Statistical Inference for Stochastic Processes (2005) 8:331–354
DOI 10.1007/s11203-005-6103-8

Bayesian Inference via Filtering for a Class of Counting Processes: Application to the Micromovement of Asset Price

YONG ZENG
Department of Mathematics and Statistics, University of Missouri, Kansas City, MO 64110, USA. e-mail: zeng@mendota.umkc.edu

Abstract. This paper highlights recent developments in a rich class of counting process models for the micromovement of asset price and in the Bayesian inference (estimation and model selection) via filtering for the class of models. A specific micromovement model built upon linear Brownian motion with jumping stochastic volatility is used to demonstrate the procedure to develop a micromovement model with specific tick-level sample characteristics. The model is further used to demonstrate the procedure to implement Bayes estimation via filtering, namely, to construct a recursive algorithm for computing the trade-by-trade Bayes parameter estimates, especially for the stochastic volatility. The consistency of the recursive algorithm model is proven. Simulation and real-data examples are provided as well as a brief example of Bayesian model selection via filtering.

Key words: Bayesian statistics, counting process, estimation, filtering, Markov chain approximation, model selection, price clustering, price discreteness, ultra high frequency data.

1. Introduction

Asset price models can be broadly classified into two categories: macro- and micro-movement models. Macro-movement refers to daily, weekly, and monthly closing price behavior and micro-movement refers to transactional price behavior. These data are refered to as *ultra high frequency data* by Engle [5]. There are strong connections as well as striking distinctions between the macro- and micro-movements. Zeng [13] proposed a general class of partially-observed micro-movement models that bridge the gap between the macro- and micro- movements caused by noise, and tie the sample characteristics of micro-movement and macro-movement of price in a consistent manner. The main appeal of the proposed model is that prices are viewed as a collection of counting processes, each of which represents a price level. Then, the model is framed as a filtering problem with counting process observations. Alternatively, the price process can be constructed from the intrinsic value process by incorporating the trading noise. Zeng [13] also developed continuous-time Bayes parameter estimation via filtering for the model and Kouritzin and Zeng [9] further developed Bayesian model selection via filtering for the class of models based on Bayes factor.

The first objective of this paper is to provide a thorough overview of the rich class of counting process models for analyzing the micromovement of asset price and of the Bayesian inference (estimation and model selection) via filtering for the class of models. The Bayesian paradigm is chosen because it provides not only additional model flexibility, but also the ability to incorporate real prior information. The second objective is to provide a worked example to help those who wish to carry out a similar analysis in practice. A specific micromovement model built upon linear Brownian motion with jumping stochastic volatility(JSV-LBM) is developed to illustrate the modeling building technique. This model is further used to demonstrate the procedure to implement Bayes estimation via filtering, namely, to construct a recursive algorithm for computing the trade-by-trade Bayes estimates. The consistency of the recursive algorithm model is proven. Simulation results which demonstrate the effectiveness of the use of Bayesian filtering to estimate trade-by-trade stochastic volatility are provided. The recursive algorithm is applied to an actual Microsoft data set to obtain trade-by-trade Bayes parameter estimates. A brief example of Bayesian model selection is also provided.

Section 2 presents the class of counting process models for the micromovement of asset price in two equivalent fashions and the equivalence of two models is proven. Section 3 illustrates continuous-time Bayesian inference (estimation and model selection) via filtering for the class of models. Section 4 focuses on developing the micromovement model built on JSV-LBM and applying the Bayesian inference via filtering to the model. Simulation and real-data examples are provided in Section 4 as well as a brief example of Bayesian model selection. Section 5 concludes the paper.

2. The Class of Counting Processes

This section presents the class of micromovement models in [13] in two equivalent ways: one as a collection of counting processes of price levels and the other as a construction of price from the intrinsic value of an asset by incorporating noises.

2.1. MODELING PRICES AS A COLLECTION OF COUNTING PROCESSES

In real world trading, the price of an asset fluctuates as the result of inflowing information and trading noise. In micromovement level, the price does not move continuously as the common asset price models such as diffusion or jump-diffusion processes suggest, but moves level-by-level due to price discreteness (caused by the minimum price variation set by trading regulation). A technical analyst usually analyzes the price movement according to price level in an attempt to extract useful information. From this viewpoint, we can formulate the prices of an asset as a collection of counting processes in the following form:

$$\vec{Y}(t)=\begin{pmatrix} N_1(\int_0^t \lambda_1(\theta(s), X(s), s)\mathrm{d}s) \\ N_2(\int_0^t \lambda_2(\theta(s), X(s), s)\mathrm{d}s) \\ \vdots \\ N_n(\int_0^t \lambda_n(\theta(s), X(s), s)\mathrm{d}s) \end{pmatrix}, \tag{1}$$

where $Y_j(t) = N_j(\int_0^t \lambda_j(\theta(s), X(s), s)\mathrm{d}s)$ is the counting process recording the cumulative number of trades that have occurred at the jth price level (denoted by y_j) up to time t. Moreover, $\theta(t)$ is a vector of parameters in the model and $X(t)$ is the intrinsic value process. Suppose that P is the physical probability measure of $(\theta, X, \vec{Y})$. We make five mild assumptions on the model.

ASSUMPTION 2.1. $\{N_j\}_{j=1}^n$ are unit Poisson processes under measure P.

Then, $Y_j(t) = N_j(\int_0^t \lambda_j(\theta(s), X(s), s)\mathrm{d}s)$ is conditional Poisson process with the stochastic intensity, $\lambda_j(\theta(t), X(t), t)$. Given $\mathcal{F}_t^{\theta,X}$, the filtration of θ and X, $Y_j(t)$ has a Poisson distribution with parameter $\int_0^t \lambda_j(\theta(s), X(s), s)\mathrm{d}s$. Moreover, $Y_j(t) - \int_0^t \lambda_j(\theta(s), X(s), s)\mathrm{d}s$ is a martingale and $\int_0^t \lambda_j(\theta(s), X(s), s)\mathrm{ds}$ is called the compensator of $Y_j(t)$ for each j.

ASSUMPTION 2.2. $(\theta, X), N_1, N_2, \ldots, N_n$ are independent under measure P.

ASSUMPTION 2.3. The total intensity process, $a(\theta, x, t)$, is uniformly bounded above, namely, there exists a positive constant, C, such that $0 \leqslant a(\theta, x, t) \leqslant C$ for all $t > 0$ and (θ, x).

The total intensity $a(\theta, x, t)$ determines the expected rate of trading at time t. These three assumptions imply that there exists a reference measure Q and that after a suitable change of measure to Q, $(\theta, X), Y_1, \ldots, Y_n$ become independent, and $Y_1, Y_2, \ldots, Y_n$ become unit Poisson processes (Bremaud [2]).

ASSUMPTION 2.4. The intensities are of the form: $\lambda_j(\theta, x, t) = a(\theta, x, t) p(y_j|x)$, where $p(y_j|x)$ is the transition probability from x to y_j, the jth price level.

This assumption imposes a desirable structure for the intensities of the model. It means that the total trading intensity $a(\theta(t), X(t), t)$ determines the overall rate of trade occurrence at time t and $p(y_j|x)$ determines the proportional intensity of trade at the price level, y_j, when the value is x. Note that $p(y_j|x)$ models how the trading noise enters the price process. The intensity structure plays an essential role in the equivalence of the two approaches of modeling.

ASSUMPTION 2.5. (θ, X) is the unique solution of a martingale problem for a generator $\mathbf{A}$ such that for a function f in the domain of $\mathbf{A}$, $M_f(t) = f(\theta(t), X(t)) - \int_0^t \mathbf{A} f(\theta(s), X(s))\mathrm{d}s$, is a $\mathcal{F}_t^{\theta,X}$-martingale, where $\mathcal{F}_t^{\theta,X}$ is the σ-algebra generated by $(\theta(s), X(s))_{0 \leqslant s \leqslant t}$.

The martingale problem and the generator approach (Ethier and Kurtz [6]) provide a powerful tool for the characterization of Markov processes. Assumption 2.5 includes all relevant stochastic processes such as diffusion and jump-diffusion processes for modeling asset price. Two examples are given in Section 4.

Under this representation, $(\theta(t), X(t))$ becomes a signal process, which cannot be observed directly, but can be partially observed through the counting processes, $\vec{Y}(t)$, corrupted by trading noise, which is modeled by $p(y_j|x)$. Hence, $(\theta, X, \vec{Y})$ is framed as a *filtering problem with counting process observations*.

2.2. CONSTRUCTING PRICE FROM INTRINSIC VALUE

In Section 2.1, we view the prices according to the price levels. Alternatively, we can construct the prices in the order of trading occurrence over time. The construction is predicated on the simple intuition that the price is formed from an intrinsic value by incorporating the noises that arise from the trading activity. In general, there are three steps in constructing the price process Y from the value process X. First, we specify X. This is done by Assumption 2.5, where (θ, X) are jointly specified in order to allow time-dependent parameters such as stochastic volatility and to prepare for parameter estimation. Next, we determine trading times $t_1, t_2, \ldots, t_i, \ldots$, which are driven by a conditional Poisson process with an intensity $a(X(t), \theta(t), t)$. Finally, $Y(t_i)$, the price at time t_i, is modeled by $Y(t_i) = F(X(t_i))$, where $y = F(x)$ is a random transformation with the transition probability $p(y|x)$, the same as $p(y_j|x)$ in Assumption 2.4 assuming y is at the jth price level.

Under this construction, the observable price is produced from the value process by combining noises when a trade occurs. Information affects $X(t)$, the value of an asset, and has a permanent influence on the price while noise affects $F(x)$ (or $p(y|x)$), and only has a transitory impact on price. The random transformation, $F(x)$, is flexible and Section 4.1 constructs one to accommodate the three important types of noise: discrete, clustering and non-clustering. The construction of price is similar to the state-space model, but here we view the price process Y as a *marked point process* (MPP), where the mark is price. Frey and Runggaldier [7] and Cvitanic et al. [3] proposed similar MPP models, but both do not incorporate trading noise.

2.3. THE EQUIVALENCE OF THE TWO APPROACHES OF MODELING

Section 2.1 models the price process as a collection of counting processes, while Section 2.2 constructs the price from the value. In this section, their equivalence in distribution is proven. This guarantees the statistical inference based on the former modeling is also valid to the latter one.

Proposition 2.1. *Suppose that The price vector process,* $\vec{Y}$, *is modeled by equation* (1) *with Assumptions* 2.1–2.5. *Suppose that the price process,* Y, *is constructed from the value process,* X, *by the three general steps in Section* 2.2, (θ, X) *satisfies Assumption* 2.5 *and* $a(\theta, x, t)$ *satisfies Assumption* 2.3. *Then,* $\vec{Y}$ *and* Y *have the same probability distribution.*

Proof. First, we show that the intensity of $Y_j(t)$ in Assumption 2.4 can be obtained from the construction approach in Section 2.2 by a heuristic argument as the following:

$$\begin{aligned}
&P\{Y_j(t+h) - Y_j(t) > 0 \mid \mathcal{F}_t^{\theta,X}\}\\
&\quad = P\{\text{a trade occurs during } (t, t+h] \mid \mathcal{F}_t^{\theta,X}\}\\
&\qquad P\{\text{at the } j\text{th level} \mid \text{a trade occurs}, \mathcal{F}_t^{\theta,X}\}\\
&\quad \approx a(\theta(t), X(t), t)\, p(y_j \mid X(t))h. \qquad (2)
\end{aligned}$$

That is, the *stochastic intensity kernels* of the two formulations of MPP are the same. Then, the two formulations have the same compensator and therefore, have the same probability distribution also.[1] □

3. Bayesian Inference via Filtering

This section presents the foundations of statistical inference for the class of models, and presents a convergence theorem, which provides a recipe to construct consistent recursive algorithms to approximate continuous-time likelihoods, posterior, and Bayes factors.

[1] Related theorems supporting the above arguments can be found in the book of Last and Brandt [11], especially, Theorems 8.3.3 and 8.2.2.

3.1. FOUNDATIONS OF STATISTICAL INFERENCE

In this section, we study the continuous-time joint likelihood, the likelihood, the integrated likelihood, the posterior of the proposed model as well as the continuous-time Bayes factors for Models 1 and 2. Moreover, we establish their relationships with the unnormalized and normalized filtering equations as well as the system of evolution equations, respectively.

3.1.1. *The Continuous-time Joint Likelihood*

The probability measure P of $\vec{Y}(t)$ can be written as $P = P_{\theta,x} \times P_{y|\theta,x}$, where $P_{\theta,x}$ is the probability measure for (θ, X) such that $M_f(t)$ in Assumption 2.5 is a $\mathcal{F}_t^{\theta,X}$-martingale, and $P_{y|\theta,x}$ is the conditional probability measure on $D_{R^n}[0,\infty)$ for $\vec{Y}$ given (θ, X) (where $D_{R^n}[0,\infty)$ is the space of right continuous with left limit functions). Under P, $\vec{Y}$ depends on (θ, X). Recall that there exists a reference measure Q such that under Q, $(\theta, X), \vec{Y}$ become independent, (θ, X) remains the same probability distribution and $Y_1, Y_2, \dots, Y_n$ become unit Poisson processes. Therefore, Q can be decomposed as $Q = P_{\theta,x} \times Q_y$, where Q_y is the probability measure for n independent unit Poisson processes. One can obtain the Radon–Nikodym derivative of the model, that is the joint likelihood of $(\theta, X, \vec{Y})$, $L(t)$, as (see [2] p. 166),

$$\begin{aligned} L(t) &= \frac{\mathrm{d}P}{\mathrm{d}Q}(t) = \frac{\mathrm{d}P_{\theta,x}}{\mathrm{d}P_{\theta,x}}(t) \times \frac{\mathrm{d}P_{y|\theta,x}}{\mathrm{d}Q_y}(t) = \frac{\mathrm{d}P_{y|\theta,x}}{\mathrm{d}Q_y}(t) \\ &= \prod_{k=1}^{n} \exp\left\{ \int_0^t \log \lambda_k(\theta(s-), X(s-), s-) \mathrm{d}Y_k(s) \right. \\ &\quad \left. - \int_0^t \Big[\lambda_k(\theta(s), X(s), s) - 1 \Big] \mathrm{d}s \right\}. \end{aligned} \tag{3}$$

or in SDE form:

$$L(t) = 1 + \sum_{k=1}^{n} \int_0^t \Big[\lambda_k(\theta(s-), X(s-), s-) - 1 \Big] L(s-) \mathrm{d}(Y_k(s) - s).$$

3.1.2. *The Continuous-time Likelihoods of $\vec{Y}$*

However, X and the stochastic components of θ (such as stochastic volatility) cannot be observed. What is needed for the statistical analysis is the likelihood of $\vec{Y}$ alone, which can be obtained by using conditional expectation. Let $\mathcal{F}_t^{\vec{Y}} = \sigma\{(\vec{Y}(s)) | 0 \leqslant s \leqslant t\}$ be all the available information up to time t. We use $E^Q[X]$ to indicate that the expectation is taken with respect to measure Q.

DEFINITION 3.1. *Let ϕ_t be the conditional measure of* $(\theta(t), X(t))$ *given* $\mathcal{F}_t^{\vec{Y}}$ *defined as*

$$\phi_t\{(\theta(t), X(t)) \in A\} = E^Q\left[\mathbf{I}_{\{(\theta(t), X(t)) \in A\}}(\theta(t), X(t))L(t)|\mathcal{F}_t^Y\right].$$

DEFINITION 3.2. *Let*

$$\phi(f, t) = E^Q[f(\theta(t), X(t))L(t)|\mathcal{F}_t^{\vec{Y}}] = \int f(\theta, x)\phi_t(\mathrm{d}\theta, \mathrm{d}x).$$

If $(\theta(0), X(0))$ is fixed, then the likelihood of Y is $E^Q[L(t) \mid \mathcal{F}^{\vec{Y}}] = \phi(1, t)$. If a prior is assumed on $(\theta(0), X(0))$ as in Bayesian paradigm, then the integrated (or marginal) likelihood of Y is also $\phi(1, t)$.

3.1.3. *The Continuous-time Posterior*

DEFINITION 3.3. *Let π_t be the conditional distribution of* $(\theta(t), X(t))$ *given* $\mathcal{F}_t^{\vec{Y}}$ *and let*

$$\pi(f, t) = E^P[f(\theta(t), X(t))|\mathcal{F}_t^{\vec{Y}}] = \int f(\theta, x)\pi_t(\mathrm{d}\theta, \mathrm{d}x).$$

If a prior is assumed on $(\theta(0), X(0))$ as in Bayesian inference, then π_t becomes the continuous-time posterior, which is determined by $\pi(f, t)$ for all continous and bounded f. Bayes Theorem (see [2], p. 171) provides the relationship between $\phi(f, t)$ and $\pi(f, t)$: $\pi(f, t) = \phi(f, t)/\phi(1, t)$. Hence, the equation governing the evolution of $\phi(f, t)$ is called the *unnormalized filtering equation*, and that of $\pi(f, t)$ is called the *normalized filtering equation.*

3.1.4. *Continuous-time Bayes Factors*

Denote Model k by $(\theta^{(k)}, X^{(k)}, \vec{Y}^{(k)})$ for $k = 1, 2$. Denote the joint likelihood of $(\theta^{(k)}, X^{(k)}, \vec{Y}^{(k)})$ by $L^{(k)}(t)$, as in equation (3). Denote $\phi_k(f_k, t) = E^{Q^{(k)}}[f_k(\theta^{(k)}(t), X^{(k)}(t))L^{(k)}(t)|\mathcal{F}_t^{\vec{Y}^{(k)}}]$. Then, the integrated likelihood of $\vec{Y}$ is $\phi_k(1, t)$, for Model k.

In general, the Bayes factor of Model 2 over Model 1, B_{21}, is defined as the ratio of integrated likelihoods of Model 2 over Model 1, i.e. $B_{21}(t) = \phi_2(1, t)/\phi_1(1, t)$. Suppose B_{21} has been calculated. Then, we can interpret it using Table I furnished by Kass and Raftery [8] as guideline. Similarly, we can define B_{12}, the Bayes factor of Model 1 over Model 2. Instead of normalizing by its integrated likelihood, when the measure ϕ_k is normalized by the integrated likelihood of the other model, a conditional finite measure is obtained.

Table I. Interpretation of Bayes factor

B_{21}	Evidence against model 1
1–3	Not worth more than a bare mention
3–12	Positive
12–150	Strong
>150	Decisive

DEFINITION 3.4. *For* $k = 1, 2$, *let* $q_t^{(k)}$ *be the conditional measure of* $(\theta^{(k)}(t), X^{(k)}(t))$ *given* $\mathcal{F}_t^{\vec{Y}^{(k)}}$:

$$q_t^{(k)}\left\{(\theta^{(k)}(t), X^{(k)}(t)) \in A\right\} = \frac{E^Q\left[\mathbf{I}_{\{(\theta^{(k)}(t), X^{(k)}(t))\in A\}}(\theta^{(k)}(t), X^{(k)}(t))L^{(k)}(t)|\mathcal{F}_t^{Y^{(k)}}\right]}{\phi_{3-k}(1,t)}$$
$$= E^Q\left[\mathbf{I}_{\{(\theta^{(k)}(t), X^{(k)}(t))\in A\}}(\theta^{(k)}(t), X^{(k)}(t))\tilde{L}^{(k)}(t)|\mathcal{F}_t^{Y^{(k)}}\right],$$

where $\tilde{L}^{(1)}(t) = L^{(1)}(t)/\phi_{3-k}(1,t)$. The second equality is because $\phi_k(1,t)$, $k = 1, 2$, depend on the same filtration $\mathcal{F}^{Y^{(1)}} = \mathcal{F}^{Y^{(2)}}$ and $\phi_k(1,t)$ can be moved inside the conditioning.

DEFINITION 3.5. *Define the filter ratio processes*:

$$q_1(f_1, t) = \frac{\phi_1(f_1, t)}{\phi_2(1, t)}, \quad \textit{and} \quad q_2(f_2, t) = \frac{\phi_2(f_2, t)}{\phi_1(1, t)}.$$

Observe that the Bayes factors, $B_{12}(t) = q_1(1,t)$ and $B_{21}(t) = q_2(1,t)$. Moreover, observe that $q_k(f_k, t)$ can be written as $q_k(f_k, t) = \int f_k(\theta^{(k)}, x^{(k)})q_t^{(k)}$ $(\mathrm{d}\theta^{(k)}, \mathrm{d}x^{(k)})$. The integral forms of $\phi(f,t)$, $\pi(f,t)$, and $q_k(f_k,t)$ are important in deriving the recursive algorithms where f (or f_k) is taken to be a lattice-point indicator function.

There are two advantages of Bayes factors. First, a Bayes factor neither requires the models to be nested, nor does it require the probability measures of the models to be absolutely continuous with respect to those of the total models in hypotheses. Second, Kass and Raftery [8] show that under some conditions, Bayes factor ≈ BIC (Bayesian Information Criterion), which penalizes according to both the number of parameters and the number of data. This suggests that Bayes factor has this desirable property also.

3.1.5. *Filtering and Evolution Equations*

The filtering equations provide an effective means to characterize the infinite dimensional continuous-time versions including likelihoods, posteriors, likelihood ratios and Bayes factors. The unnormalized filtering equation characterizes the likelihoods. The normalized filtering equation, which is the optimum filter in the sense of least mean square error, characterizes the posteriors. The system of evolution equations for $q_k(f_k, t)$, $k = 1, 2$ characterizes the likelihood ratios or the Bayes factors. The following two theorems summarize all the useful equations.

THEOREM 3.1. *Suppose that* $(\theta, X, \vec{Y})$ *satisfies Assumptions 2.1–2.5. Then,* ϕ_t *is the unique measure-valued solution of the SDE, the unnormalized filtering equation,*

$$\begin{aligned}\phi(f,t) = \phi(f,0) &+ \int_0^t \phi(\mathbf{A}f - (a-n)f, s)\mathrm{d}s \\ &+ \sum_{k=1}^{n} \int_0^t \phi((ap_k - 1)f, s-)\mathrm{d}Y_k(s),\end{aligned} \tag{4}$$

for $t > 0$ *and* $f \in D(\mathbf{A})$, *the domain of generator* $\mathbf{A}$, *where* $a = a(\theta(t), X(t), t)$, *is the trading intensity, and* $p_k = p(y_k|x)$ *is the transition probability from* x *to* y_k.

π_t *is the unique measure-valued solution of the SDE, the normalized filtering equation*,

$$\begin{aligned}\pi(f,t) = \pi(f,0) &+ \int_0^t \big[\pi(\mathbf{A}f, s) - \pi(fa, s) + \pi(f,s)\pi(a,s)\big]\mathrm{d}s \\ &+ \sum_{k=1}^{n} \int_0^t \left[\frac{\pi(fap_k, s-)}{\pi(ap_k, s-)} - \pi(f, s-)\right]\mathrm{d}Y_k(s).\end{aligned} \tag{5}$$

Moreover, when the trading intensity is deterministic, that is, $a(\theta(t), X(t), t) = a(t)$, the normalized filtering equation is simplified as

$$\begin{aligned}\pi(f,t) = \pi(f,0) &+ \int_0^t \pi(\mathbf{A}f, s)\mathrm{d}s \\ &+ \sum_{k=1}^{n} \int_0^t \left[\frac{\pi(fp_k, s-)}{\pi(p_k, s-)} - \pi(f, s-)\right]\mathrm{d}Y_k(s).\end{aligned} \tag{6}$$

THEOREM 3.2. *Suppose Model* k $(k = 1, 2)$ *has generator* $\mathbf{A}^{(k)}$ *for* $(\theta^{(k)}, X^{(k)})$, *the trading intensity* $a_k = a_k(\theta^{(k)}(t), X^{(k)}(t), \vec{Y}^{(k)}(t))$, *and the transition probability* $p_j^{(k)} = p^{(k)}(y_j|x)$ *from* x *to* y_j *for the random transformation* $F^{(k)}$. *Suppose that* $(\theta^{(k)}, X^{(k)}, \vec{Y}^{(k)})$ *satisfies Assumptions 2.1–2.5. Then,*

$\left(q_t^{(1)}, q_t^{(2)}\right)$ *are the unique measure-valued pair solution of the following system of SDEs,*

$$q_1(f_1,t) = q_1(f_1,0) + \int_0^t \left[q_1(\mathbf{A}^{(1)} f_1, s) - q_1(a_1 f_1, s) + \frac{q_1(f_1,s) q_2(a_2,s)}{q_2(1,s)} \right] \mathrm{d}s + \sum_{j=1}^{n} \int_0^t \left[\frac{q_1(f_1 a_1 p_j^{(1)}, s-)}{q_2(a_2 p_j^{(2)}, s-)} q_2(1, s-) - q_1(f_1, s-) \right] \mathrm{d}Y_j(s), \tag{7}$$

$$q_2(f_2,t) = q_2(f_2,0) + \int_0^t \left[q_2(\mathbf{A}^{(2)} f_2, s) - q_2(a_2 f_2, s) + \frac{q_2(f_2,s) q_1(a_1,s)}{q_1(1,s)} \right] \mathrm{d}s + \sum_{j=1}^{n} \int_0^t \left[\frac{q_2(f_2 a_2 p_j^{(2)}, s-)}{q_1(a_1 p_j^{(1)}, s-)} q_1(1, s-) - q_2(f_2, s-) \right] \mathrm{d}Y_j(s) \tag{8}$$

for all $t > 0$ *and* $f_k \in D(\mathbf{A}^{(k)})$ *for* $k = 1, 2$. *When* $a_1(\theta^{(1)}(t), X^{(1)}(t), t) = a_2(\theta^{(2)}(t), X^{(2)}(t), t) = a(t)$, *the above two equations are simplified to*

$$q_1(f_1,t) = q_1(f_1,0) + \int_0^t q_1(\mathbf{A}^{(1)} f_1, s) \mathrm{d}s + \sum_{j=1}^{n} \int_0^t \left[\frac{q_1(f_1 p_j^{(1)}, s-)}{q_2(p_j^{(2)}, s-)} q_2(1, s-) - q_1(f_1, s-) \right] \mathrm{d}Y_j(s), \tag{9}$$

$$q_2(f_2,t) = q_2(f_2,0) + \int_0^t q_2(\mathbf{A}^{(2)} f_2, s) \mathrm{d}s + \sum_{j=1}^{n} \int_0^t \left[\frac{q_2(f_2 p_j^{(2)}, s-)}{q_1(p_j^{(1)}, s-)} q_1(1, s-) - q_2(f_2, s-) \right] \mathrm{d}Y_j(s). \tag{10}$$

The proof of Theorem 3.1 is in [13] and that of Theorem 3.2 is [9].

Note that $a(t)$ disappears in equations (6), (9) and (10). This reduces the computation greatly in computing the Bayes estimates and Bayes factors. The tradeoff of taking a_i independent of $(\theta^{(k)}, X^{(k)})$ is that the relationship between trading intensity and other parameters (such as stochastic volatility) is excluded.

Let the trading times be $t_1, t_2, \ldots$, then, for example, equation (6) can be written in two parts. The first is called the *propagation equation*, describing the evolution without trades and the second is called the *updating equation*, describing the update when a trade occurs. The propagation equation has no random component and is written as

$$\pi(f, t_{i+1}-) = \pi(f, t_i) + \int_{t_i}^{t_{i+1}-} \pi(\mathbf{A} f, s) \mathrm{d}s. \tag{11}$$

This implies that when there are no trades, the posterior evolves deterministically.

Assume the price at time t_{i+1} occurs at the jth price level, then the updating equation is

$$\pi(f, t_{i+1}) = \frac{\pi(fp_j, t_{i+1}-)}{\pi(p_j, t_{i+1}-)}. \tag{12}$$

It is random because the price level j, which is the observation, is random. Similarly, Equations for Bayes factors can be written in such two parts.

3.2. A CONVERGENCE THEOREM AND RECURSIVE ALGORITHMS

Theorems 3.1 and 3.2 provide the evolutions of the continuous-time versions. To compute them, one constructs algorithms to approximate them. The algorithms, based on the evolutions of SDEs, are naturally recursive, handling a datum at a time. Thus, the algorithm can make real-time updates and handle large data sets. One basic requirement for the recursive algorithms is consistency: The approximate versions, computed by the recursive algorithms, converges to the true ones. The following theorem proves the consistency of the approximate versions and provides a recipe for constructing consistent algorithms through Kushner's Markov chain approximation methods.

For $k=1,2$, let $(\theta_\epsilon^{(k)}, X_\epsilon^{(k)})$ be an approximation of $(\theta^{(k)}, X^{(k)})$. Then, we define

$$\vec{Y}_\epsilon^{(k)}(t) = \begin{pmatrix} N_1^{(k)}(\int_0^t \lambda_1(\theta_\epsilon^{(k)}(s), X_\epsilon^{(k)}(s), s)\mathrm{d}s) \\ N_2^{(k)}(\int_0^t \lambda_2(\theta_\epsilon^{(k)}(s), X_\epsilon^{(k)}(s), s)\mathrm{d}s) \\ \vdots \\ N_{n_k}^{(k)}(\int_0^t \lambda_n(\theta_\epsilon^{(k)}(s), X_\epsilon^{(k)}(s), s)\mathrm{d}s) \end{pmatrix}, \tag{13}$$

set $\mathcal{F}_t^{\vec{Y}_\epsilon^{(k)}} = \sigma(\vec{Y}_\epsilon^{(k)}(s), 0 \leqslant s \leqslant t)$, take $L_\epsilon^{(k)}(t) = L((\theta_\epsilon^{(k)}(s), X_\epsilon^{(k)}(s), Y_\epsilon^{(k)}(s))_{0 \leqslant s \leqslant t})$ as in equation(3), and use the notation, $X_\epsilon \Rightarrow X$, to mean X_ϵ converges weakly to X in the Skorohod topology as $\epsilon \to 0$. Suppose that $(\theta_\epsilon^{(k)}, X_\epsilon^{(k)}, \vec{Y}_\epsilon^{(k)})$ lives on $(\Omega_\epsilon^{(k)}, \mathcal{F}_\epsilon^{(k)}, P_\epsilon^{(k)})$, and Assumptions 2.1–2.5 also hold for $(\theta_\epsilon^{(k)}, X_\epsilon^{(k)}, \vec{Y}_\epsilon^{(k)})$. Then, there also exists a reference measure $Q_\epsilon^{(k)}$ with similar properties. Next, we define the approximations of $\phi_k(f_k, t)$, $\pi_k(f_k, t)$, and $q_k(f_k, t)$.

DEFINITION 3.6. *For $k=1,2$, let*
$\phi_{\epsilon,k}(f_k, t) = E^{Q_\epsilon^{(k)}}\left[f_k\big(\theta_\epsilon^{(k)}(t), X_{\epsilon_x}^{(k)}(t)\big)L_\epsilon^{(k)}(t)|\mathcal{F}_t^{\vec{Y}_\epsilon^{(k)}}\right]$,
$\pi_{\epsilon,k}(f_k, t) = E^{P_\epsilon^{(k)}}\left[f_k\big(\theta_\epsilon^{(k)}(t), X_{\epsilon_x}^{(k)}(t)\big)|\mathcal{F}_t^{\vec{Y}_\epsilon^{(k)}}\right]$,
$q_{\epsilon,1}(f_1, t) = \phi_{\epsilon,1}(f_1, t)/\phi_{\epsilon,2}(1, t)$ *and* $q_{\epsilon,2}(f_2, t) = \phi_{\epsilon,2}(f_2, t)/\phi_{\epsilon,1}(1, t)$.

THEOREM 3.3. *Suppose that Assumptions* 2.1–2.5 *hold for the models* $(\theta^{(k)}, X^{(k)}, \vec{Y}^{(k)})_{k=1,2}$ *and that Assumptions 2.1–2.5 hold for the approximate models* $(\theta_\epsilon^{(k)}, X_\epsilon^{(k)}, \vec{Y}_\epsilon^{(k)})$. *Suppose* $(\theta_\epsilon^{(k)}, X_\epsilon^{(k)}) \Rightarrow (\theta^{(k)}, X^{(k)})$ *as* $\epsilon \to 0$. *Then, as* $\epsilon \to 0$, *for all bounded continuous functions,* f_1 *and* f_2, *and* $k = 1, 2$,

(i) $\vec{Y}_\epsilon^{(k)} \Rightarrow \vec{Y}^{(k)}$; *(ii)* $\phi_{\epsilon,k}(f_k, t) \Rightarrow \phi_k(f_k, t)$; *(iii)* $\pi_{\epsilon,k}(f_k, t) \Rightarrow \pi_k(f_k, t)$; *(iv)* $q_{\epsilon,1}(f_1, t) \Rightarrow q_1(f_1, t)$ *and* $q_{\epsilon,2}(f_2, t) \Rightarrow q_2(f_2, t)$ *simultaneously.*

The proofs for (i) and (iii) are in [13] and those for (ii) and (iv) are in [9]. Another short proof of the whole theorem is provided in [14].

Part (ii) implies the consistency of the integrated likelihood, part (iii) implies the consistency of posterior and part (iv) implies the consistency of the Bayes factors.

This theorem provides a three-step recipe for constructing a consistent recursive algorithm based on Kushner's Markov chain approximation method to compute the continuous-time versions. For example, to compute the posterior and Bayes estimates for a model (then the superscript, "(k)" is excluded), step 1 is to construct $(\theta_\epsilon, X_\epsilon)$, the Markov chain approximation to (θ, X), and obtain $p_{\epsilon,j} = p(y_j|\theta_\epsilon, x_\epsilon)$ as an approximation to $p_j = p(y_j|\theta, x)$, where $(\theta_\epsilon, x_\epsilon)$ is restricted to the state space of $(\theta_\epsilon, X_\epsilon)$. Step 2 is to obtain the filtering equation for $\pi_\epsilon(f, t)$ corresponding to $(\theta_\epsilon, X_\epsilon, Y_\epsilon, p_{\epsilon,j})$ by applying Theorem 3.1. For simplicity, one only considers the case when $a = a(t)$. Similar to Remark 3.9, the filtering equation for the approximate model can be separated into the propagation equation:

$$\pi_\epsilon(f, t_{i+1}-) = \pi_\epsilon(f, t_i) + \int_{t_i}^{t_{i+1}-} \pi_\epsilon(\mathbf{A}_\epsilon f, s)\mathrm{d}s \tag{14}$$

and the updating equation (assuming that a trade at jth price level occurs at time t_{i+1}):

$$\pi_\epsilon(f, t_{i+1}) = \frac{\pi_\epsilon(f p_{\epsilon,j}, t_{i+1}-)}{\pi_\epsilon(p_{\epsilon,j}, t_{i+1}-)}. \tag{15}$$

Step 3 converts equations (14) and (15) to the recursive algorithm in discrete state space and time by two substeps: (a) represents $\pi_\epsilon(\cdot, t)$ as a finite array with the components being $\pi_\epsilon(f, t)$ for lattice-point indicator f and (b) approximates the time integral in (14) with an Euler scheme.

4. An Example with Simulation and Real Data Results

This section illustrates how to develop a specific micromovement model and how to construct a recursive algorithm. Simulation and real-data examples are supplied.

4.1. A SPECIFIC MICROMOVEMENT MODEL

Since it is intuitive to construct the model, it is presented constructively. Recall that there are three steps. First, $X(t)$, is assumed to be JSV-LBM, or in SDE form:

$$\mathrm{d}X(t)=\mu\mathrm{d}t+\sigma(t)\mathrm{d}W(t),\qquad \mathrm{d}\sigma(t)=\big(J_{N(t-)+1}-\sigma(t-)\big)\mathrm{d}N(t),\tag{16}$$

where $W(t)$ is a standard Brownian motion, $N(t)$ is a Poisson process with intensity λ and is independent of $W(t)$, and $\{J_i\}$ is a sequence of i.i.d. random variables independent of $W(t)$ and $N(t)$. We further assume that each J_i is uniformly distributed on a range, $[\alpha_\sigma,\beta_\sigma]$. Suppose that the ith Poisson event happens at time t_i and J_i is drawn. Then, the volatility changes from $\sigma(t_{i-1})$ to J_i at time t_i (because $\sigma(t_i)=(J_i-\sigma(t_{i-1}))+\sigma(t_i-)$), and then remains the same until the next Poisson event occurs. The generator is

$$\begin{aligned}\mathbf{A}f(\sigma,x)=&\,\mu\frac{\partial f}{\partial x}(\sigma,x)+\frac{1}{2}\sigma^2\frac{\partial^2 f}{\partial x^2}(\sigma,x)\\&+\lambda\int_{\alpha_\sigma}^{\beta_\sigma}\big(f(z,x)-f(\sigma,x)\big)\frac{1}{\beta_\sigma-\alpha_\sigma}\mathrm{d}z.\end{aligned}$$

Second, the trading intensity is assumed to be deterministic. Then, $a(t)$ drops out in the normalized filtering equation for posterior and the evolution equations for Bayes factor. A time-dependent deterministic intensity $a(t)$ fits the data of trade duration better than the time-invariant one since trading activity is higher in the opening and the closing periods.

Lastly, we incorporate the trading noises on the intrinsic values at trading times to obtain the price process. There are three important types of noise have been identified and extensively studied: discrete, clustering, and non-clustering. First, intraday prices move discretely, resulting in "discrete noise", because the tick size, $\$1/M$, is set by trading regulation. Second, because prices do not distribute evenly on all ticks, but cluster on integer and half ticks, "price clustering" is obtained. Third, the distribution of price changes (the difference between the current price and the latest previous price) and the outliers in prices show the existence of the "non-clustering noise", which includes all other unspecified noise. To simplify notation, at a trading time t_i, set $x=X(t_i)$, $y=Y(t_i)$, and $y'=Y'(t_i)=R[X(t_i)+V_i,1/M]$, where V_i is to be defined as the non-clustering noise. Instead of directly specifying $p(y|x)$, we define $y=F(x)$ in three steps:

Step 1. Incorporate non-clustering noise by adding V; $x'=x+V$, where V is the non-clustering noise of trade i at time t_i. We assume $\{V_i\}$, are independent of the value process, and they are i.i.d. with a doubly geometric distribution:

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

Step 2. Incorporate discrete noise by rounding off x' to its closest tick, $y'=R[x',1/M]$.

Step 3. Incorporate clustering noise by biasing y' through a random biasing function $b(\cdot)$. $\{b_i(\cdot)\}$ is assumed independent of $\{y_i'\}$. To be consistent with the data analyzed in Section 4.4, we construct a simple random biasing function only for the tick of 1/8 dollar (i.e. $M=8$). The data to be analyzed has this clustering phenomenon: integers and halves are most likely and have about the same frequencies; odd quarters are the second most likely and have about the same frequencies; and odd eighthes are least likely and have about the same frequencies. To generate such clustering, a random biasing function is constructed based on the following rules: if the fractional part of y' is an even eighth, then y stays on y' with probability one; if the fractional part of y' is an odd eighth, then y stays on y' with probability $1-\alpha-\beta$, y moves to the closest odd quarter with probability α, and moves to the closest half or integer with probability β.

In summary, the construction of the price from the value at a trading time is

$$Y(t_i)=b_i\left(R\left[X(t_i)+V_i,\frac{1}{M}\right]\right)=F(X(t_i)).$$

In this way, the random transformation, $F(x)$, which models the impact of trading noise, is specified. The detail of $b(\cdot)$, and the explicit transition probability $p(y|x)$ of F can be found in Appendix A. Simulations can demonstrate that the constructed $F(x)$ are able to capture the tick-level sample characteristics of transaction data for more details see [13].

When the value process is JSV-GBM, the model is studied in [14]. For different tick size M, the modeling of trading noise is different and we can generalize the above modeling to other ticks such as 1/16 or 1/100\$. A more complicated F for 1/16 is constructed in [12]. Larger M may increase the number of lattice for the state space of X and thus increases the compuatation of the recursive algorithm. The parameters of clustering noise, α and β, can be estimated through the method of relative frequency. The other parameters, $\mu,\sigma(t)$, λ and ρ, are estimated by Bayes estimation via filtering through the recurisve algorithm to be constructed.

4.2. THE RECURSIVE ALGORITHM FOR BAYES ESTIMATION

Following the three-step recipe provided by Theorem 3.3, we apply the Markov chain approximation method to construct a consistent algorithm in discrete state space and time to calculate the posteriors and Bayes estimates recursively.

Step 1. Construct a Markov chain, $(\theta_\epsilon(t), X_\epsilon(t))$ as an approximation to $(\theta(t), X(t))$. Here, $\theta(t)=(\mu, \sigma(t), \lambda, \rho)$, which is time-dependent.

First, applying an important idea in Bayesian analysis for latent variables, we treat the unknown parameters as part of the unobserved signal. Then, we discretize the parameter spaces of μ, λ, ρ and the state spaces of $\sigma(t)$ and $X(t)$. Suppose there are $n_\mu+1$, $n_\sigma+1$, $n_\lambda+1$, $n_\rho+1$ and n_x+1 lattices in the discretized spaces of μ, σ, λ, ρ and X, respectively. For example, the discretization for μ in $[\alpha_\mu, \beta_\mu]$ is $\{\alpha_\mu, \alpha_\mu+\epsilon_\mu, \alpha_\mu+2\epsilon_\mu, \dots, \alpha_\mu+j\epsilon_\mu, \dots, \alpha_\mu+n_\mu\epsilon_\mu\}$ where $\alpha_\mu+n_\mu\epsilon_\mu=\beta_\mu$. Define $\mu_j=\alpha_\mu+j\epsilon_\mu$, the jth lattice in the discretized parameter space of μ. Similarly, define $\sigma_k=\sigma_k(t)=\alpha_\sigma+k\epsilon_\sigma$, $\lambda_l=\alpha_\lambda+l\epsilon_\lambda$, $\rho_m=\alpha_\rho+m\epsilon_\rho$, $x_w=x_w(t)=\alpha_x+w\epsilon_x$, and let $\varepsilon=\max(\epsilon_\mu, \epsilon_\sigma, \epsilon_\lambda, \epsilon_\rho, \epsilon_x)$. The discretized space of (μ, λ, ρ) is a natural approximation for the parameter space. But, Markov chain is a natural approximation for the stochastic processes $(\sigma(t), X(t))$.

Second, we observe that the construction of a Markov chain approximation can be achieved by constructing a Markov chain with generator, $\mathbf{A}_\varepsilon$, such that $\mathbf{A}_\varepsilon \to \mathbf{A}$ as $\varepsilon \to 0$. Accommodating other parameters, the generator of the model becomes

$$\begin{aligned}\mathbf{A}f(\mu,\sigma,\lambda,\rho,x) &= \mu\frac{\partial f}{\partial x}(\mu,\sigma,\lambda,\rho,x)+\frac{1}{2}\sigma^2\frac{\partial^2 f}{\partial x^2}(\mu,\sigma,\lambda,\rho,x)\\ &\quad +\lambda\int_{\alpha_\sigma}^{\beta_\sigma}\big(f(\mu,z,\lambda,\rho,x)-f(\mu,\sigma,\lambda,\rho,x)\big)\frac{1}{\beta_\sigma-\alpha_\sigma}\mathrm{d}z. \quad (17)\end{aligned}$$

The generator can be decomposed into two parts: one part is a diffusion generator involving first- and second-order differentiation and the other is a jump generator involving integration. The finite difference approximation is applied for differentiation and the rectangle approximation for integration. One constructs $\mathbf{A}_\varepsilon$ as follows:

$$\begin{aligned}&\mathbf{A}_\varepsilon f(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w)\\ &\quad=\mu_j\Big(\frac{f(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w+\epsilon_x)-f(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w-\epsilon_x)}{2\epsilon_x}\Big)\\ &\qquad+\frac{1}{2}\sigma_k^2\Big(\frac{f(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w+\epsilon_x)+f(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w-\epsilon_x)}{\epsilon_x^2}\Big)\\ &\qquad-\frac{1}{2}\sigma_k^2\Big(\frac{2f(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w)}{\epsilon_x^2}\Big)\end{aligned}$$

$$
\begin{aligned}
&+\lambda_l\sum_{i=0}^{n_\sigma}\big(f(\mu_j,\sigma_i,\lambda_l,\rho_m,x_w)-f(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w)\big)\frac{1}{n_\sigma+1}\\
&=b(\mu_j,\sigma_k,x_w)(f(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w+\epsilon_x)-f(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w))\\
&\quad+d(\mu_j,\sigma_k,x_w)(f(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w-\epsilon_x)-f(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w))\\
&\quad+\lambda_l\big(\bar f(\mu_j,\lambda_l,\rho_m,x_w)-f(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w)\big), \qquad (18)
\end{aligned}
$$

where

$$
b(\mu_j,\sigma_k,x_w)=\frac{1}{2}\left(\frac{\sigma_k^2}{\epsilon_x^2}+\frac{\mu_j}{\epsilon_x}\right),\qquad \mathrm{d}(\mu_j,\sigma_k,x_w)=\frac{1}{2}\left(\frac{\sigma_k^2}{\epsilon_x^2}-\frac{\mu_j}{\epsilon_x}\right)
$$

and

$$
\bar f(\mu_j,\lambda_l,\rho_m,x_w)=\frac{1}{n_\sigma+1}\sum_{i=0}^{n_\sigma}f(\mu_j,\sigma_i,\lambda_l,\rho_m,x_w).
$$

Note that $b(\mu_j,\sigma_k,x_w)$ and $d(\mu_j,\sigma_k,x_w)$ are birth and death rates, which should be kept nonnegative by making ϵ_x sufficiently small. $\bar f(\mu_j,\lambda_l,\rho_m,x_w)$ is the mean of f on σ with fixed $(\mu_j,\lambda_l,\rho_m,x_w)$.

We can choose the range for each component large enough so that the transition probability to anything outside the range is negligible. Then, the generators, $\mathbf{A}_\varepsilon$, characterizes the continuous-time Markov chain satisfying Assumption 2.5. After having the discretized state space, we obtained $p_{\epsilon,j}=p(y_j|\theta_\epsilon,x_\epsilon)$ subsequently.

Step 2. Obtain the normalized filtering equation for the approximate model. When (θ,X) is replaced by $(\theta_\epsilon,X_\epsilon)$, $\mathbf{A}$ by $\mathbf{A}_\varepsilon$, $\vec{Y}$ by $\vec{Y}_\varepsilon$, p_j by $p_{\epsilon,j}$ and there also exists a probability measure P_ε to replace P, then Assumptions 2.1–2.5 also hold for $(\theta_\epsilon,X_\epsilon,\vec{Y}_\varepsilon)$. Hence, the filtering equation is obtained in equations (14) and (15) using Theorem 3.1.

The procedure hitherto brings forth $\pi_\varepsilon(f,t)$, which is the discretized approximation of $\pi(f,t)$. Let $(\mu_\epsilon,\sigma_\epsilon,\lambda_\epsilon,\rho_\epsilon,X_\epsilon)$ denote the discretized signal.

DEFINITION 4.1. *Let $\pi_{\varepsilon,t}$ be the conditional probability mass function of $(\mu_\epsilon,\sigma_\epsilon(t),\lambda_\epsilon,\rho_\epsilon,X_\epsilon(t))$ on the discrete state space given $\mathcal{F}_t^{\vec{Y}_\varepsilon}$. Let*

$$
\begin{aligned}
\pi_\varepsilon(f,t)&=E^{P_\varepsilon}[f(\mu_\epsilon,\sigma_\epsilon(t),\lambda_\epsilon,\rho_\epsilon,X_\epsilon(t))\,|\,\mathcal{F}_t^{\vec{Y}_\varepsilon}]\\
&=\sum_{\mu,\sigma,\lambda,\rho,x}f(\mu,\sigma,\lambda,\rho,x)\pi_{\varepsilon,t}(\mu,\sigma,\lambda,\rho,x),
\end{aligned}
$$

where the summation goes over all lattices in the discretized state spaces.

Step 3. Convert equations (14) and (15) to the recursive algorithm. First, we define the approximate posterior that the recursive algorithm computes.

DEFINITION 4.2. *The posterior of the approximate model at time t is denoted by*

$$\begin{aligned}&p_\varepsilon(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w;\,t)\\&\quad=\pi_{\varepsilon,t}\{\mu_\epsilon=\mu_j,\sigma_\epsilon(t)=\sigma_k,\lambda_\epsilon=\lambda_l,\rho_\epsilon=\rho_m,X_\epsilon(t)=x_w\}.\end{aligned}$$

Then, there are two substeps. The core of the first substep is to take f as the following lattice-point indicator function:

$$\begin{aligned}&\mathbf{I}_{\{\mu_\epsilon=\mu_j,\sigma_\epsilon(t)=\sigma_k,\lambda_\epsilon=\lambda_l,\rho_\epsilon=\rho_m,X_\epsilon(t)=x_w\}}(\mu_\epsilon,\sigma_\epsilon(t),\lambda_\epsilon,\rho_\epsilon,X_\epsilon(t))\\&\quad\overset{\text{def}}{=}\mathbf{I}(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w).\end{aligned}\tag{19}$$

Then, the following fact emerges:

$$\begin{aligned}&\pi_\varepsilon\Big(b(\mu_\epsilon,\sigma_\epsilon(t),X_\epsilon(t))\mathbf{I}(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w+\epsilon_x),t\Big)\\&\quad=b(\mu_j,\sigma_k,x_{w-1})p_\varepsilon(\mu_j,\sigma_k,\lambda_l,\rho_m,x_{w-1};\,t).\end{aligned}$$

Along with similar results, $\pi_\varepsilon(\mathbf{A}_\varepsilon\mathbf{I},t)$ in equation (14) becomes explicit and equation (14) becomes

$$\begin{aligned}&p_\varepsilon(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w;\,t_{i+1}-)\\&\quad=p_\varepsilon(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w;\,t_i)\\&\quad\quad+\int_{t_i}^{t_{i+1}-}\Big(b(\mu_j,\sigma_k,x_{w-1})p_\varepsilon(\mu_j,\sigma_k,\lambda_l,\rho_m,x_{w-1};\,t)\\&\quad\quad-(b(\mu_j,\sigma_k,x_w)+d(\mu_j,\sigma_k,x_w))p_\varepsilon(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w;\,t)\\&\quad\quad+d(\mu_j,\sigma_k,x_{w+1})p_\varepsilon(\mu_j,\sigma_k,\lambda_l,\rho_m,x_{w+1};\,t)\\&\quad\quad+\lambda_l\big(\bar{p}(\mu_j,\lambda_l,\rho_m,x_w;\,t)-p(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w;\,t)\big)\Big)\mathrm{d}t,\end{aligned}\tag{20}$$

where

$$\bar{p}(\mu_j,\lambda_l,\rho_m,x_w;\,t)=\frac{1}{n_\sigma+1}\sum_{i=0}^{n_\sigma}p(\mu_j,\sigma_i,\lambda_l,\rho_m,x_w;\,t).$$

If a trade at kth price level occurs at time t_{i+1}, the updating equation (15) can be written as,

$$\begin{aligned}&p_\varepsilon(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w;\,t_{i+1})\\&\quad=\frac{p_\varepsilon(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w;\,t_{i+1}-)p(y_k|x_w,\rho_m)}{\sum_{j',k',l',m',w'}p_\varepsilon(\mu_{j'},\sigma_{k'},\lambda_{l'},\rho_{m'},x_{w'};\,t_{i+1}-)p(y_k|x_{w'},\rho_{m'})},\end{aligned}\tag{21}$$

where the summation goes over the total discretized space, and $p(y_k|x_w, \rho_m)$, the transition probability from x_w to y_k, is specified by equation (A2) in Appendix A.

In the second substep, we approximate the time integral in equation (20) with an Euler scheme to obtain a recursive algorithm further discrete in time. After excluding the probability-zero event that two or more jumps occur at the same time, there are two possible cases for the inter-trading time. Case 1, if $t_{i+1} - t_i \leqslant LL$, the length controller in the Euler scheme, then we approximate $p(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w; t_{i+1}-)$ as

$$\begin{aligned}
&p(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w; t_{i+1}-)\\
&\quad\approx p(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w; t_i)\\
&\qquad+\Big[b(\mu_j, \sigma_k, x_{w-1})p(\mu_j, \sigma_k, \lambda_l, \rho_m, x_{w-1}; t_i)\\
&\qquad-\big(b(\mu_j, \sigma_k, x_w)+d(\mu_j, \sigma_k, x_w)\big)p(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w; t_i)\\
&\qquad+d(\mu_j, \sigma_k, x_{w+1})p(\mu_j, \sigma_k, \lambda_l, \rho_m, x_{w+1}; t_i)\\
&\qquad+\lambda_l\big(\bar{p}(\mu_j, \lambda_l, \rho_m, x_w; t) - p(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w; t)\big)\Big](t_{i+1} - t_i). \qquad (22)
\end{aligned}$$

Case 2, if $t_{i+1} - t_i > LL$, then we can choose a fine partition $\{t_{i,0} = t_i, t_{i,1}, \dots, t_{i,n} = t_{i+1}\}$ of $[t_i, t_{i+1}]$ such that $\max_j |t_{i,j+1} - t_{i,j}| < LL$ and then approximate $p(\mu_j, \sigma_k, \lambda_l, \rho_m, x_l; t_{i+1}-)$ by applying repeatedly equation (22) from $t_{i,0}$ to $t_{i,1}$, then $t_{i,2}, \dots,$ until $t_{i,n} = t_{i+1}$.

Equations (21) and (22) consist of the recursive algorithm we employ to calculate the approximate posterior at time t_{i+1} for $(\mu, \sigma(t_{i+1}), \lambda_l, \rho, X(t_{i+1}))$ based on the posterior at time t_i. At time t_{i+1}, the Bayes estimates of μ, $\sigma(t_{i+1})$, λ, ρ and $X(t_{i+1})$ are the means of the corresponding marginal posteriors.

Finally, we choose a reasonable prior. The Bayesian paradigm allows us to actively incorporate prior information such as researchers past experience, personal intuition, or expert opinion. Assume independence between $X(0)$ and $(\mu, \sigma(0), \lambda, \rho)$. The prior for $X(0)$ can be set by $P\{X(0) = Y(t_1)\} = 1$, where $Y(t_1)$ is the first trade price of a data set because they are very close. For other parameters, in the case that no prior information is available, we can simply assign *noninformative priors* on the parameters, namely, assign uniform distributions to the discretized state space of $(\mu, \sigma(0), \lambda, \rho)$ and obtain the prior, which is used in the rest of this section, at $t = 0$ as,

$$p(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w; 0) = \begin{cases} \frac{1}{(1+n_\mu)(1+n_\sigma)(1+n_\lambda)(1+n_\rho)} & \text{if } x_w = Y(t_1), \\ 0 & \text{otherwise.} \end{cases}$$

4.2.1. *Consistency of the Recursive Algorithm*

There are two approximations in the construction of the recursive algorithm. The second is to approximate the time integral in the propagation equation (14) by Euler scheme, whose convergence is well-known. The first, more important one is the approximation of the filtering equations (11) and (12) by equations (14) and (15) of the approximate model. Since $(\theta_\epsilon, X_\epsilon) \Rightarrow (\theta, X)$ by construction, Theorem 3.3 guarantees the weak convergence of the filtering equation of the approximate model to that of the assumed model.

4.3. A SIMULATION EXAMPLE

A Fortran program for the recursive algorithm is constructed to calculate, at each trading time t_i, the joint posterior of $(\mu, \sigma(t), \lambda, \rho, X(t))$, their marginal posteriors, their Bayes estimates and their standard errors (SE), respectively. The recursive algorithm for computing the Bayes factors is fast enough to generate real-time Bayes estimates. The algorithm is extensively tested and validated on simulated data. In the simulation, we know the true posterior. We run the Fortran program through simulated data and calculate the approximate posteriors. By comparing the approximate posteriors with the true one, we can choose the appropriate ϵ in the discretizations of the state spaces. Also simulation studies suggest that the posterior is not sensitive to the prior. Namely, for reasonable priors, as long as the range covers the true values and the lattices are reasonably fine, the Bayes estimates will converges to the true values. One simulation example is provided to demonstrate the effectiveness of Bayes estimates, especially for stochastic volatility.

For time-invariant parameters (μ, λ, ρ), they converge to their true values and the two-SE bounds become smaller and smaller, and goes to zero as in the case of GBM in Zeng [13]. Hence, only the final Bayes estimates, their SE, and true values are presented in Table II. The true values are close to the Bayes estimates and all within two SE bounds. Estimating stochastic volatility is the focus. Figure 1 shows how the Bayes estimates of volatility evolve in comparison of the time-varying true volatilities, $\sigma(t)$, for all 70,000 simulated data. Overall, we see that the Bayes estimates of volatility are close to their true values, which are within the two SE bounds most of time. The Bayes estimates can catch up with the movement, as the true volatility changes.

4.4. A REAL DATA EXAMPLE

The tested recursive algorithm is applied to a two-month (January and February, 1994, 40 business days) transaction data set of Microsoft. The

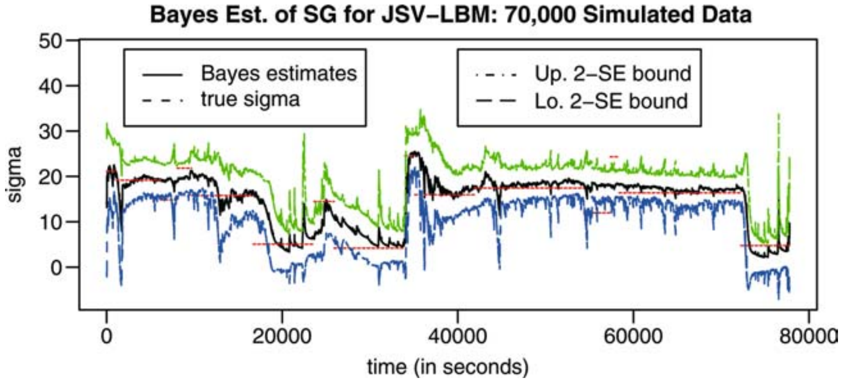

Figure 1.

Table II. Final Bayes estimates of SVJ-LBM: 70,000 simulated data.

	Final Bayes estimates				Summary of σ		
Parameters	μ	σ	λ	ρ	Mean	St.Dev.	Median
True Value	760.5	4.20	228.15	0.20	13.68	6.06	15.97
Bayes Est.	499.43	9.51	301.14	0.2017	14.81	5.62	17.22
SE	(156.14)	(7.23)	(46.89)	(0.0015)			

data are extracted from the Trade and Quote (TAQ) database distributed by NYSE. We apply standard procedures to filter the data. The final sample has 49,937 observations with a minimum \$78, a maximum \$87, and a mean \$82.38.

Based on the relative frequencies of the fractional parts of price, we use the method of relative frequencies to estimate $\alpha = 0.2414$, and $\beta = 0.3502$. The empirical distribution of trade waiting times does not support a pure exponential distribution for duration, but does support a mixed exponential distribution, which is consistent with the assumption that the total trading intensity, $a(t)$, can depend on time. The mean duration is 18.63 s with a standard error 30.01 s.

For comparison, the recursive algorithm for the simple model with straight LBM as the value process incorporated with the same noises described in Section 4.1 is applied to the same data set to obtain Bayes estimates. Assuming 260 business days per year, the annualized estimates are obtained. Final Bayes estimates of both models for the same Microsoft data are summarized in Table III. For JSV-LBM, the jump intensity, λ, is 3964.30, which means there are 3964.30 changes in volatility annually, or 15.25 times daily. It is observed that intraday volatility has "U-shaped" :

Table III. Final Bayes estimates: MSFT data.

	Final Bayes est.				Summary of σ		
Models	μ	σ	λ	ρ	Mean	St.Dev.	Median
JSV-LBM	\$16.26	\$15.20	3964.30	0.2046	\$29.07	\$ 16.44	\$24.48
	(\$32.71)	(\$12.12)	(433.74)	(0.0036)			
LBM	\$13.44	\$21.76	NA	0.2264			
	(\$53.96)	(\$0.26)	NA	(0.0023)			

higher volatility in opening and closing periods of the market. So volatility changes at least twice a day. The large λ is consistent with this observation and indicates the volatility changes even more frequent within the tradings of a day. The parameter for non-clustering noise, ρ, is reduced in JSV-LBM, probably due to more price variation is explained by stochastic volatility.

Figure 2 shows how the Bayes estimates of the volatilities for JSV-LBM and LBM evolve and demonstrate how different the volatility estimates are for the two models. In LBM, the volatility is assumed to be constant, and its estimates tend to be stable and fail to capture the movement of volatility. In JSV-GBM, the stochastic volatility feature in the Microsoft data is clearly demonstrated. This is also clearly confirmed by the empirical result according to the Bayesian model selection via filtering in Section 4.5. To avoid crowdedness in picture, Figure 3 presents only the last 1000 Bayes estimates of volatility for JSV-LBM and their estimated two-SE bounds. We can see the volatility estimates varies greatly. Sometimes it moves continuously and sometimes it jumps just as the model of suggests.

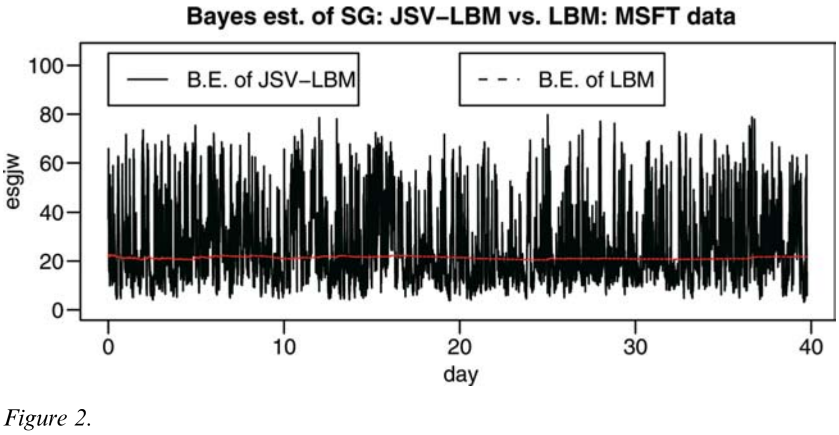

Figure 2.

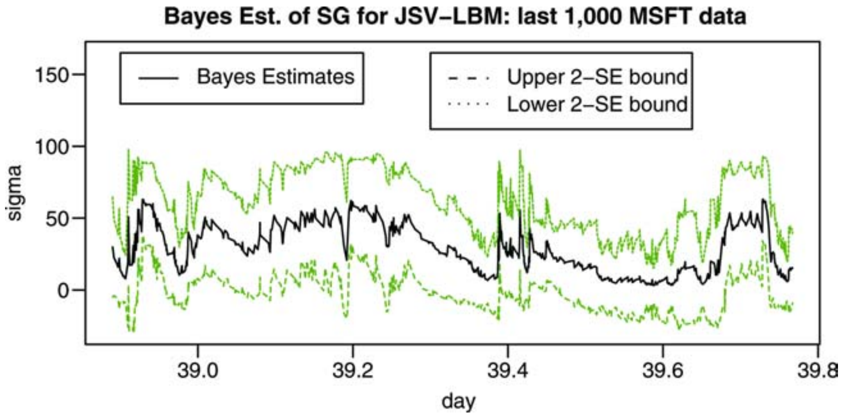

Figure 3.

4.5. A BRIEF EXAMPLE FOR BAYESIAN MODEL SELECTION

We consider a model selection problem between two micromovement models: Model 1 is just the simple micromovement model based on straight LBM incorporated with the same noises described in Section 4.1 (more details and estimation results of the model are in [15]). Model 2 is the model based on JSV-LBM. For the trading intensity functions, we assume both of them are deterministic. In the model selection setup, we take $a_1(t) = a_2(t) = a(t)$ because both models are for the same asset and the same data set. As shown in equations (9) and (10), $a(t)$ drops out in the evolution equations of Bayes factor.

Similarly to the construction of recursive algorithm for Bayes estimation, a recursive algorithm can be constructed to compute the Bayes factors of Model 2 over Model 1 and of Model 1 over Model 2. For the detail of how to construct such a recursive algorithm and the effectiveness of Bayes factor for the model selection of this class of micromovement models, interested readers are referred to [9]. The recursive algorithm is applied to the same Microsoft data set to obtain the Bayes factor of Model 2 over Model 1. Although the calculated Bayes factor of JSV-LBM vs. LBM fluctuates, it has an exponential trend of growth and is already 9.019×10^5 at the end of first day (at the 826 datum), which is much larger than 150, the decisive banchmark against Model 1. This clearly demonstrates that there are volatility changes within one trading day and is consistent with the empirical finding of higher volatility in opening and closing periods of the market sessions.

5. Conclusions

This paper highlights recent development of a rich class of counting process models for the micromovement of asset price and the Bayesian inference via filtering for the class of models. A specific micromovement model built upon JSV-LBM is used to demonstrate the procedure to develop a micromovement model with specific tick-level sample characteristics, and further demonstrate the procedure to implement Bayes estimation via filtering with simulation and real-data examples. A further important topic for research is on how to assess the model fit formally.

The applications of the class of micromovement models and its continuous-time Bayesian inference via filtering is not just mere to obtain trade-by-trade Bayes parameter estimates, nor to show stochastic volatility matters, but rather to provide a general framework for modeling price micromovement with powerful Bayesian inference. The general model provides strong potential to relate or illuminate aspects of the rich theoretical literiture on market microstructure and trading mechanism. Furthermore, Bayesian model selection via filtering provides a general, powerful tool to test related market microstructure theories, represented by the micromovement models. We may test whether NASDAQ has less trading noise after a market reform as in Barclay et al. [1] (see [11]), test whether information affects trading intensity as in Easley and O'Hara [4] and Engle [5].

Acknowledgements

The author would like to thank Mike Wiper, the guest editor for the special issue, and two anonymous referees for detailed comments and suggestions on earlier versions, improving the quality of the paper.

Appendix A: More on Modeling Clustering Noise

To formulate the biasing rule, we first define a classifying function $r(\cdot)$,

$$r(y)=\begin{cases}3 \text{ if the fractional part of y is odd eighth,}\\ 2 \text{ if the fractional part of y is odd quarter,}\\ 1 \text{ if the fractional part of y is a half or zero.}\end{cases} \tag{A1}$$

The biasing rules specify the transition probabilities from y' to y, $p(y|y')$. Then, $p(y|x)$, the transition probability can be computed through $p(y|x)=\sum_{y'} p(y|y')p(y'|x)$ where $p(y'|x)=P\{V=y'-R[x,1/8]\}$. Suppose $D=8|y-R[x,1/8]|$. Then, $p(y|x)$ can be calculated as, for example, when $r(y)=1$,

$$p(y|x)=\begin{cases}(1-\rho)(1+\beta\rho) & \text{if } r(y)=1 \text{ and } D=0,\\ \frac{1}{2}(1-\rho)[\rho+\beta(2+\rho^2)] & \text{if } r(y)=1 \text{ and } D=1,\\ \frac{1}{2}(1-\rho)\rho^{D-1}[\rho+\beta(1+\rho^2)] & \text{if } r(y)=1 \text{ and } D\geqslant 2.\end{cases} \tag{A2}$$

References

1. Barclay, M. J., Christie, W. G., Harris, J. H., Kandel, E. and Schultz, P. H.: The effects of market reform on the trading costs and depths of NASDAQ stocks, *J. Fin.* **54**(1) (1999), 1–34.
2. Bremaud, P.: *Point Processes and Queues:Martingale Dynamics*. Springer-Verlag, New York, 1981.
3. Cvitanic, J., Liptser R., and Rozovskii, B.: A filtering approach to tracking volatility from prices observed at random times, Working paper. Southern California University, 2003.
4. Easley, D. and O'Hara, M.: Time and the process of security price adjustment, *J. Fin.* **47** (1992), 577–605.
5. Engle, R.: The econometrics of ultra-high-frequency data, *Econometrica*, **68** (2000), 1–22.
6. Ethier, S. N. and Kurtz, T. G.: *Markov Processes: Characterization and Convergence*. Wiley, New York, 1986.
7. Frey, R. and Runggaldier, W.: A nonlinear filtering approach to volatility estimation with a view towards high frequency data, *Int, J. Theor. Appl. Fin.* **4** (2001), 199–210.
8. Kass, R. E. and Raftery, A. E.: Bayes factors and model uncertainty, *J. Am. Stat. Assoc.*, **90** (1995), 773–795.
9. Kouritzin, M. and Zeng, Y.: Bayesian model selection via filtering for a class of micro-movement models of asset price, To appear in *Int. J. Theor. Appl. Fin.* **8** (1) (2005), 97–122.
10. Michael A. Kouritzin and Yong Zeng: Week convergence for a type of conditional expectation: application to the inference for a class of asset price models, *Nonlinear Analysis Series A: Theory and Methods* **60** (2) (2005), 231–239.
11. Last, G. and Brandt, A.: *Marked Point Processes on the Real Line:The Dynamic Approach*. Springer, New York, 1995.
12. Spalding, R., Tsui, K. W. and Zeng, Y.: A micro-movement model with Bayes estimation via filtering: Applications to measuring trading noises costs, Working Paper, Univeristy of Missouri at Kansas City, 2004.
13. Zeng, Y.: A partially-observed model for micro-movement of asset prices with Bayes estimation via filtering. *Math. Fin.*, **13** (2003), 411–444.
14. Zeng, Y.: Estimating stochastic volatility via filtering for the micro-movement of asset prices, *IEEE Trans. Automatic Control* **49** (2004), 338–348.
15. Yong Zeng: Bayes Estimation via filtering for a simple micro-movement model of asset price with discrete noises, to appear in *Nonlinear Analysis Series A: Theory and Methods*.