} - \frac{1}{2}c_2^2 f \frac{1}{\lambda_z^2} \cdot \left(\frac{\partial \lambda_z}{\partial v}\right)^2.
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
\mathbf{II} \triangleq \ln\lambda_z \cdot \mathbf{A}f &= \ln\lambda_z \cdot b_1 \frac{\partial f}{\partial s} + \ln\lambda_z \cdot \frac{1}{2}c_1^2 \frac{\partial^2 f}{\partial s^2} \\
&+ \ln\lambda_z \cdot b_2 \frac{\partial f}{\partial v} + \ln\lambda_z \cdot \frac{1}{2}c_2^2 \frac{\partial^2 f}{\partial v^2}.
\end{aligned}
$$

Note that

$$
\begin{aligned}
\mathbf{A}(\ln\lambda_z) &= b_1 \cdot \frac{1}{\lambda_z} \cdot \frac{\partial \lambda_z}{\partial s} + b_2 \cdot \frac{1}{\lambda_z} \cdot \frac{\partial \lambda_z}{\partial v} + \frac{1}{2}c_1^2 \frac{1}{\lambda_z} \cdot \frac{\partial^2 \lambda_z}{\partial s^2} - \frac{1}{2}c_1^2 \frac{1}{\lambda_z^2} \cdot \left(\frac{\partial \lambda_z}{\partial s}\right)^2 \\
&+ \frac{1}{2}c_2^2 \frac{1}{\lambda_z} \cdot \frac{\partial^2 \lambda_z}{\partial v^2} - \frac{1}{2}c_2^2 \frac{1}{\lambda_z^2} \cdot \left(\frac{\partial \lambda_z}{\partial v}\right)^2.
\end{aligned}
$$

It follows

$$
\begin{aligned}
&\mathbf{III} \triangleq f \cdot \mathbf{A}(\ln\lambda_z) = f \cdot b_1 \cdot \frac{1}{\lambda_z} \cdot \frac{\partial \lambda_z}{\partial s} + f \cdot b_2 \cdot \frac{1}{\lambda_z} \cdot \frac{\partial \lambda_z}{\partial v} + f \cdot \frac{1}{2}c_1^2 \frac{1}{\lambda_z} \cdot \frac{\partial^2 \lambda_z}{\partial s^2} \\
&- f \cdot \frac{1}{2}c_1^2 \frac{1}{\lambda_z^2} \cdot \left(\frac{\partial \lambda_z}{\partial s}\right)^2 + f \cdot \frac{1}{2}c_2^2 \frac{1}{\lambda_z} \cdot \frac{\partial^2 \lambda_z}{\partial v^2} - f \cdot \frac{1}{2}c_2^2 \frac{1}{\lambda_z^2} \cdot \left(\frac{\partial \lambda_z}{\partial v}\right)^2.
\end{aligned}
$$

Combining the above steps together, we get

$$
[\ln\lambda_z, f]^t = \mathbf{I} - \mathbf{II} - \mathbf{III} = c_1^2 \cdot \frac{\partial f}{\partial s} \cdot \left(\frac{1}{\lambda_z}\frac{\partial \lambda_z}{\partial s}\right) + c_2^2 \cdot \frac{\partial f}{\partial v} \cdot \left(\frac{1}{\lambda_z}\frac{\partial \lambda_z}{\partial v}\right).
$$

Therefore

$$
\begin{aligned}
\widehat{\mathbf{A}}^Y f &= \mathbf{A}f - \int_E [\ln\lambda_z, f] Y(dz, s) \\
&= \frac{1}{2}c_2^2 \frac{\partial^2 f}{\partial v^2} + \frac{1}{2}c_1^2 \frac{\partial^2 f}{\partial s^2} + \left(b_1 - c_1^2 \int_E \left(\frac{1}{\lambda_z}\frac{\partial \lambda_z}{\partial s}\right) Y(dz, s)\right) \frac{\partial f}{\partial s} \\
&+ \left(b_2 - c_2^2 \int_E \left(\frac{1}{\lambda_z}\frac{\partial \lambda_z}{\partial v}\right) Y(dz, s)\right) \frac{\partial f}{\partial v}.
\end{aligned}
$$

Thus

$$
\widehat{\mathbf{A}}^Y = \overline{\mathbf{A}}^Y.
$$

Hence the result. ∎

It will become clear that in this paper, we focus only on the case

$$\lambda_z = \lambda_z(s, \theta),$$

that is, the tick intensity does not depend on v, thus $\frac{\partial \lambda_z}{\partial v} = 0$ and

Corollary 3.1. *Suppose X satisfies GDSV and $\lambda_j = \lambda_j(s, \theta)$, then under $\widehat{\mathbb{Q}}^Y$, X is governed by the path-dependent equation*

$$dS_t = b_1^Y(S_t, V_t, \theta)dt + c_1(S_t, V_t, \theta)dW_t, \tag{3.47}$$

$$dV_t = b_2(S_t, V_t, \theta)dt + c_2(S_t, V_t, \theta)dB_t, \tag{3.48}$$

where

$$b_1^Y(S_t, V_t, \theta) = b_1(S_t, V_t, \theta) - c_1^2(S_t, V_t, \theta) \cdot \int_E \left(\frac{1}{\lambda_z}\frac{\partial \lambda_z}{\partial s}\right)(S_t, \theta)Y(dz, t). \tag{3.49}$$

Remark 3.7. *Note that the distribution of V does not change in this case.*

As to the JSV-GBM model, we have an analogous result

Proposition 3.2. *Suppose X satisfies the JSV-GBM model and $\lambda_z = \lambda_z(s, \theta)$, then under $\widehat{\mathbb{Q}}^Y$, X is governed by the path-dependent equation*

$$dS_t = \mu^Y(S_t, V_t, \theta)dt + S_t V_t dW_t, \tag{3.50}$$

$$dV_t = (J_{N_{t-}+1} - V_{t-})dN_t. \tag{3.51}$$

Here, the Brownian motion W_t, Poisson process N_t are independent and

$$\mu^Y(S_t, V_t, \theta) = \mu S_t - V_t^2 S_t^2 \cdot \int_E \left(\frac{1}{\lambda_z}\frac{\partial \lambda_z}{\partial s}\right)(S_t, \theta)Y(dz, t). \tag{3.52}$$

Proof. The proof is similar to that of Proposition 3.1 as we still have

$$[\ln \lambda_z, f]^t = \mathbf{I} - \mathbf{II} - \mathbf{III}$$

where

$$\mathbf{I} = \mathbf{A}^{(6)}(\ln \lambda_z \cdot f)$$
$$\mathbf{II} = (\ln \lambda_z) \cdot \mathbf{A}^{(6)} f$$
$$\mathbf{III} = f \cdot \mathbf{A}^{(6)}(\ln \lambda_z),$$

and the generator of Example 3.6 is

$$\begin{aligned}\mathbf{A}^{(6)} f(s,v,\theta) &= \frac{1}{2}v^2 s^2 \frac{\partial^2 f}{\partial s^2}(s,v,\theta) + \mu s \frac{\partial f}{\partial s}(s,v,\theta)\\ &+ \lambda \int_{\alpha_J}^{\beta_J} (f(s,z,\theta) - f(s,v,\theta)) \frac{1}{\beta_J - \alpha_J} dz.\end{aligned}$$

Note that $\lambda_z = \lambda_z(s,\theta)$ does not depend on v. Therefore, we have

$$\begin{aligned}\mathbf{I} &= \mathbf{A}^{(6)}(\ln \lambda_z \cdot f)\\ &= \frac{1}{2}v^2 s^2 \frac{\partial^2 (\ln \lambda_z \cdot f)}{\partial s^2}(s,v,\theta) + \mu s \frac{\partial (\ln \lambda_z \cdot f)}{\partial s}(s,v,\theta)\\ &+ \lambda \int_{\alpha_J}^{\beta_J} ((\ln \lambda_z \cdot f)(s,z,\theta) - (\ln \lambda_z \cdot f)(s,v,\theta)) \frac{1}{\beta_J - \alpha_J} dz\\ &= \frac{1}{2}v^2 s^2 \frac{\partial^2 (\ln \lambda_z \cdot f)}{\partial s^2}(s,v,\theta) + \mu s \frac{\partial (\ln \lambda_z \cdot f)}{\partial s}(s,v,\theta)\\ &+ \lambda \cdot \ln \lambda_z(s,\theta) \int_{\alpha_J}^{\beta_J} (f(s,z,\theta) - f(s,v,\theta)) \frac{1}{\beta_J - \alpha_J} dz.\end{aligned}$$

$$\begin{aligned}\mathbf{II} &= (\ln \lambda_z) \cdot \mathbf{A}^{(6)} f\\ &= \ln \lambda_z \cdot \frac{1}{2}v^2 s^2 \frac{\partial^2 f}{\partial s^2}(s,v,\theta) + \ln \lambda_z \cdot \mu s \frac{\partial f}{\partial s}(s,v,\theta)\\ &+ \ln \lambda_z \cdot \lambda \int_{\alpha_J}^{\beta_J} (f(s,z,\theta) - f(s,v,\theta)) \frac{1}{\beta_J - \alpha_J} dz\end{aligned}$$

$$\begin{aligned}\mathbf{III} &= f \cdot \mathbf{A}^{(6)}(\ln \lambda_z)\\ &= f \cdot \frac{1}{2}v^2 s^2 \frac{\partial^2 \ln \lambda_z}{\partial s^2}(s,v,\theta) + f \cdot \mu s \frac{\partial \ln \lambda_z}{\partial s}(s,v,\theta).\end{aligned}$$

Combining above terms together, it follows that the integration terms of pure jumping component are canceled. Thus, we still have

$$[\ln \lambda_z, f]^t = \left(v^2 s^2 \frac{\partial f}{\partial s}\right) \cdot \left(\frac{1}{\lambda_z}\frac{\partial \lambda_z}{\partial s}\right).$$

Hence the result. □

Our new particle filter that can be thought of as a generalization of Del Moral, Noyer and Salut (1995). Note that the state process $X_t = (S_t, V_t)$ and the parameters to be estimated consist of two components: the state parameter $\theta \in \mathbb{R}^{n_\theta}$ and microstructure noise parameter $(\alpha, \beta, \gamma_1, \gamma_2; \alpha_Z, h, \epsilon)$. In this paper, the clustering noise parameter $(\alpha, \beta, \gamma_1, \gamma_2)$ are estimated empirically by the relative frequency analysis as discussed before while the magnitude parameter α_Z, h, ϵ can be estimated jointly with (X, θ) through the particle filtering. For sake of convenience, introduce $\vartheta \triangleq (\theta; \alpha_Z, h, \epsilon) \in \mathbb{R}^{n_\theta+3}$ as the augmented parameter vector that will be estimated through the particle filtering.

Consider the simple N-equalized time partitions

$$\{\tau_0 = 0, \tau_1 = \frac{T}{N}, \ldots \tau_i = \frac{iT}{N}, \cdots \tau_N = T\} \tag{3.53}$$

to make the particle approximation converge regardless the trading times $\{t_1, t_2, \cdots\}$.

Now, let $\{\varphi_N(t);\ t \geq 0\}$ be a sequence of measure-valued processes which stands for empirical distributions of the particle system to be constructed. In general, the particle filtering algorithm can be divided into three consecutive phases: *initialization, evolution (prediction) and re-sampling (updating).*

Initialization

- At $\tau_0 = 0$, we draw m_N independent particles $\{P^k\}_{k=1}^{m_N}$ with the equal weights $\frac{1}{m_N}$ from the joint prior distribution $\pi_0(\cdot)$ of $(S_0, V_0; \vartheta) \in \mathbb{R}^{n_\theta+3}$. The states of these particles at time t are denoted as $\{P_t^k\}_{k=1}^{m_N}$. Here, m_N is some positive integer that satisfies

$$\lim_{N \longrightarrow \infty} m_N = \infty.$$

The empirical (occupation) measure of this particle system at $\tau_0 = 0$ is

$$\varphi_N(0) \triangleq \frac{1}{m_N} \sum_{k=1}^{m_N} \delta_{P_0^k}(\cdot), \tag{3.54}$$

which satisfies

$$\lim_{N \longrightarrow \infty} (\varphi_N(0), f) = \pi_0(f) \quad \forall f \in B(\mathbb{R}^{n_\theta+3}).$$

Here, $\delta_x(\cdot)$ is the Dirac measure at x.

Remark 3.8. *Note that* $L_0 = 1$ *and* $Y_j(0) = 0$ *for all* $j = 0, 1, \cdots$ *thus*

$$\pi_0(f) = \sigma_0(f) = \nu_0(f) \quad \forall f \in B(\mathbb{R}^{n_\theta+3}).$$

Remark 3.9. *When there is no special information, it is convenient to assign uniform distributions to* $(S_0, V_0; \vartheta)$ *where the range* $[\alpha_{S_0}, \beta_{S_0}], [\alpha_{V_0}, \beta_{V_0}] \subset \mathbb{R}$ *and* $[\alpha_\vartheta, \beta_\vartheta] \subset \mathbb{R}^{n_\theta+3}$ *are determined from empirical study.*

Evolution

- During the interval $[\tau_{i-1}, \tau_i)$, $i = 1, 2, \cdots, N$, all particles move independently according to the same law of $(S, V; \vartheta)$. Considering ϑ is time-invariant, so it suffices to consider only the dynamics of (S, V) during these intervals. Suppose X satisfied the GDSV model, then from Proposition 3.1, the particles $\{P_t^k = (S_t^k, V_t^k, \theta^k)\}_{k=1}^{m_N}$ evolve according to the following diffusion

$$dS_t^k = b_1^Y(S_t^k, V_t^k, \theta^k)dt + c_1(S_t^k, V_t^k, \theta^k)dW_t^k, \tag{3.55}$$

$$dV_t^k = b_2^Y(S_t^k, V_t^k, \theta^k)dt + c_2(S_t^k, V_t^k, \theta^k)dB_t^k \tag{3.56}$$

for $k = 1, 2, \cdots, m_N$. Here, $\{W^k\}_{k=1}^{m_N}, \{B^k\}_{k=1}^{m_N}$ are independent Brownian motions. Instead, if X satisfied the JSV-GBM model, then from Proposition 3.2, the particles evolve according to the following diffusion

$$dS_t^k = \mu^Y(S_t^k, V_t^k, \theta^k)dt + S_t^k V_t^k dW_t^k, \tag{3.57}$$

$$dV_t^k = (J_{N_{t-}^k+1} - V_{t-}^k)dN_t^k. \tag{3.58}$$

for $k = 1, 2, \cdots, m_N$. Here, the Brownian motions $\{W^k\}_{k=1}^{m_N}$ and Poisson processes $\{N^k\}_{k=1}^{m_N}$ are independent and

$$\mu^Y(s, v, \theta) = \mu s - v^2 s^2 \cdot \int_E \left(\frac{1}{\lambda_z} \frac{\partial \lambda_z}{\partial s} \right) (s, \theta) Y(dz, t).$$

Testing Particle Weight

- For $k = 1, 2, \cdots, m_N$, each particle $P^k_{\tau_i}$ is given a weight $\omega^k_i(\tau_i)$ at the ending point τ_i based on the likelihood of the observation depending on its trajectory realized on $[\tau_{i-1}, \tau_i)$. Note that our filter is robust filter, so we can just make the simple time partitions regardless the trading times. Therefore, from the representation (2.80) of ν_t, we have

$$\begin{aligned} \omega^k_i(\tau_i) \quad &\triangleq \quad \exp(\int_{\tau_{i-1}}^{\tau_i} \{- \left[\int_E \mathbf{A}(\ln \lambda_z)(S^k_s, V^k_s, \theta^k_s, s) Y(dz, s) + a(s) - \kappa \right] \\ &+ \quad \int_{E \times E} \frac{[\lambda_z, \lambda_\zeta]^s}{2 \lambda_z \lambda_\zeta} (S^k_s, V^k_s, \theta^k_s, s) \} ds). \end{aligned}$$

- If there is no trade occurring within $[\tau_{i-1}, \tau_i)$, then Y_i, Y_j stay unchanged. The trajectory of each particle is determined by the Proposition 3.1 or 3.2. Suppose X satisfied the GDSV model, then from the Proposition 3.1 and Euler scheme, for $k = 1, 2, \cdots, m_N$, a discrete version of the diffusion processes (S^k_t, V^k_t) is

(3.59)

$$\begin{aligned} \Delta S^k_{\tau_i} = &\left[b_1(S^k_{\tau_{i-1}}, V^k_{\tau_{i-1}}) - c_1^2(S^k_{\tau_{i-1}}, V^k_{\tau_{i-1}}) \cdot \int_E \left(\frac{1}{\lambda_z} \frac{\partial \lambda_z}{\partial s} \right) Y(dz, \tau_{i-1}) \right] \cdot \Delta \tau_i \\ &+ c_1(S^k_{\tau_{i-1}}, V^k_{\tau_{i-1}}, \theta^k) \Delta W^k_{\tau_i}, \end{aligned}$$

(3.60)

$$\begin{aligned} \Delta V^k_{\tau_i} = &\left[b_2(S^k_{\tau_{i-1}}, V^k_{\tau_{i-1}}) - c_2^2(S^k_{\tau_{i-1}}, V^k_{\tau_{i-1}}) \cdot \int_E \left(\frac{1}{\lambda_z} \frac{\partial \lambda_z}{\partial s} \right) Y(dz, \tau_{i-1}) \right] \cdot \Delta \tau_i \\ &+ c_2(S^k_{\tau_{i-1}}, V^k_{\tau_{i-1}}, \theta^k) \Delta B^k_{\tau_i} \end{aligned}$$

where $\Delta W^k_{\tau_i} \sim N(0,\ \Delta \tau_i)$, $\ \Delta B^k_{\tau_i} \sim N(0,\ \Delta \tau_i)$ for $i = 1, 2, \cdots, N$.

- Instead, if X satisfies the JSV-GBM model, then from Proposition 3.2,

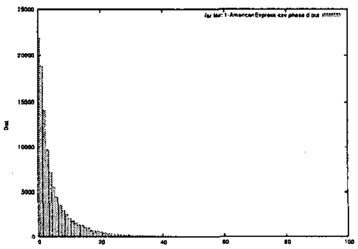
$$\Delta S^k_{\tau_i} = \left[\mu S^k_{\tau_{i-1}} - (S^k_{\tau_{i-1}} V^k_{\tau_{i-1}})^2 \cdot \int_E \left(\frac{1}{\lambda_z}\frac{\partial \lambda_z}{\partial s}\right) Y(dz, \tau_{i-1})\right] \cdot \Delta\tau_i + c_1(S^k_{\tau_{i-1}}, \theta^k)\Delta W^k_{\tau_i}, \quad V^k_{\tau_i} = J_{N^k_{\tau_i}} \tag{3.61}$$

  where $\Delta W^k_{\tau_i} \sim N(0,\, \Delta\tau_i)$; $N^k_0 = 0$, $N^k_{\tau_i} - N^k_{\tau_{i-1}}$ is a Poisson random variable with intensity $\lambda(\Delta\tau_i)$; and each J_n is uniformly distributed on $[\alpha_J, \beta_J]$ for $i = 1, 2, \cdots, N$.

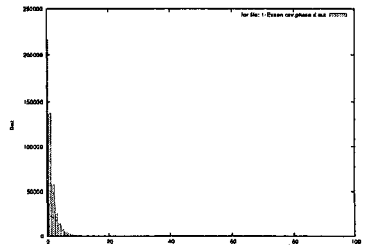
- For the details of the re-sampling step, we refer Chapter 2.

3.3.1 Numerical Results

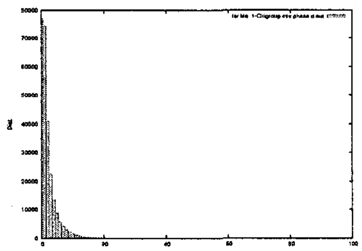
The particle filtering algorithm is extensively implemented on real data to calibrate the SV models $1-6$ of Section 3.2. The data is still the one-month (March, 2005, 23 business days) transaction price of the listed NYSE stocks. Here, to simplify the analysis, the total intensity function $a(x, \theta, t)$ is confined to be the deterministic case $a(x, \theta, t) = a(t)$ where the intensity function $a(t)$ is estimated through the inter-trade durations of the tick data. Figure 3.6 plots histogram of the inter-trade durations of 8 stocks from the listed NYSE stocks. As discussed before, the clustering noise parameters are estimated through the relative frequency method. The relative frequency method is a variant of the method of moments thus it follows that the relative frequency estimates are consistent and asymptotically normal, as discussed in Zeng (2003). The state parameters θ and the remaining microstructure noise parameters are estimated via the particle filtering algorithm and the final Bayes



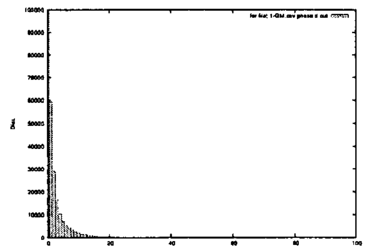
(a) American Express



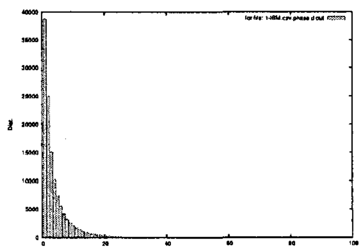
(b) Exxon



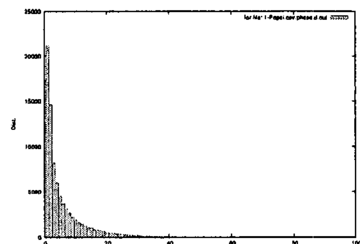
(c) Citigroup



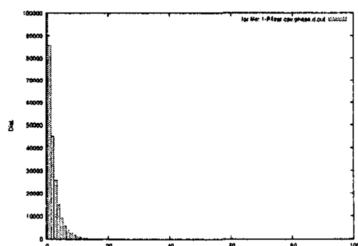
(d) GM



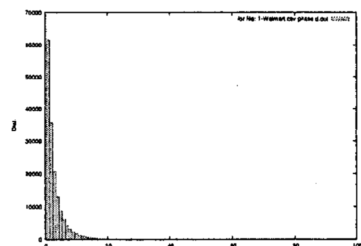
(e) IBM



(f) Pepsi



(g) Pfizer



(h) Walmart

Figure 3.6: The Histogram of Intertrade Times of Some NYSE Stocks, March 2005.

estimates (in seconds) are reported in the following tables.

PEP	HW	LOU	Nelson	Heston	JSV
μ	$-2.23E-8$	$-1.86E-8$	$-2.69E-8$	$-2.21E-8$	$-3.77E-8$
σ	–	–	–	–	–
ν	$-5.55E-8$	$4.89E-4$	$6.04E-9$	$5.50E-10$	–
ϱ	–	$3.39E-8$	$4.56E-8$	$4.90E-8$	–
ϖ	–	$2.48E-4$	–	–	–
κ	$2.99E-3$	$5.05E-3$	$5.41E-3$	$8.91E-8$	–
λ	–	–	–	–	$4.68E-6$
α_J	–	–	–	–	$1.40E-5$
β_J	–	–	–	–	$6.18E-5$

XRX	HW	LOU	Nelson	Heston	JSV
μ	$-2.86E-8$	$-3.40E-8$	$-3.80E-8$	$-3.35E-8$	$-3.73E-8$
σ	–	–	–	–	–
ν	$-9.45E-8$	$1.80E-4$	$5.12E-10$	$4.13E-11$	–
ϱ	–	$3.89E-8$	$1.78E-7$	$1.01E-7$	–
ϖ	–	$1.25E-4$	–	–	–
κ	$3.72E-4$	$1.34E-4$	$3.68E-4$	$6.61E-8$	–
λ	–	–	–	–	$3.62E-6$
α_J	–	–	–	–	$2.10E-5$
β_J	–	–	–	–	$1.09E-4$

The parameter estimations of GBM model are respectively: PEP, $(\mu, \sigma) = (-3.37E-8, 3.47E-5)$; XRX, $(\mu, \sigma) = (-1.57E-8, 6.00E-5)$. The data we used here are respectively the market transaction prices of Pepsi and Xorex (March, 2005). The estimation of the intrinsic value, volatility based on the market prices of Pepsi, Xerox are reported respectively in Figure $3.7-3.10$. The noise estimations are given in Figure $3.11-3.12$. It can be seen that these five SV models can track the volatility changes in the transaction prices and give some real-time updated to our posterior estimations of the value and volatility processes. However, their tracking speeds are different. Thus, it is necessary to compare their behaviors within a unified framework and this will be done in the following section using Bayes model selection.

Figure 3.7: The Value Estimation for SV models of Pepsi, March. 2005.

Please purchase 'docPrint PDF Driver' on http://www.verypdf.com/artprint/index.html to remove this message.

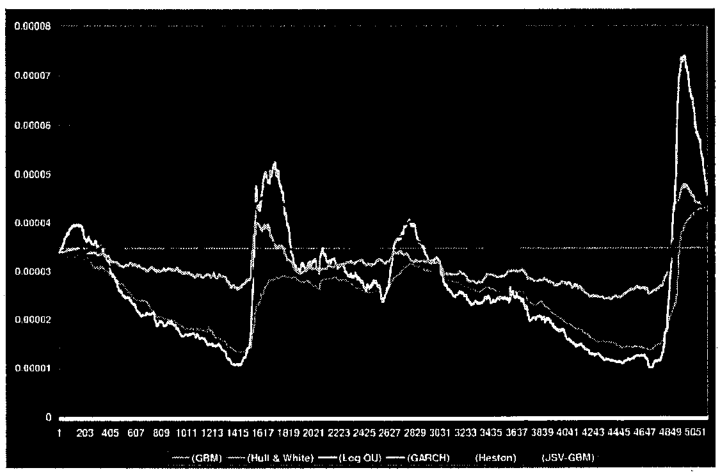


Figure 3.8: The Volatility Estimation for SV models of Pepsi, March. 2005.

Figure 3.9: The Value Estimation for SV models of Xerox, March. 2005.

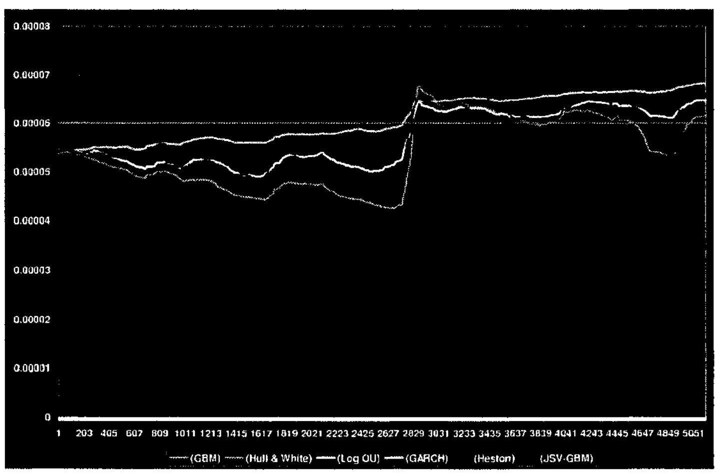


Figure 3.10: The Volatility Estimation for SV models of Xerox, March 2005.

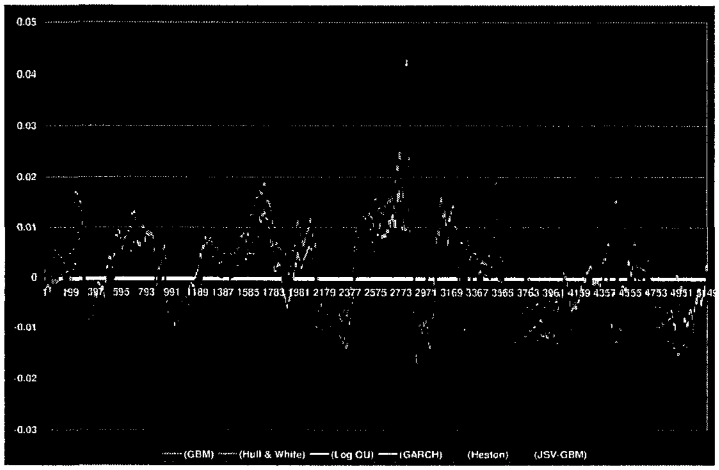


Figure 3.11: The Noise Process Estimation for SV models of Pepsi, March 2005.

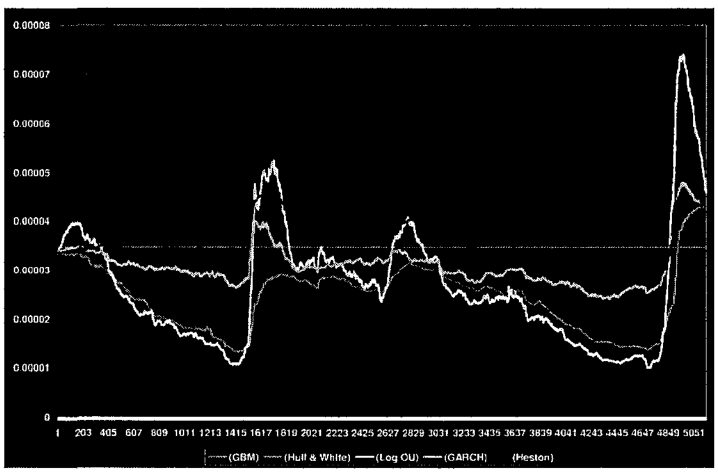


Figure 3.12: The Noise Process Estimation for SV models of Xerox, March 2005.

3.4 Bayes Model Selection of Stochastic Volatility Models

3.4.1 Bayes Factor and Model Selection

The available information in microstructure market is the observation process Y which represents the cumulative transaction records throughout all tick price levels. The Bayes factor determines which class of models best fits such observed datum by doing pairwise comparisons. Our methods is based on the robust filter derived in Section 3.3. The underlying robust filtering equation has only one time scale "dt" so only common calculus formulas are required to derive the dynamics of the robust Bayes factor. Suppose there are two microstructure models

$$M^{(k)} \triangleq (X^{(k)}, \theta^{(k)}, Y^{(k)})$$

with total intensity function $a^{(k)} = a^{(k)}$, tick intensity $\{\lambda_z^{(k)}\}_{z=0}^{\infty}$ and generators $\mathbf{A}^{(k)}$ for $k = 1, 2$ respectively. The filter ratio processes are defined in Chapter 2 and the following result holds true:

Theorem 3.3. *For $i = 1, 2$, $(\widetilde{q}_1, \widetilde{q}_2)$ is the unique solution to*

$$\begin{aligned}
d\widetilde{q}_i(f_i, t) &= \widetilde{q}_i\left(\widehat{\mathbf{A}}_t^{Y,\,(i)} f_i - a^{(i)} f_i,\ t\right) dt \\
&+ \widetilde{q}_i\left(f_i\left(-\int_E \mathbf{A}_t^{(i)} \ln \lambda_z^{(i)} Y(dz) + \int_{E\times E} \frac{[\lambda_z, \lambda_\zeta]^{(i),\,t}}{2\lambda_z^{(i)}\lambda_\zeta^{(i)}} Y(dz) Y(d\zeta)\right)\right) dt \\
&- \frac{\widetilde{q}_i(f_i)}{\widetilde{q}_{3-i}(g^{(3-i)})} \cdot \widetilde{q}_{3-i}\left(\widehat{\mathbf{A}}_t^{Y,\,(3-i)} g^{(3-i)} - a^{(3-i)} g^{(3-i)}\right) dt - \frac{\widetilde{q}_{3-i}(f_i, t)}{\widetilde{q}_{3-i}(g^{(3-i)}, t)} \\
&\cdot \widetilde{q}_{3-i}\left(g^{(3-i)}\left(-\int_E \mathbf{A}_t^{(3-i)} \ln \lambda_z^{(3-i)} Y(dz) + \int_{E\times E} \frac{[\lambda_z, \lambda_\zeta]^{(3-i),\,t}}{2\lambda_z^{(3-i)}\lambda_\zeta^{(3-i)}} Y(dz, d\zeta)\right)\right) dt.
\end{aligned}$$

Now we consider the problem of selecting the best of our 6 microstructure volatility models:

$$M^{(k)} \triangleq (X^{(k)}, \theta^{(k)}, \vec{Y}^{(k)})$$

with the total intensity $a^{(k)}$, tick intensity $\{\lambda_z^{(k)}\}_{z=0}^{\infty}$ and generators $\mathbf{A}^{(k)}$ for $k = 1, 2, 3, 4, 5, 6$ respectively specified as in Section 3.2. We compare these 6 models in pairwise to determine which one can best represent the market data. The above equations do not have a closed form solution. Instead, we describe the particle filter system which approximates the pair of measure-valued processes $(\widetilde{q}_{k_1}, \widetilde{q}_{k_2})$, where $\widetilde{q}_{k_1}, \widetilde{q}_{k_2}$ are respectively the Bayes factors of model $M^{(k_1)}, M^{(k_2)}$. The particle filter system for Bayesian model selection is similar to that of Bayesian estimation discussed in Section 3.3. Following the same lines, we consider the following N equally-spaced partitions on $[0, T]$:

$$\{\tau_0 = 0, \tau_1 = \frac{T}{N}, \dots \tau_i = \frac{iT}{N}, \cdots \tau_N = T\} \tag{3.62}$$

and let $\{(\psi_N^{(k_1)}(t), \psi_N^{(k_2)}(t)); t \geq 0\}$ be the sequence of empirical measure process of particle system which converges to $(\widetilde{q}_{k_1}, \widetilde{q}_{k_2})$ weakly as the number of particles converges to infinity. Here, for $k \in \{1, 2, 3, 4, 5, 6\}$, note that $L_0^{(k)} = 1$ and $Y_j^{(k)}(0) = 1$ for all $j = 0, 1, 2, \cdots$, then we have

$$(\pi_0^{(k)}, f) = (\sigma_0^{(k)}, f) = (\widetilde{q}_k, f)$$

for $f \in B(\mathbb{R}^{n_\theta+4})$.

Evolution

- During the interval $[\tau_{i-1}, \tau_i)$, $i = 1, 2, \cdots, N$, the particles $\{P_t^{(k_1)}, P_t^{(k_2)}\}_{k=1}^{m_N}$ move according to the law of model $M^{(k_1)}, M^{(k_2)}$ given by Proposition 3.1 or 3.2. Intuitively, the particle system explores the state space following the law of the state process.

- For $i = 1, 2, \cdots, N$, the particles $\{P_t^{(k_1)}\}_{k=1}^{m_N}$ are respectively given a weight $\{\omega_i^{(k_1), k}\}_{k=1}^{m_N}$ at time τ_i based on its trajectory realized on $[\tau_{i-1}, \tau_i)$:

$$\omega_i^{(k_1), k} = \Upsilon_i^{(k_1)}(\tau_i) \cdot \Lambda_i^{(k_1), k}(\tau_i) \cdot \exp(\int_{\tau_{i-1}}^{\tau_i} \{- \left[\int_E \mathbf{A}(\ln \lambda_z^{(k_1)}) Y(dz, s)\right]$$
$$- a(X_s^{(k_1), k}, \theta_s^{(k_1), k}, s) + \int_{E \times E} \frac{[\lambda_z^{(k_1)}, \lambda_\zeta^{(k_1)}]}{2\lambda_z^{(k_1)} \lambda_\zeta^{(k_1)}} Y(dz) Y(d\zeta)\} ds),$$

where

(3.63)

$$\Upsilon_i^{(k_1)}(t) = \exp\left(-\int_{\tau_{i-1}}^{t} \frac{\sum_{k=1}^{m_N} \left(\widehat{\mathbf{A}}^{y,\,(k_2)} g^{(k_2)} - a^{(k_2)} g^{(k_2)}\right)\left(X_s^{(k_2),\,k}\right)}{\sum_{k=1}^{m_N} g^{(k_2)}\left(X_s^{(k_2),\,k}\right)} ds \right),$$

$$\Lambda_i^{(k_1)}(t) = \left(-\int_{\tau_{i-1}}^{t} \frac{\sum_{k=1}^{m_N} \left(g^{(k_2)}\Gamma^{(k_2)}\right)\left(X_s^{(k_2),\,k}, \theta_s^{(k_2),\,k}, s\right)}{\sum_{k=1}^{m_N} g^{(k_2)}\left(X_s^{(k_2),\,k}, \theta_s^{(k_2),\,k}, s\right)} ds \right).$$

Here,

$$\Gamma^{(k_1)} \triangleq -\int_E \mathbf{A}^{(k_2)} \ln \lambda_z^{(k_2)} Y(dz) + \int_{E\times E} \frac{[\lambda_z^{(k_2)}, \lambda_\zeta^{(k_2)}]}{2\lambda_z^{(k_2)}\lambda_\zeta^{(k_2)}} Y(dz)Y(d\zeta).$$

- For $i = 1, 2, \cdots, N$, the particles $\{P_t^{(k_2)}\}_{k=1}^{m_N}$ are given a weight $\{\omega_i^{(k_2),k}\}_{k=1}^{m_N}$ similarly. The remaining procedures of particle filtering are just an analog of that given in Section 3.3 and we make the re-sampling step to successively incorporate the new information into the Bayes factor.

3.4.2 Numerical Results

The particle filtering algorithm proposed is extensively tested on real data to study the model selection between the microstructure models. The candidate models include the GDSV model (Example 3.1 – 3.5) and JSV-GBM model (Example 3.6). The data is the transaction price of Pepsi and Xerox, March, 2005. We use the GBM model as the calibration model for our model selection. The evolution of the Bayes factor are reported by Figure 3.13 – 3.14 and according to the interpretation table given in Section 2.4, we know the Bayes factor enables us to capture the volatility movement and prefer the Heston's model based on these two stock prices.

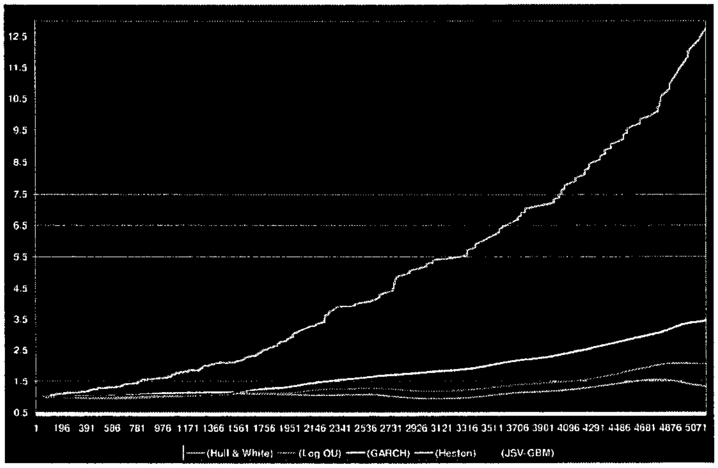


Figure 3.13: The Bayes Factor Estimation for SV models of Pepsi, March 2005.

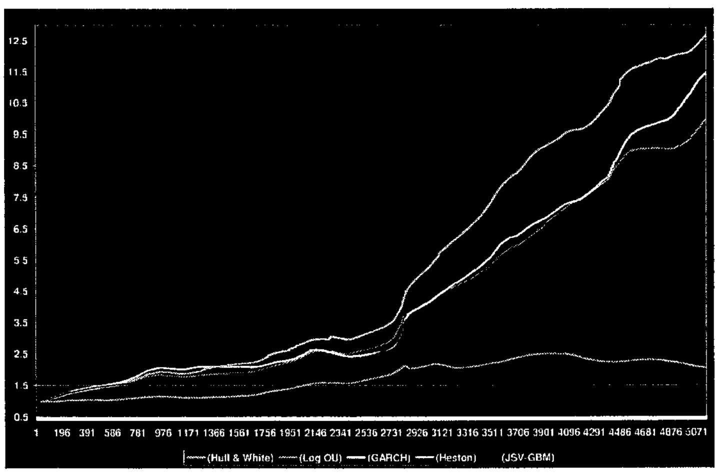


Figure 3.14: The Bayes Factor Estimation for SV models of Xerox, March 2005.

Bibliography

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

[2] Aït-Sahalia, Y., Kimmel, R. (2007). Maximum Likelihood Estimation of Stochastic Volatility Models. *Journal of Financial Economics* **83**, 413-452.

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

[4] Barclay, M. (1997). Bid-ask Spreads and the Avoidance of Odd-eighth Quotes on Nasdaq: An Examination of Exchange Listings. *Journal of Financial Economics* **45**, 35-60.

[5] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., Shephard, N. (2007). Subsampling Realised Kernels. Working paper.

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

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

[8] Bremaud, P. (1981). *Point Process and Queues: Martingale Dynamics.* Springer-Verlag, New York.

[9] Chan, L. K. C., Lakonishok, J. (1993). Institutes Trades and Intraday Stock Prices Behavior. *Journal of Financial Econometrics* **33**, 173-199.

[10] Chung, K. H., Van Ness, B., Van Ness, R. (2004). Trading Costs and Quotes Clustering on the NYSE and NASDAQ after Decimalization. *Journal of Financial Research* **27**, 309-328.

[11] Chung, K. L. (1982). *Lectures from Markov Process to Brownian Motion.* Springer-Verlag, New York.

[12] Chung, K. H., Kim, K. A., Kitsabunnarat. (2005). Liquidity and Quote Clustering in a Market with Multiple Tick Sizes. *Journal of Financial Research* **28**, 177-195.

[13] Clark, J. M. C. (1978). The Design of Robust Approximations to the Stocchastic Differential Equations of Nonlinear Filtering. *Communication systems and random process theory* (Proc. 2nd NATO Advanced Study Inst., Darlinton, 1977), 721-734. NATO Advanced Study Inst. Ser., Ser. E: Appl. Sci., No. 25, Sijthoff & Noordhoff, Alphen aan den Rijn.

[14] Davis, M. H. A. (1980). On a Multiplicative Functional Transformation Arising in Nonlinear Filtering Theory. *Z. Wahrsch. Verw. Gebiete.* **54**, 125-139.

[15] Davis, M. H. A. (1981). Factorization of a Multiplicative Functional of Nonlinear Filtering Theory. *System & Control Letters* **25(2)**, 260-278.

[16] Dawson, D., Perkins, E. (1991). Historical Processes. *Memoirs of the Amer. Math. Soc.* **93**, No. 454.

[17] Dellacherie, C., Meyer, P. A. (1978). *Probabilities and Potentials.* North-Holland, Amsterdam.

[18] Del Moral, P., Jacod, J., Protter, Ph. (2001). The Monte-Carlo Method for Filtering with Discrete-time Observations. *Probability Theory and Realted Fields* **120**, 346-368.

[19] Del Moral, P., Noyer, J. C., Salut, G. (1995). Maslov Optimization Theory: Stochastic Interpretation, Particle Resolution. *Lecture Note in Control and Information Sciences* **199**, Springer-Verlag, Berlin.

[20] Desai, R., Viens, F., Lele, T. (2003). A Monte-Carlo Method for Portfolio Optimization under Partially Observed Stochastic Volatility. *IEEE International Conference on Computational Intelligence for Financial Engineering 2003. Proceedings*, 257 - 263.

[21] Drost, F. C., Werker, J. M. (1996). Closing the GARCH Gap: Continuous Time GARCH Modeling. *Journal of Econometrics* **74**, 31-57.

[22] Duan, J. C., Fulop, A. (2007). Estimating the Structural Credit Risk Model When Equity Prices are Contaminated by Trading Noises. *Journal of Econometrics* Forthcoming.

[23] Duffie, D., Lando, D. (2001). Term Structures of Credit Spreads with Incomplete Accounting Information. *Econometrica* **69**, 633-664.

[24] Dufour, A., Engle, R. (2000). Time and the Price Impact of a Trade. *Journal of Finance* **55**, 2467-2498.

[25] Duncan, T. E., Hu, Y. Z., Pasik-Duncan, B. (2000). Stochastic Calculus for Fractional Brownian Motion. *SIAM J. Control Optim.* **38**, 582-612.

[26] Dynkin, E. B. (1991). Path Processes and Historical Superprocesses. *Probab. Theory Rel. Fields* **90**, 1-36.

[27] Elliot, R. J., Malcolm, W. P., Tsoi, A. H. (2003). Robust Parameter Estimation for Asset Price Models with Markov Modulated Volatilities. *Journal of Econ. Dyna. Control* **27**, 1391-1409.

[28] Engle, R. (2000). The Econometrics of Ultra-high-frequency Data. *Econometrica* **68**, 1-22.

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

[30] Ethier, S. N., Kurtz, T. G. (1986). *Markov Processes: Characterization and Convergence.* John Wiley & Sons, New York.

[31] Frey, R., Runggaldier, W. J. (2001). A Nonliear Filtering Approach to Volatility Estimation with a View towards High Frequency Data. *International Journal of Theoretical and Applied Finance* **4**, 199-210.

[32] Grother, O., Müller, C. (2007). Liquidity-Based Estimation of Stochastic Volatility under Microstructure Noise. Working paper.

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

[34] Hasbrouck, J. (1991). Measuring the Information Content of Stock Trades. *Journal of Finance* **46**, 179-207.

[35] Hasbrouck, J. (1996). Modelling Market Microstructure Time Series. *Handbook of Statistics. North Holland, Amsterdam.* **14**, 647-692.

[36] Hasbrouck, J. (1999). Security Bid/Ask Dynamics with Discreteness and Clustering: Simple Strategies for Modeling and Estimation. *Journal of Financial Markets* **2**, 1-28.

[37] Heston, S. (1993). A Closed-form Solutions for Options with Stochastic Volatility with Applications to Bond and Currency Options. *Rev. Finan. Stud.* **6**, 327-343.

[38] Heunis, A. J. (1990). On the Stochastic Differential Equations of Filtering Theory. *Applied Mathematics and Computation* **37**, 185-218.

[39] Hu. Y. Z., Øksendal, B., Sulem, A. (2003). Optimal Portfolio in a Fractional Black-Scholes Market Driven by Fractional Brownian Motion. *Infinite Dimensional Analysis, Quantum Probability and Related Topics* **6**, 519-536.

[40] Huang, R., Stoll, A. (2001). Tick Size, Bid-ask Spreads, and Market Structure. *Journal of Financial and Quantitative Analysis* **36**, 503-522.

[41] Hull, J., White, A. (1987). The Pricing of Options on Assets with Stochastic Volatilities. *Journal of Finance* **42**, 281-300.

[42] Jackwerth, J., Rubinstein, M. (1996). Recovering Probability Distributions from Contemporaneous Security Prices. *Journal of Finance* **51**, 1611-1631.

[43] Jacod, J., Shirayev, A. N. (1987). *Limit Theorems for Stochastic Processes.* Springer-Verlag, Berlin.

[44] Kallenberg, O. (1997) *Foundations of Modern Probability.* Springer-Verlag, New York.

[45] Kass, R. E., Raftery, A. E. (1995). Bayes Factor and Model Uncertainty. *Journal of the American Statistical Association* **90**, 773-795.

[46] Klieman, W., Koch, G., Marchetti, F. (1990). On the Unnormalized Solution of the Filtering Problem with Counting Process Observations. *IEEE Trans. Info. Theory* **36**, 1415-1425.

[47] Kouritzin, M. A., Long, H., Sun, W. (2003). Nonlinear Filtering for Diffusions in Random Environments. *Journal of Theoretical Probability* **16**, 1-20.

[48] Kouritzin, M. A., 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.

[49] Lin, S. J. (1995). Stochastic Analysis of Fractional Brownian Motion. *Stochastics Stochastic Rep.* **55**, 121-140.

[50] Mandelbrot, B. B., Van Ness, J. W. (1968). Fractional Brownian Motions, Fractional Brownian Noises and Applications. *SIAM Review* **10**, 422-437.

[51] Mandelbrot, B. B. (1971). When can Price be Arbitraged Efficiently? A Limit to the Validity of the Random Walk and Martingale Models. *Review of Economics and Statistics* **53**, 225-236.

[52] Merton, R. C. (1973). The Theory of Rational Option Pricing. *The Bell Journal of Economics and Management Science* **4**, 141-183.

[53] Nelson, D. B. (1990). ARCH Models as Diffusion Approximations. *Journal of Econometrics* **45**, 7-39.

[54] Nielsen, J., Vestergaard, M. (2000). Estimation in Continuous-time Stochastic Volatility Models Using Nonlinear Filters. *International Journal of Theoretical and Applied Finance* **3**, 279-308.

[55] Pardoux, E. (1979). Stochastic Partial Differential Equations and Filtering of Diffusion Processes. *Stochastics* **3**, 127-167.

[56] Protter, P. (1990). *Stochastic Integration and Differential Equations.* Springer-Verlag, Heidelberg.

[57] Scott, L. O. (1987). Option Pricing when the Virance Changes Randomly: Theory, Estimation and Application. *Journal of Financial Quantitative Analysis* **22**, 419-438.

[58] Stroock, D. W., Varadhan, S. R. S. (1979). *Multidimensional Diffusion Processes.* Springer-Verlag, Berlin.

[59] Viens, F., Tindel, S. (2004). Convergence of a Branching Particle System to the Solution of a Parabolic Stochastic PDE. (2004) *Rand. Operators Stoch. Eqs* **12**, 129-144.

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