International Journal of Theoretical and Applied Finance
Vol. 8, No. 1 (2005) 97–121

BAYESIAN MODEL SELECTION VIA FILTERING FOR A CLASS OF MICRO-MOVEMENT MODELS OF ASSET PRICE

MICHAEL A. KOURITZIN
Department of Mathematical and Statistical Sciences
University of Alberta, Edmonton
Alberta T6G 2G1, Canada
mkouritz@math.ualberta.ca

YONG ZENG*
Department of Mathematics and Statistics
University of Missouri at Kansas City
Kansas City, MO 64110, USA
zeng@mendota.umkc.edu

Received 2 December 2003
Accepted 8 June 2004

This paper develops the Bayesian model selection based on Bayes factor for a rich class of partially-observed micro-movement models of asset price. We focus on one recursive algorithm to calculate the Bayes factors, first deriving the system of SDEs for them and then applying the Markov chain approximation method to yield a recursive algorithm. We prove the consistency (or robustness) of the recursive algorithm. To illustrate the construction of such a recursive algorithm, we consider a model selection problem for two micro-movement models with and without stochastic volatility, and provide simulation and real-data examples to demonstrate the effectiveness of the Bayes factor in the model selection for this class of models.

Keywords: Bayes factor; counting process; filtering; Markov chain approximation method; model selection; price clustering; Transaction data.

1. Introduction

Asset price models can be classified into two broad categories: macro- and micro-movement models. Macro-movement refers to daily, weekly, and monthly closing price behavior and micro-movement refers to transactional (trade-by-trade) price behavior. There are strong connections as well as striking distinctions between the macro- and micro-movements. Zeng [12] proposed a class of partially-observed

*Corresponding author.

micro-movement models that can tie the sample characteristics of micro- and macro-movements together in a consistent manner. The most prominent feature of the class of models is that the micromovement model is characterized as a filtering problem with counting process observations. Under this representation, Zeng also developed Bayes estimation via filtering for the class of models.

Model selection, a significant and persistent area of research, evaluates which of competing models best fits transaction price data. Moreover, to the degree that economic theory can be modeled statistically, model selection provides a powerful tool for testing the economic theories related to market microstructure. The Bayesian approach provides a powerful methodology for hypothesis testing and model selection based on Bayes factor. Kass and Raftery [8] survey Bayes factor in both methodology and applications. We adopt the Bayesian approach for the model selection of the class of micro-movement models for three reasons. Firstly, the continuous-time likelihood functions of the class of models do not satisfy the common regularity conditions [5]. Thus, model selection approaches based on maximum likelihood are not justified, but Bayes factor methods are. Secondly, Bayes factors neither require the models in hypotheses testing to be nested, nor the corresponding probability measures to be absolutely continuous. This is essential to the model selection of stochastic processes, because absolute continuity is not common in the measures of stochastic processes. Finally, Kass and Raftery [8] shows that under some conditions, Bayes factor is approximately equal to BIC (Bayesian Information Criterion), which penalizes according to both the number of parameters and the number of data. This suggests that Bayes factor has the similar desirable property also.

In this paper, we introduce one two-step approach to calculate the Bayes factor. The first step is to derive the system of two stochastic differential equations (SDEs) that govern the evolution of the Bayes factors of Models 1 vs. 2 and of Models 2 vs. 1. The second step is to apply the Markov chain approximation method to the system of SDEs to develop a recursive algorithm for computing the Bayes factors. We prove a convergence theorem guaranteeing the consistency of the recursive algorithms. In one important case when the trading intensity is deterministic, this direct approach has a strong computational advantage, over the indirect approach of calculating Bayes factors as a ratio of approximated integrated likelihoods, because the trading intensity drops out of the Bayes factor. Furthermore, we do not even need to estimate the trading intensity.

To illustrate the construction of such a recursive algorithm, we consider a model selection problem for two micro-movement models of asset price. One model is built upon geometric Brownian motion (GBM) and the other upon GBM with jumping stochastic volatility (JSV-GBM). Simulation is conducted to demonstrate the effectiveness of the Bayes factor in the model selection. As a real-world application, the model selection procedure is applied to an actual Microsoft transaction data set and the Bayes factor calculated provides abundant evidence that JSV-GBM fits the data better.

The rest of the paper goes as follows. Section 2.1 reviews the methodology of Bayesian model selection and Secs. 2.2 and 2.3 review the class of micro-movement models and related results in Zeng [12]. Section 3 derives the system of SDEs for the Bayes factors and proves an important convergence theorem. Section 4 contains an illustrative model selection problem and constructs the recursive algorithm for the Bayes factors in detail. Simulation results and an actual data application are also presented in Sec. 4. We conclude in Sec. 5.

2. Review

We begin with the definition of Bayes factor [8].

2.1. *Bayes factor and its interpretation*

Suppose that data set D is generated by one of the two models: M_1 (Model 1) or M_2 (Model 2), where Model k, $k = 1, 2$, has the vector parameter, θ_k. We let $pr(D|\theta_k, M_k)$ be the likelihood function for D given the parameter θ_k, and $\pi(\theta_k|M_k)$ be the prior density of θ_k under Model k. Define the integrated (or marginal) likelihood of Model k as

$$pr(D|M_k) = \int pr(D|\theta_k, M_k)\pi(\theta_k|M_k)\, d\theta_k. \tag{2.1}$$

A Bayesian statistician uses the integrated likelihood, $pr(D|M_k)$, to measure the chance that D is generated by Model k with the prior $\pi(\theta_k|M_k)$. The Bayes factor of Model 1 over Model 2 is defined as the ratio of integrated likelihoods:

$$B_{12} = \frac{pr(D|M_1)}{pr(D|M_2)}. \tag{2.2}$$

Jeffreys [7] argued that the Bayes factor is a summary of the evidence provided by the data in favor of Model 1 over Model 2. Then, he developed a Bayes factor methodology to quantify the evidence of scientific theories, represented by statistical models. Suppose the prior opinion is that Model 1 is true with probability $pr(M_1)$, and $pr(M_2) = 1 - pr(M_1)$. The data produce posterior probabilities $pr(M_1|D)$ and $pr(M_2|D) = 1 - pr(M_1|D)$. Since any prior opinion [i.e., $pr(M_k)$] is transformed to a posterior opinion [i.e., $pr(M_k|D)$] through the consideration of the data, the transformation itself represents the evidence provided by the data. The transformation can be obtained from Bayes' Theorem:

$$pr(M_k|D) = \frac{pr(D|M_k)pr(M_k)}{pr(D|M_1)pr(M_1) + pr(D|M_2)pr(M_2)}, \quad (k = 1, 2).$$

In fact, the above transformation is used to acquire the posterior probabilities [i.e., $pr(M_k|D)$], regardless of the prior probabilities [i.e., $pr(M_k)$]. Once we transfer to the odds scale (odds = probability/(1 − probability)), the Bayes factor B_{12} appears

$$\frac{pr(M_1|D)}{pr(M_2|D)} = B_{12}\,\frac{pr(M_1)}{pr(M_2)}, \quad \text{where } B_{12} = \frac{pr(D|M_1)}{pr(D|M_2)},$$

Table 1. Interpretation of Bayes factor.

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

and this transformation takes the simple form:

$$\text{posterior odds} = \text{Bayes factor} \times \text{prior odds}.$$

Therefore, the Bayes factor can be alternatively defined as the ratio of the posterior odds of Model 1 to its prior odds, regardless of the value of the prior odds. From the above definition and arguments, we see that Bayes factors do not require the two models to be nested, nor their distributions to be absolutely continuous with respect to each other.

Once the Bayes factor is defined, the key question is how to comprehend it. Kass and Raftery [8] furnish Table 1 as the guideline for interpretation. Similarly, we can define B_{21}, the Bayes factor of Model 2 over Model 1. Obviously, $B_{12} \times B_{21} = 1$. If $B_{12} < 1$, then we can calculate B_{21} and interpret it according to the guidelines in Table 1.

2.2. *The class of micro-movement models*

The model proposed by Zeng [12] is predicated on the simple intuition that the price is formed from an intrinsic value process by incorporating the noises that arise from the trading activity. Suppose that the intrinsic value process X of an asset can not be observed directly, but can be partially observed through the price process, Y. X lives in a continuous state space while Y lives in a discrete state space given by the multiples of the minimum price variation, *a tick*, which is assumed to be $\frac{1}{M}$ for some positive integer M. The combination of (X, Y) provides a natural partially-observed framework for the micro-movement process. Prices can only be observed at irregularly spaced trading times, which are modeled by a conditional Poisson process with the trading intensity function $a(\theta(t), X(t), t)$, where θ is a vector of parameters in the model.

We assume that (θ, X, Y) is defined on a complete, filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$ and impose a mild assumption on (θ, X):

Assumption 2.1. (θ, X) *is the unique solution of a* $\mathbb{R}^{d+1}$*-valued martingale problem for a generator* $\mathbf{A}$ *and initial distribution* μ_0*, such that* $P\{(\theta(0), X(0)) \in d\theta \times dx\} = \mu_0(d\theta \times dx)$ *and*

$$M_f(t) = f(\theta(t), X(t)) - \int_0^t \mathbf{A} f(\theta(s), X(s)) ds,$$

is a $\mathcal{F}_t^{\theta,X}$-martingale for each $f \in D(\mathbf{A})$, where $\mathcal{F}_t^{\theta,X}$ is the σ-algebra generated by $(\theta(s), X(s))_{0\leq s\leq t}$ and $D(\mathbf{A})$ is the domain of $\mathbf{A}$. Specially, we assume that $D(\mathbf{A})$ contains a class of bounded, continuous functionals that separates points, is closed under multiplication and contains the constant function, and that the martingale problem admits a cadlag solution.

There are two equivalent methods to build the price model from the value process. The first constructs Y from X by incorporating noises. The second formulates (X, Y) as a filtering problem with counting process observations. The former approach is intuitive, while the latter approach is useful for statistical analysis. The equivalence of the two approaches of modeling is proven in Zeng [14].

2.2.1. *Construction of Y from X*

There are three general steps in constructing Y from X. First, we specify the value process $X(t)$. Next, one determines 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 determined by

$$Y(t_i) = F(X(t_i)),$$

where $y = F(x)$ is a random transformation with the transition probability $p(y|x)$. For example, $F(x)$ can accommodate the three important types of noise: discrete, clustering and non-clustering as shown in Sec. 4.1. Under this construction, information affects $X(t)$, the asset value, and has a permanent influence on the price while the noise affects $F(x)$, and only has a transitory impact on price.

2.2.2. *Counting process observations*

In the above construction, we view the prices in the order of trading occurrence over time. Alternatively, we can view them in terms of price levels. Namely, we view the prices as a collection of counting processes in the following form:

$$\vec{Y}(t) = \begin{pmatrix} N_1(\int_0^t \lambda_1(\theta(s), X(s), s)ds) \\ N_2(\int_0^t \lambda_2(\theta(s), X(s), s)ds) \\ \vdots \\ N_n(\int_0^t \lambda_n(\theta(s), X(s), s)ds) \end{pmatrix}, \tag{2.3}$$

where $Y_j(t) = N_j(\int_0^t \lambda_j(\theta(s), X(s), s)ds)$ 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.

Note that $Y_j(t) = N_j(\int_0^t \lambda_j(\theta(s), X(s), s)ds)$ is a conditional Poisson process with the stochastic intensity $\lambda_j(\theta(t), X(t), t)$, and $Y_j(t) - \int_0^t \lambda_j(\theta(s), X(s), s)ds$ is a martingale. We invoke four mild assumptions so that the model in Sec. 2.2.1 is equivalent in distribution to the counting process observations in Eq. (2.3). The equivalence ensures that the statistical analysis based on the latter specification can be applied to the former.

Assumption 2.2. *N_j's are unit Poisson processes under measure P.*

Assumption 2.3. *$(\theta, X), N_1, N_2, \ldots, N_n$ are independent under measure P.*

Assumption 2.4. *There exist a constant, C, such that $0 < a(\theta, x, t) \leq C$ for all θ, x and $t > 0$.*

Assumption 2.5. *The intensity, $\lambda_j(\theta, x, t) = a(\theta, x, t)p(y_j|x)$, where $a(\theta, x, t)$ is the total intensity at time t and $p(y_j|x)$ is the transition probability from x to y_j, the jth price level, as in $F(x)$.*

Under this representation, $(\theta(t), X(t))$ becomes the unobserved signal process, and $\vec{Y}(t)$ becomes the observation process corrupted by noise, which is modeled by $p(y|x)$. Hence, $(\theta, X, \vec{Y})$ is formulated as a *filtering problem with counting process observations.* The class of models presented here are closely related to the models in [6] and [2], but their models do not include microstructure noises.

2.3. *Related theoretical results*

We review the continuous-time integrated likelihood for the micro-movement model, and the related filtering equation from Zeng [12].

Recall that (θ, X, Y) is defined on a complete, filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P)$. Assumptions 2.2–2.4 imply that there exists a reference measure Q such that P is absolutely continuous with respect to Q. Under Q, the counting processes, $(Y_1, \ldots, Y_n)$, are n independent unit Poisson processes, and they are independent of (θ, X). Furthermore, $Q = P_{\theta,x} \times Q_{\vec{y}}$, where $P_{\theta,x}$ is the probability measure for (θ, X) and $Q_{\vec{y}}$ is the probability measure for the n independent unit Poisson processes. Suppose $P_{y|\theta,x}$ is the conditional probability measure of Y given (θ, X) under P. Then, the Radon–Nikodym derivative of the model is

$$\begin{aligned} L(t) &= L\big((\theta(s), X(s), Y(s))_{0 \leq s \leq t}\big) = \frac{dP}{dQ}(t) \\ &= \frac{dP_{\theta,x}}{dP_{\theta,x}}(t) \times \frac{dP_{y|\theta,x}}{dQ_{\vec{y}}}(t) = \frac{dP_{y|\theta,x}}{dQ_{\vec{y}}}(t) \\ &= \prod_{j=1}^{n} \exp\bigg\{ \int_0^t \log \lambda_j(\theta(s-), X(s-), s-) dY_j(s) \\ &\qquad - \int_0^t \Big[\lambda_j(\theta(s), X(s), s) - 1\Big] ds \bigg\}. \end{aligned} \tag{2.4}$$

Note that $L(t)$ is the joint likelihood of $(\theta, X, \vec{Y})$. To obtain the integrated likelihood of the model, or the marginal likelihood of $\vec{Y}$, we may integrate $L(t)$ on (θ, X), or equivalently, write the integrated likelihood in terms of conditional expectation.

Definition 2.1. Let $\mathcal{F}_t^{\vec{Y}} = \sigma\{(\vec{Y}(s)) | 0 \leq s \leq t\}$ and $\phi(f, t) = E^Q[f(\theta(t), X(t))L(t)|\mathcal{F}_t^{\vec{Y}}]$.

Then, $\mathcal{F}_t^{\vec{Y}}$ is all the available information up to time t. For a prior on $(\theta(0), X(0))$, the integrated likelihood of $\vec{Y}$ is $E^Q[L(t) \,|\, \mathcal{F}^{\vec{Y}}] = \phi(1,t)$, which corresponds to $P(D|M_k)$ in Sec. 2.1. Note that $\phi(1,t)$ is uniquely characterized by the unnormalized filtering equation, presented in Theorem 2.1.

Theorem 2.1. *Suppose that $(\theta, X, \vec{Y})$ satisfies Assumptions 2.1–2.5. Then, $\phi(f,t)$ is the unique solution of the following SDE, called the unnormalized filtering equation,*

$$\begin{aligned}\phi(f,t) = \phi(f,0) &+ \int_0^t \phi(\mathbf{A}f - (a-n)f, s)ds \\ &+ \sum_{j=1}^{n} \int_0^t \phi((ap_j - 1)f, s-)dY_j(s),\end{aligned} \tag{2.5}$$

for every $t > 0$ and $f \in D(\mathbf{A})$, where $a = a(\theta(t), X(t), t)$, is the trading intensity, and $p_j = p(y_j|x)$ is the transition probability from x to y_j, the jth price level.

For a proof of Theorem 2.1, we refer the reader to Appendix A of Zeng [12].

3. Model Selection for the Class of Models

In this section, we study the model selection problem for the class of models reviewed in Sec. 2.2. The data is a sequence of pairs $\{(Y(t_i), t_i)\}$, where t_i is the ith trading time and $Y(t_i)$ is the price. Alternatively, the data is viewed as $\vec{Y}(t)$. To decide which of the two models better fits the data using Bayesian methodology, we calculate the Bayes factor and then interpret it accordingly.

The first approach to using Bayes factors is to calculate the integrated likelihood of each model and then take the ratio. As shown in Theorem 2.1, the integrated likelihood is characterized by the unnormalized filtering equation. Thus, we may apply the Markov chain approximation method, similar to the case of Bayes estimation via filtering in Zeng [12], to the unnormalized filtering equation to construct a recursive algorithm to calculate the integrated likelihood of each model. However, this approach is not always computational efficient as shown in an important case to be studied in sequel.

In this paper, we concentrate on an alternative two-step approach. The first step is to derive the system of two SDEs governing the evolution of the Bayes factors and the second step is to develop a consistent recursive algorithm to compute the Bayes factors directly.

3.1. *The evolution of Bayes factors*

Suppose Model k is denoted by $(\theta^{(k)}, X^{(k)}, \vec{Y}^{(k)})$ for $k = 1, 2$. Using the notation in Sec. 2.3, we denote the joint likelihood of $(\theta^{(k)}, X^{(k)}, \vec{Y}^{(k)})$ by $L^{(k)}(t)$, which is given by Eq. (2.4). Let $\phi_k(f_k, t) = E^{Q^{(k)}}\left[f_k(\theta^{(k)}(t), X^{(k)}(t))L^{(k)}(t)|\mathcal{F}_t^{\vec{Y}^{(k)}}\right]$. Then, the integrated likelihood of $\vec{Y}$ is $\phi_k(1,t)$.

Definition 3.1. The filter ratio processes are defined as:

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

Remark 3.1. We observe that the Bayes factors, $B_{12}(t) = q_1(1,t)$ and $B_{21}(t) = q_2(1,t)$.

Remark 3.2. Let $\pi_t^{(k)}$ be the conditional distribution of $(\theta^{(k)}(t), X^{(k)}(t))$ given $\mathcal{F}_t^{\vec{Y}^{(k)}}$. Then, Bayes Theorem implies that $\frac{\phi_k(f_k,t)}{\phi_k(1,t)} = \int f_k(\theta,x)\pi_t^{(k)}(d\theta,dx) = E^P\left[f_k(\theta(t),X(t))|\mathcal{F}_t^{\vec{Y}^{(k)}}\right]$. Now, when the measure ϕ_1 is normalized by $\phi_2(1,t)$, the total measure of ϕ_2 at time t, a random (conditional) finite measure is obtained and denoted by $q_t^{(1)}$. Similarly, $q_t^{(2)}$ is defined. Then, for $k = 1,2$, $q_k(f_k,t)$ can be written in the integral form as

$$q_k(f_k,t) = \int f_k(\theta^{(k)}, x^{(k)}) q_t^{(k)}(d\theta^{(k)}, dx^{(k)}). \tag{3.1}$$

The integral form of $q_k(f_k,t)$ is important in deriving the recursive algorithm.

Now, we present the theorem characterizing the evolution of $q_t^{(k)}$.

Theorem 3.1. *Suppose Model k ($k = 1,2$) has generator:* $\mathbf{A}^{(k)}$ *for* $(\theta^{(k)}, X^{(k)})$, *trading intensity* $a_k = a_k(\theta^{(k)}(t), X^{(k)}(t), \vec{Y}^{(k)}(t))$, *and transition probability* $p_j^{(k)} = p^{(k)}(y_j|x)$ *from* x *to* y_j *for the random transformation* $F^{(k)}$. *Suppose that Model k ($k = 1,2$) satisfies Assumptions* 2.1 *to* 2.5. *Then,* $\left(q_t^{(1)}, q_t^{(2)}\right)$ *are the unique measure-valued pair solution of the following system of SDEs,*

$$\begin{aligned} q_1(f_1,t) &= q_1(f_1,0) + \int_0^t \left[q_1\left(\mathbf{A}^{(1)} f_1, s\right) - q_1(a_1 f_1, s) + \frac{q_1(f_1,s) q_2(a_2,s)}{q_2(1,s)} \right] ds \\ &\quad + \sum_{j=1}^n \int_0^t \left[\frac{q_1\left(f_1 a_1 p_j^{(1)}, s-\right)}{q_2\left(a_2 p_j^{(2)}, s-\right)} q_2(1, s-) - q_1(f_1, s-) \right] dY_j(s), \end{aligned} \tag{3.2}$$

for all $t > 0$ and $f_1 \in D(\mathbf{A^{(1)}})$, and for all $t > 0$ and $f_2 \in D(\mathbf{A^{(2)}})$,

$$\begin{aligned} q_2(f_2,t) &= q_2(f_2,0) + \int_0^t \left[q_2(\mathbf{A}^{(2)} f_2, s) - q_2(a_2 f_2, s) + \frac{q_2(f_2,s) q_1(a_1,s)}{q_1(1,s)} \right] ds \\ &\quad + \sum_{j=1}^n \int_0^t \left[\frac{q_2\left(f_2 a_2 p_j^{(2)}, s-\right)}{q_1\left(a_1 p_j^{(1)}, s-\right)} q_1(1, s-) - q_2(f_2, s-) \right] dY_j(s). \end{aligned} \tag{3.3}$$

When $a_k(\theta^{(k)}(t), X^{(k)}(t), t) = a(t)$, $k = 1,2$, the above two equations becomes

$$\begin{aligned} q_1(f_1,t) &= q_1(f_1,0) + \int_0^t q_1(\mathbf{A}^{(1)} f_1, s) ds \\ &\quad + \sum_{j=1}^n \int_0^t \left[\frac{q_1\left(f_1 p_j^{(1)}, s-\right)}{q_2\left(p_j^{(2)}, s-\right)} q_2(1, s-) - q_1(f_1, s-) \right] dY_j(s), \end{aligned} \tag{3.4}$$

$$q_2(f_2,t) = q_2(f_2,0) + \int_0^t q_2(\mathbf{A}^{(2)} f_2, s)ds$$
$$+ \sum_{j=1}^{n} \int_0^t \left[\frac{q_2\big(f_2 p_j^{(2)}, s-\big)}{q_1\big(p_j^{(1)}, s-\big)} q_1(1, s-) - q_2(f_2, s-) \right] dY_j(s). \quad (3.5)$$

Proof. Applying Itô's formula for semimartingale, recalling $\phi_k(f_k,t)$ satisfies Eq. (2.5) for $k=1,2$ and simplifying, we have

$$\frac{\phi_1(f_1,t)}{\phi_2(1,t)} = \frac{\phi_1(f_1,0)}{\phi_2(1,0)} + \int_0^t \left[\frac{\phi_1(\mathbf{A}^{(1)} f_1, s)}{\phi_2(1,s)} - \frac{\phi_1(a_1 f_1, s)}{\phi_2(1,s)} + \frac{\phi_1(f_1,s)\phi_2(a_2,s)}{\phi_2^2(1,s)} \right] ds$$
$$\times \sum_{j=1}^{n} \int_0^t \left[\frac{\phi_1(f_1,s)}{\phi_2(1,s)} - \frac{\phi_1(f_1,s-)}{\phi_2(1,s-)} \right] dY_j(s). \quad (3.6)$$

We make two observations: The first is that

$$\frac{\phi_1(f_1,s)\phi_2(a_2,s)}{\phi_2^2(1,s)} = \frac{\frac{\phi_1(f_1,s)}{\phi_2(1,s)} \frac{\phi_2(a_2,s)}{\phi_1(1,s)}}{\frac{\phi_2(1,s)}{\phi_1(1,s)}} = \frac{q_1(f_1,s)q_2(a_2,s)}{q_2(1,s)}.$$

Next, we show that the integrand of the last integral in Eq. (3.6) is predictable. Indeed, assuming that a trade at Y_j occurs at time s, one has by Eq. (2.5) that

$$\frac{\phi_1(f_1,s)}{\phi_2(1,s)} = \frac{\phi_1(f_1,s-) + \phi_1\big(\big(a_1 p_j^{(1)} - 1\big) f_1, s-\big)}{\phi_2(1,s-) + \phi_2\big(a_2 p_j^{(2)} - 1, s-\big)} = \frac{\phi_1\big(f_1 a_1 p_j^{(1)}, s-\big)}{\phi_2\big(a_2 p_j^{(2)}, s-\big)}$$
$$= \frac{q_1\big(f_1 a_1 p_j^{(1)}, s-\big)}{q_2\big(a_2 p_j^{(2)}, s-\big)} q_2(1, s-).$$

Then, Eq. (3.6) implies Eq. (3.2).

When both models have the common trading intensity $a(t) = a_1 = a_2$, Eq. (3.2) clearly is simplified to Eq. (3.4).

Finally, the uniqueness of Eqs. (3.2) and (3.3) comes from the following lemma. □

Lemma 3.1. *Suppose that there are two pairs of cadlag measure-valued processes* (q_1^m, q_2^m), $m = 1,2$, *satisfying the assumptions of Theorem* 3.1 *and Eqs.* (3.2) *and* (3.3). *Then,* $q_1^1 = q_1^2$ *and* $q_2^1 = q_2^2$.

The proof of Lemma 3.1 is provided in Appendix A.

Remark 3.3. Note that $a(t)$ disappears in Eqs. (3.4) and (3.5). This reduces the calculations greatly in computing the Bayes factors. Hence, this convenient case is studied in detail in this paper. 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.

Remark 3.4. Suppose the trading times are $t_1, t_2, \ldots$. Then, Eqs. (3.4) and (3.5) 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 equations have no random component and are written for $k = 1, 2$ as

$$q_k(f_k, t_{i+1}-) = q_k(f_k, t_i) + \int_{t_i}^{t_{i+1}-} q_k(\mathbf{A}^{(k)} f_k, s)ds.$$

This implies that when there are no trades, the Bayes factors evolve deterministically.

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

$$q_1(f_1, t_{i+1}) = \frac{q_1\big(f_1 p_j^{(1)}, t_{i+1}-\big)}{q_2\big(p_j^{(2)}, t_{i+1}-\big)} q_2(1, t_{i+1}-),$$

and

$$q_2(f_2, t_{i+1}) = \frac{q_2\big(f_2 p_j^{(2)}, t_{i+1}-\big)}{q_1\big(p_j^{(1)}, t_{i+1}-\big)} q_1(1, t_{i+1}-).$$

They are random because the price level is random.

3.2. *A convergence theorem*

Theorem 3.1 provides the evolution of the Bayes factors. To compute the Bayes factors, one constructs an algorithm to approximate $q_k(f_k, t)$, where $q_1(1, t) = B_{12}(t)$. The algorithm, based on the evolution of SDEs, is naturally recursive, handling a datum at a time. Thus, the algorithm makes real-time updates and can handle large data sets.

One basic requirement for the recursive algorithm is consistency: The approximate q_k, computed by the recursive algorithm, must converge to the true one. The following theorem provides not only the theoretical foundation for consistency, but also a recipe for constructing consistent recursive algorithms.

Let $\big(\theta_\epsilon^{(k)}, X_\epsilon^{(k)}\big)$ be an approximation of $\big(\theta^{(k)}, X^{(k)}\big)$, namely, $\big(\theta_\epsilon^{(k)}, X_\epsilon^{(k)}\big)$ converges weakly to (X, θ). Then, we define

$$\vec{Y}_\epsilon^{(k)}(t) = \begin{pmatrix} N_1\big(\int_0^t \lambda_1\big(\theta_\epsilon^{(k)}(s), X_\epsilon^{(k)}(s), s\big)ds\big) \\ N_2\big(\int_0^t \lambda_2\big(\theta_\epsilon^{(k)}(s), X_\epsilon^{(k)}(s), s\big)ds\big) \\ \vdots \\ N_n\big(\int_0^t \lambda_n\big(\theta_\epsilon^{(k)}(s), X_\epsilon^{(k)}(s), s\big)ds\big) \end{pmatrix}, \tag{3.7}$$

set $\mathcal{F}_t^{\vec{Y}_\epsilon^{(k)}} = \sigma\big(\vec{Y}_\epsilon^{(k)}(s),\ 0 \le s \le t\big)$, take $L_\epsilon^{(k)}(t) = L\Big(\big(\theta_\epsilon^{(k)}(s),\ X_\epsilon^{(k)}(s),\ Y_\epsilon^{(k)}(s)\big)_{0\le s\le t}\Big)$, 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 $\left(\theta_\epsilon^{(k)}, X_\epsilon^{(k)}, \vec{Y}_\epsilon^{(k)}\right)$ lives on $\left(\Omega_\epsilon^{(k)}, \mathcal{F}_\epsilon^{(k)}, P_\epsilon^{(k)}\right)$, with Assumptions 2.1–2.5 holding. Then, there also exists a reference measure $Q_\epsilon^{(k)}$ with similar properties. Before we present the convergence theorem, we define the approximations of $\phi_k(f_k, t)$ and $q_k(f_k, t)$ for $k = 1, 2$.

Definition 3.2. For $k = 1, 2$, let $\phi_{\epsilon,k}(f_k, t) = E^{Q_\epsilon^{(k)}}\left[f_k(\theta_\epsilon^{(k)}(t), X_{\epsilon_x}^{(k)}(t))L_\epsilon^{(k)}(t) | \mathcal{F}_t^{\vec{Y}_\epsilon^{(k)}}\right]$, $q_{\epsilon,1}(f_1, t) = \phi_{\epsilon,1}(f_1, t)/\phi_{\epsilon,2}(1, t)$ and $q_{\epsilon,2}(f_2, t) = \phi_{\epsilon,2}(f_2, t)/\phi_{\epsilon,1}(1, t)$.

Remark 3.5. Suppose $\phi_{\epsilon,t}^{(k)}(\theta, x)$ $\left(q_{\epsilon,t}^{(k)}(\theta, x)\right)$ is the measure corresponding to $\phi_{\epsilon,k}$ (f_k, t) $\left(q_{\epsilon,k}(f_k, t)\right)$. We observe that, for $k = 1$, 2,

$$q_{\epsilon,k}(f_k, t) = \frac{\phi_{\epsilon,k}(f_k, t)}{\phi_{\epsilon,3-k}(1, t)} = \frac{\sum_{\theta,x} f_k(\theta, x)\phi_{\epsilon,t}^{(k)}(\theta, x)}{\phi_{\epsilon,3-k}(1, t)} = \sum_{\theta,x} f_k(\theta, x) q_{\epsilon,t}^{(k)}(\theta, x),$$

and use $q_{\epsilon,k}(f_k, t) = \sum_{\theta,x} f_k(\theta, x) q_{\epsilon,t}^{(k)}(\theta, x)$ in the construction of recursive algorithm.

Theorem 3.2. *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* $\left(\theta_\epsilon^{(k)}, X_\epsilon^{(k)}, \vec{Y}_\epsilon^{(k)}\right)_{k=1,2}$. *Suppose* $\left(\theta_\epsilon^{(k)}, X_\epsilon^{(k)}\right) \Rightarrow (\theta^{(k)}, X^{(k)})$ *as* $\epsilon \to 0$. *Then, as* $\epsilon \to 0$, *for continuous and bounded* f_1 *and* f_2,

(i) $\phi_{\epsilon,k}(f_k, t) \Rightarrow \phi_k(f_k, t)$, *for* $k = 1, 2$;
(ii) $q_{\epsilon,1}(f_1, t) \Rightarrow q_1(f_1, t)$ *and* $q_{\epsilon,2}(f_2, t) \Rightarrow q_2(f_2, t)$ *simultaneously.*

Remark 3.6. Part (i) implies the consistency of the integrated likelihood and part (ii) implies the consistency of Bayes factors.

The proof relies on *Kurtz and Protter's Theorem* on the Convergence of Stochastic Integral (see Theorem 2.2 of [10]) as well as the following lemma.

Lemma 3.2. *Suppose that* S_1 *and* S_2 *are complete, separable metric spaces and that* (X^N, Y^N), $N = 1, 2, \ldots,$ *and* (X, Y) *are* $S_1 \times S_2$*-valued random variables defined on the probability spaces* $(\Omega^N, \mathcal{F}^N, P^N)$ *and* $(\Omega, \mathcal{F}, P)$, *respectively. Suppose that* $\{(X^N, Y^N)\}$ *converges in distribution to* (X, Y), *that* $P^N \ll Q^N$ *on* $\sigma(X^N, Y^N)$ *with* $dP^N/dQ^N = L^N(X^N, Y^N)$, *and that* $Q^N(Q)$ *is a probability measure on* $\mathcal{F}^N(\mathcal{F})$ *such that* X^N, Y^N (X, Y) *are independent under* $Q^N(Q)$.

Suppose that the Q^N*-distribution of* $(X^N, Y^N, L^N(X^N, Y^N))$ *converges weakly to the* Q*-distribution of* $(X, Y, L(X, Y))$, *where* $E^Q[L(X, Y)] = 1$. *Then, the following hold*:

(i) $P \ll Q$ *on* $\sigma(X, Y)$ *and* $dP/dQ = L(X, Y)$;
(ii) *For every bounded continuous function* F: $S_1 \to R$, $E^{Q^N}[F(X^N) L^N(X^N, Y^N)|Y^N]$ *converges weakly to* $E^Q[F(X)L(X, Y)|Y]$ *as* $N \to \infty$.

Lemma 3.2 is proved in Theorem 1 in [9].

Proof of Theorem 3.2. (i) Since this part only involves one model, we exclude the superscript, (k), which identifies the model.

The assumptions for $(\theta_\epsilon, X_\epsilon, \vec{Y}_\epsilon)$ imply that there exists a reference measure Q_ϵ such that under Q_ϵ, $\{Y_{\epsilon,j}\}_{j=1}^n$ with $Y_{\epsilon,j}$ being the jth component of $\vec{Y}_\epsilon$, are independent unit Poisson processes, and they are independent of $(\theta_\epsilon, X_\epsilon)$. Similarly, under Q, $\vec{Y}$ is a collection of independent unit Poisson processes, and they are independent of (θ, X). The Radon–Nikodym derivative dP_ϵ/dQ_ϵ is

$$L_\epsilon(t) = \prod_{j=1}^n \exp\left\{ \int_0^t \log \lambda_j(\theta_\epsilon(s-), X_\epsilon(s-), s-) dY_{\epsilon,j}(s) \right. \\ \left. - \int_0^t [\lambda_j(\theta_\epsilon(s), X_\epsilon(s), s) - 1] ds \right\}.$$

Noting $(\theta_\epsilon, X_\epsilon, \vec{Y}_\epsilon) \Rightarrow (\theta, X, \vec{Y})$ under the reference measures, one finds that the *Continuous Mapping theorem* as well as *Kurtz and Protter's Theorem* imply that

$$\int_0^t \log \lambda_j(\theta_\epsilon(s-), X_\epsilon(s-), s-) dY_{\epsilon,j}(s) \Rightarrow \int_0^t \log \lambda_j(\theta(s-), X(s-), s-) dY_k(s), \\ \int_0^t [\lambda_j(\theta_\epsilon(s), X_\epsilon(s), s) - 1] ds \Rightarrow \int_0^t [\lambda_j(\theta(s), X(s), s) - 1] ds.$$

and $((\theta_\epsilon, X_\epsilon), \vec{Y}_\epsilon, L_\epsilon) \Rightarrow ((\theta, X), \vec{Y}, L)$ under the reference measures (Condition C2.2(i) of Kurtz and Protter [10] holds in our case, see Example 3.3 there.). So, Lemma 3.2 implies part (i).

Part (ii) comes from part (i), the unique (c.f. Theorem 3.1) representation $q_{\varepsilon,1} = \frac{\phi_{\varepsilon,1}(f_1,t)}{\phi_{\varepsilon,2}(1,t)}$, and the *Continuous Mapping Theorem.* □

Remark 3.7. This theorem provides a three-step recipe for constructing a consistent recursive algorithm to compute the Bayes factors. Step 1 is to construct $\left(\theta_\epsilon^{(k)}, X_\epsilon^{(k)}\right)$, the Markov chain approximation to $(\theta^{(k)}, X^{(k)})$ with generator $\mathbf{A}_\epsilon^{(k)}$, and obtain $p_{\epsilon,j}^{(k)} = p^{(k)}(y_j|\theta_\epsilon, x_\epsilon)$ as an approximation to $p_j^{(k)} = p^{(k)}(y_j|\theta, x)$, where $(\theta_\epsilon, x_\epsilon)$ is restricted to the state space of $(\theta_\epsilon^{(k)}, X_\epsilon^{(k)})$, for $k = 1, 2$. Step 2 is to obtain the evolution equations for $q_{\epsilon,k}(f_k, t)$, $k = 1, 2$, the filter ratio processes, using Theorem 3.1. For simplicity, we only consider the case when $a_1 = a_2 = a(t)$. Similar to Remark 3.4, the evolution equations can be separated into the propagation equations, for $k = 1, 2$,

$$q_{\epsilon,k}(f_k, t_{i+1}-) = q_{\epsilon,k}(f_k, t_i) + \int_{t_i}^{t_{i+1}-} q_{\epsilon,k}(\mathbf{A}_\epsilon^{(k)} f_k, s) ds, \tag{3.8}$$

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

$$\begin{aligned} q_{\epsilon,1}(f_1, t_{i+1}) &= \frac{q_{\epsilon,1}\left(f_1 p_{\epsilon,j}^{(1)}, t_{i+1}-\right)}{q_{\epsilon,2}\left(p_{\epsilon,j}^{(2)}, t_{i+1}-\right)} q_{\epsilon,2}(1, t_{i+1}-), \\ q_{\epsilon,2}(f_2, t_{i+1}) &= \frac{q_{\epsilon,2}\left(f_2 p_{\epsilon,j}^{(2)}, t_{i+1}-\right)}{q_{\epsilon,1}\left(p_{\epsilon,j}^{(1)}, t_{i+1}-\right)} q_{\epsilon,1}(2, t_{i+1}-). \end{aligned} \tag{3.9}$$

Here, f_k is a test function on the state space of $\left(\theta_\epsilon^{(k)}, X_\epsilon^{(k)}\right)$ and is in the domain of $\mathbf{A}_\epsilon^{(k)}$, for $k = 1, 2$. Step 3 converts Eqs. (3.8) and (3.9) to the recursive algorithm in discrete time by two substeps: (a) Represent $q_{\epsilon,k}(\cdot, t)$ as a finite array with components being $q_{\epsilon,k}(f_k, t)$ for lattice-point indicator f_k and (b) approximate the time integral in (3.8) with an Euler scheme.

4. An Example with Simulation and Application

To illustrate the Bayesian model selection procedure, we consider a model selection problem between two micromovement models: one's value process is GBM and the other's is jumping stochastic volatility GBM (JSV-GBM). The first model is chosen because GBM is the standard model for the first approximation of asset price. The second model is chosen not only because it is built on GBM with the feature of stochastic volatility, but also because it has an advantage of showing the effectiveness of Bayes factor for model selection, which is demonstrated in Sec. 4.3.

4.1. *Two micromovement models*

Recall the construction of a micromovement model has three steps. Firstly, we specify the value process. For Model 1, $X^{(1)}(t)$ is GBM, which in SDE form is

$$\frac{dX^{(1)}(t)}{X^{(1)}(t)} = \mu^{(1)} dt + \sigma^{(1)} dW(t)\,, \tag{4.1}$$

where $W(t)$ is a standard Brownian motion. Its generator is

$$\mathbf{A}^{(1)} f_1(x^{(1)}) = \mu^{(1)} x^{(1)} \frac{\partial f}{\partial x^{(1)}}(x^{(1)}) + \frac{1}{2}(\sigma^{(1)})^2 (x^{(1)})^2 \frac{\partial^2 f}{\partial (x^{(1)})^2}(x^{(1)}). \tag{4.2}$$

Model 2's value process is JSV-GBM, given by

$$\frac{dX^{(2)}(t)}{X^{(2)}(t)} = \mu^{(2)} dt + \sigma^{(2)}(t)\, dW(t), \quad d\sigma^{(2)}(t) = (J_{N(t-)+1} - \sigma^{(2)}(t-))\, dN(t), \tag{4.3}$$

where $N(t)$ 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, $[\alpha_\sigma^{(2)}, \beta_\sigma^{(2)}]$. Suppose that the ith Poisson event happens at time t_i and J_i is drawn. Then, the volatility changes from $\sigma^{(2)}(t_{i-1})$ to J_i at time t_i, and then remains the same until the next Poisson event occurs. The generator of Model 2 is

$$\begin{aligned}\mathbf{A}^{(2)} f_2(\sigma^{(2)}, x^{(2)}) &= \mu^{(2)} x^{(2)} \frac{\partial f_2}{\partial x^{(2)}}(\sigma^{(2)}, x^{(2)}) + \frac{1}{2}(\sigma^{(2)})^2 (x^{(2)})^2 \frac{\partial^2 f_2}{\partial (x^{(2)})^2}(\sigma^{(2)}, x^{(2)}) \\ &\quad + \lambda \int_{\alpha_\sigma^{(2)}}^{\beta_\sigma^{(2)}} \left(f_2(z, x^{(2)}) - f_2(\sigma^{(2)}, x^{(2)})\right) \frac{1}{\beta_\sigma^{(2)} - \alpha_\sigma^{(2)}}\, dz. \end{aligned} \tag{4.4}$$

Secondly, we assume both of the trading intensity functions 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 same data set. As shown in Eqs. (3.4) and (3.5), $a(t)$ drops out in the evolution equations of 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. The three important types of noise: discrete, clustering, and non-clustering, are to be incorporated (for reasons see [12]). For simplicity, we assume both models have the same noises. To simplify notation, at a trading time t_i, set $x = X(t_i)$, $y = Y(t_i)$, and $y' = Y'(t_i) = R[X(t_i) + V_i, \frac{1}{M}]$, 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:

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

Step 2. Incorporate discrete noise by rounding off x' to its closest tick, $y' = R[x', \frac{1}{M}] = R[x + V, \frac{1}{M}]$.

Step 3. Incorporate clustering noise by biasing y' through a random biasing function $b(\cdot)$. $\{b_i(\cdot)\}$ are assumed independent of $\{y'_i\}$ and serially independent. To be consistent with the data analyzed in Sec. 4.4, we construct a simple random biasing function only for the tick of 1/8 dollar. We can generalize it to other ticks such as 1/100 dollar easily. 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 rule: 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, $F(x)$, which models the impact of financial noise, is specified. The detail of $b(\cdot)$, and the explicit transition probability $p(y|x)$ of F can be found in Appendix A of [13].

The parameters of clustering noise, α and β, can be estimated through the method of relative frequency. The other parameters are estimated by Bayes estimation via filtering. The Bayes estimation for $(\mu_1, \sigma_1, \rho_1)$ of Model 1 is studied in [13] and that for $(\mu_2, \sigma, \lambda, \rho_1)$ of Model 2 is in [13].

4.2. *The recursive algorithm for Bayes factors*

Similar to Bayes estimation via filtering, the recursive algorithm for Bayes factors is constructed through the Markov chain approximation method. Following the three-step recipe provided by Theorem 3.2, we will construct a consistent recursive algorithm to calculate the Bayes factors. Since Model 2 is more complicated than Model 1, we only provide the detail construction related to Model 2. The corresponding part related to Model 1 can be derived similarly.

Step 1. Construct Markov chains, $\left(\theta_\epsilon^{(k)}(t), X_\epsilon^{(k)}(t)\right)$ as approximations to $(\theta^{(k)}(t), X^{(k)}(t))$, for $k = 1, 2$. Here, $\theta^{(1)}(t) = (\mu_1, \sigma_1, \rho_1)$ and $\theta^{(2)}(t) = (\mu_2, \sigma(t), \lambda, \rho_2)$.

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

Second, we observe that the construction of a Markov chain approximation can be achieved by constructing a Markov chain with generator, $\mathbf{A}_\varepsilon^{(2)}$, such that $\mathbf{A}_\varepsilon^{(2)} \to \mathbf{A}^{(2)}$ as $\varepsilon^{(2)} \to 0$. Accommodating other parameters, the generator of Model 2 becomes

$$\begin{aligned}
&\mathbf{A}^{(2)} f_2(\mu^{(2)}, \sigma^{(2)}, \lambda, \rho^{(2)}, x^{(2)}) \\
&\quad = \mu^{(2)} x^{(2)} \frac{\partial f_2}{\partial x^{(2)}}(\mu^{(2)}, \sigma^{(2)}, \lambda, \rho^{(2)}, x^{(2)}) + \frac{1}{2}(\sigma^{(2)})^2 (x^{(2)})^2 \frac{\partial^2 f_2}{\partial (x^{(2)})^2}(\mu^{(2)}, \sigma^{(2)}, \lambda, \rho^{(2)}, x^{(2)}) \\
&\qquad + \lambda \int_{\alpha_\sigma^{(2)}}^{\beta_\sigma^{(2)}} \left(f_2(\mu^{(2)}, z, \lambda, \rho^{(2)}, x^{(2)}) - f(\mu^{(2)}, \sigma^{(2)}, \lambda, \rho^{(2)}, x^{(2)})\right) \frac{1}{\beta_\sigma^{(2)} - \alpha_\sigma^{(2)}}\, dz. \qquad (4.5)
\end{aligned}$$

The diffusion part involves first- and second-order differentiation and the jump part involves integration. The finite difference approximation is applied for differentiation and the rectangle approximation for integration. One constructs $\mathbf{A}_\varepsilon^{(2)}$ as follows:

$$\begin{aligned}
&\mathbf{A}_\varepsilon^{(2)} f_2\left(\mu_j^{(2)}, \sigma_k^{(2)}, \lambda_l, \rho_m^{(2)}, x_w^{(2)}\right) \\
&= \mu_j^{(2)} x_w^{(2)} \left(\frac{f_2\left(\mu_j^{(2)}, \sigma_k^{(2)}, \lambda_l, \rho_m^{(2)}, x_w^{(2)} + \epsilon_x^{(2)}\right) - f_2\left(\mu_j^{(2)}, \sigma_k^{(2)}, \lambda_l, \rho_m^{(2)}, x_w^{(2)} - \epsilon_x^{(2)}\right)}{2\epsilon_x^{(2)}} \right) \\
&\quad + \frac{1}{2}\left(\sigma_k^{(2)}\right)^2 \left(x_w^{(2)}\right)^2 \left(\frac{f_2\left(\mu_j^{(2)}, \sigma_k^{(2)}, \lambda_l, \rho_m^{(2)}, x_w^{(2)} + \epsilon_x^{(2)}\right) + f_2\left(\mu_j^{(2)}, \sigma_k^{(2)}, \lambda_l, \rho_m^{(2)}, x_w^{(2)} - \epsilon_x^{(2)}\right)}{(\epsilon_x^{(2)})^2} \right)
\end{aligned}$$

$$
\begin{aligned}
&-\frac{1}{2}\left(\sigma_k^{(2)}\right)^2\left(x_w^{(2)}\right)^2\left(\frac{2f_2\left(\mu_j^{(2)},\sigma_k^{(2)},\lambda_l,\rho_m^{(2)},x_w^{(2)}\right)}{\left(\epsilon_x^{(2)}\right)^2}\right)\\
&+\lambda_l\sum_{i=0}^{n_\sigma^{(2)}}\left(f_2\left(\mu_j^{(2)},\sigma_i^{(2)},\lambda_l,\rho_m^{(2)},x_w^{(2)}\right)-f_2\left(\mu_j^{(2)},\sigma_k^{(2)},\lambda_l,\rho_m^{(2)},x_w^{(2)}\right)\right)\frac{1}{n_\sigma^{(2)}+1}\\
&=b_2\left(\mu_j^{(2)},\sigma_k^{(2)},x_w^{(2)}\right)\left(f_2\left(\mu_j^{(2)},\sigma_k^{(2)},\lambda_l,\rho_m^{(2)},x_w^{(2)}+\epsilon_x^{(2)}\right)-f_2\left(\mu_j^{(2)},\sigma_k^{(2)},\lambda_l,\rho_m^{(2)},x_w^{(2)}\right)\right)\\
&+d_2\left(\mu_j^{(2)},\sigma_k^{(2)},x_w^{(2)}\right)\left(f_2\left(\mu_j^{(2)},\sigma_k^{(2)},\lambda_l,\rho_m^{(2)},x_w^{(2)}-\epsilon_x^{(2)}\right)-f_2\left(\mu_j^{(2)},\sigma_k^{(2)},\lambda_l,\rho_m^{(2)},x_w^{(2)}\right)\right)\\
&+\lambda_l\left(\bar f_2\left(\mu_j^{(2)},\lambda_l,\rho_m^{(2)},x_w^{(2)}\right)-f_2\left(\mu_j^{(2)},\sigma_k^{(2)},\lambda_l,\rho_m^{(2)},x_w^{(2)}\right)\right),
\end{aligned}
\tag{4.6}
$$

where

$$
b_2\left(\mu_j^{(2)},\sigma_k^{(2)},x_w^{(2)}\right)=\frac{1}{2}\left(\frac{\left(\sigma_k^{(2)}\right)^2\left(x_w^{(2)}\right)^2}{\left(\epsilon_x^{(2)}\right)^2}+\frac{\mu_j^{(2)}x_w^{(2)}}{\epsilon_x^{(2)}}\right),
$$

$$
d_2\left(\mu_j^{(2)},\sigma_k^{(2)},x_w^{(2)}\right)=\frac{1}{2}\left(\frac{\left(\sigma_k^{(2)}\right)^2\left(x_w^{(2)}\right)^2}{\left(\epsilon_x^{(2)}\right)^2}-\frac{\mu_j^{(2)}x_w^{(2)}}{\epsilon_x^{(2)}}\right),
$$

and

$$
\bar f_2\left(\mu_j^{(2)},\lambda_l,\rho_m^{(2)},x_w^{(2)}\right)=\frac{1}{n_\sigma^{(2)}+1}\sum_{i=0}^{n_\sigma^{(2)}}f_2\left(\mu_j^{(2)},\sigma_i^{(2)},\lambda_l,\rho_m^{(2)},x_w^{(2)}\right).
$$

Remark 4.1. $b_2(\mu_j^{(2)},\sigma_k^{(2)},x_w^{(2)})$ and $d_2(\mu_j^{(2)},\sigma_k^{(2)},x_w^{(2)})$ can be thought of as birth and death rates, and $\bar f_2(\mu_j^{(2)},\lambda_l,\rho_m^{(2)},x_w^{(2)})$ is the mean of f_2 on $\sigma^{(2)}$ with fixed $(\mu_j^{(2)},\lambda_l,\rho_m^{(2)},x_w^{(2)})$.

As for boundary conditions, we can choose the range of each component large enough so that the transition probability to anything outside the range is negligible. Then, the generator, $\mathbf{A}_\varepsilon^{(2)}$, characterizes a continuous-time Markov chains satisfying Assumption 2.5. After having the discretized state space, $p_{\epsilon,j}^{(k)}=p^{(k)}(y_j|\theta_\epsilon,x_\epsilon)$ is obtained subsequently.

Step 2. Obtain the evolution of Bayes factors for the approximate models. For $k=1,2$, when $(\theta^{(k)},X^{(k)})$ is replaced by $\left(\theta_\epsilon^{(k)},X_\epsilon^{(k)}\right)$, $\mathbf{A}^{(k)}$ by $\mathbf{A}_\varepsilon^{(k)}$, $\vec{Y}^{(k)}$ by $\vec{Y}_\varepsilon^{(k)}$, $p_j^{(k)}$ by $p_{\epsilon,j}^{(k)}$ and there also exists a probability measure $P_\varepsilon^{(k)}$ to replace $P^{(k)}$, then Assumptions 2.1–2.5 also hold for $\left(\theta_\epsilon^{(k)},X_\epsilon^{(k)},\vec{Y}_\varepsilon^{(k)}\right)$. Hence, the evolution equations are obtained in Eqs. (3.8) and (3.9) using Theorem 3.1.

The procedure hitherto brings forth $q_\varepsilon^{(k)}(f_k,t)$, which is the discrete-spaced approximation of $q_k(f_k,t)$ for $k=1,2$. Let $\left(\mu_\epsilon^{(2)},\sigma_\epsilon^{(2)},\lambda_\epsilon,\rho_\epsilon^{(2)},X_\epsilon^{(2)}\right)$ denote the discretized signal for Model 2.

Definition 4.1. Let $q_{\varepsilon,t}^{(2)}$ be the conditional finite measure of $\left(\mu_\epsilon^{(2)},\sigma_\epsilon^{(2)},\lambda_\epsilon,\rho_\epsilon^{(2)},X_\epsilon^{(2)}(t)\right)$ on the discrete state space given $\mathcal{F}_t^{\vec{Y}_\varepsilon^{(2)}}$. $q_{\varepsilon,t}^{(k)}$, $k=1,2$, are the approximates of $q_t^{(k)}$ in Remark 3.2.

Definition 4.2. Let the approximate filter ratio process be

$$q_\varepsilon^{(2)}(f,t) = \sum_{\mu^{(2)},\sigma^{(2)},\lambda,\rho^{(2)},x^{(2)}} f_2\big(\mu^{(2)},\sigma^{(2)},\lambda,\rho^{(2)},x^{(2)}\big) q_{\varepsilon,t}^{(2)}\big(\mu^{(2)},\sigma^{(2)},\lambda,\rho^{(2)},x^{(2)}\big),$$

where the sum goes over all corresponding lattices in the discretized state spaces.

Step 3. Convert Eqs. (3.8) and (3.9) to the recursive algorithm. We begin with defining the discrete conditional measure in Model 2 that the recursive algorithm computes.

Definition 4.3.

$$\begin{aligned} &q_\varepsilon^{(2)}\big(\mu_j^{(2)},\sigma_k^{(2)},\lambda_l,\rho_m^{(2)},x_w^{(2)};\,t\big) \\ &\quad = q_{\varepsilon,t}^{(2)}\big\{\mu_\epsilon^{(2)} = \mu_j^{(2)}, \sigma_\epsilon^{(2)}(t) = \sigma_k^{(2)}, \lambda_\epsilon = \lambda_l, \rho_\epsilon^{(2)} = \rho_m^{(2)}, X_\epsilon^{(2)}(t) = x_w^{(2)}\big\}. \end{aligned}$$

Step 3(a), the core of the conversion, is to take f_2 as the lattice-point indicator as below:

$$\begin{aligned} &\mathbf{I}^{(2)}_{\{\mu_\epsilon^{(2)}=\mu_j^{(2)},\sigma_\epsilon^{(2)}(t)=\sigma_k^{(2)},\lambda_\epsilon=\lambda_l,\rho_\epsilon^{(2)}=\rho_m^{(2)},X_\epsilon^{(2)}=x_w^{(2)}\}}\big(\mu_\epsilon^{(2)},\sigma_\epsilon^{(2)},\lambda_\epsilon,\rho_\epsilon^{(2)},X_\epsilon^{(2)}\big) \\ &\quad \stackrel{\text{def}}{=} \mathbf{I}^{(2)}\big(\mu_j^{(2)},\sigma_k^{(2)},\lambda_l,\rho_m^{(2)},x_w^{(2)}\big). \end{aligned} \tag{4.7}$$

Then,

$$q_\varepsilon^{(2)}\big(\mathbf{I}^{(2)}\big(\mu_j^{(2)},\sigma_k^{(2)},\lambda_l^{(2)},\rho_m^{(2)},x_w^{(2)}\big),t\big) = q_\varepsilon^{(2)}\big(\mu_j^{(2)},\sigma_k^{(2)},\lambda_l,\rho_m^{(2)},x_w^{(2)};t\big).$$

So, we have a discrete-state-space finite array of $q_{\varepsilon,t}^{(k)}$ to approximate the measure $q_t^{(k)}$ of $(\theta(t), X(t))$, and the finite array evolves continuously in time according to Eqs. (3.8) and (3.9).

For more detail of Step 3(a) and Step 3(b), which is to approximate the continuous-time evolution of the finite array by an (discrete-time) Euler scheme, interested readers are referred to the corresponding parts of Bayes estimation via filtering in [13]. The final propagation part of the recursive algorithm is obtained as the following. Case 1, if $t_{i+1} - t_i \le LL$, the length controller for Euler scheme, then

$$\begin{aligned} &q_\varepsilon^{(2)}\big(\mu_j^{(2)},\sigma_k^{(2)},\lambda_l,\rho_m^{(2)},x_w^{(2)};t_{i+1}-\big) \\ &\quad \approx q_\varepsilon^{(2)}\big(\mu_j^{(2)},\sigma_k^{(2)},\lambda_l,\rho_m^{(2)},x_w^{(2)};t_i\big) \\ &\qquad + \big[b_2\big(\mu_j^{(2)},\sigma_k^{(2)},x_{w-1}^{(2)}\big)q_\varepsilon^{(2)}\big(\mu_j^{(2)},\sigma_k^{(2)},\lambda_l,\rho_m^{(2)},x_{w-1}^{(2)};t_i\big) \\ &\qquad - \big(b_2\big(\mu_j^{(2)},\sigma_k^{(2)},x_w^{(2)}\big) + d_2\big(\mu_j^{(2)},\sigma_k^{(2)},x_w^{(2)}\big)\big)q_\varepsilon^{(2)}\big(\mu_j^{(2)},\sigma_k^{(2)},\lambda_l,\rho_m^{(2)},x_w^{(2)};t_i\big) \\ &\qquad + d_2\big(\mu_j^{(2)},\sigma_k^{(2)},x_{w+1}^{(2)}\big)q_\varepsilon^{(2)}\big(\mu_j^{(2)},\sigma_k^{(2)},\lambda_l,\rho_m^{(2)},x_{w+1}^{(2)};t_i\big) \\ &\qquad + \lambda_l\big(\bar{q}_\varepsilon^{(2)}\big(\mu_j^{(2)},\lambda_l^{(2)},\rho_m^{(2)},x_w^{(2)};t\big) - q_\varepsilon^{(2)}\big(\mu_j^{(2)},\sigma_k^{(2)},\lambda_l,\rho_m^{(2)},x_w^{(2)};t\big)\big)\big](t_{i+1}-t_i). \end{aligned} \tag{4.8}$$

where

$$\bar{q}_\varepsilon^{(2)}\big(\mu_j^{(2)}, \lambda_l, \rho_m^{(2)}, x_w^{(2)}; t\big) = \frac{1}{n_\sigma^{(2)}+1}\sum_{i=0}^{n_\sigma^{(2)}} q_\varepsilon^{(2)}\big(\mu_j^{(2)}, \sigma_i^{(2)}, \lambda_l, \rho_m^{(2)}, x_w^{(2)}; t\big).$$

We choose $\big[\alpha_x^{(2)}, \beta_x^{(2)}\big]$, the range for $X^{(2)}$, large enough so that the finite measure, $q_{\epsilon,t}^{(2)}$, is negligible outside $\big[\alpha_x^{(2)}, \beta_x^{(2)}\big]$. Then, when $w < 0$ or $w > n_x^{(2)} + 1$, the corresponding $q_{\epsilon,t}^{(2)}$s are set to be zero.

Case 2, if $t_{i+1} - t_i > LL$, then we can choose a finer partition $\{t_{i,0} = t_i, t_{i,1}, \ldots, t_{i,n} = t_{i+1}\}$ of $[t_i, t_{i+1}]$ such that $\max_j |t_{i,j+1} - t_{i,j}| < LL$ and then apply repeatedly the recursive algorithms given by the above equations from $t_{i,0}$ to $t_{i,1}$, then $t_{i,2}, \ldots$, until $t_{i,n} = t_{i+1}$.

Suppose a trade at j_0th price level occurs at time t_{i+1}. Then, the updating equation of Model 2 in (3.9) becomes,

$$\begin{aligned} &q_\varepsilon^{(2)}\big(\mu_j^{(2)}, \sigma_k^{(2)}, \lambda_l, \rho_m^{(2)}, x_w^{(2)}; t_{i+1}\big) \\ &\quad= \frac{q_\varepsilon^{(2)}\big(\mu_j^{(2)}, \sigma_k^{(2)}, \lambda_l, \rho_m^{(2)}, x_w^{(2)}; t_{i+1}-\big)p^{(2)}\big(y_{j_0}|x_w^{(2)}, \rho_m^{(2)}\big)}{\sum_{j',k',l',m',w'} q_\varepsilon^{(2)}\big(\mu_{j'}^{(2)}, \sigma_{k'}^{(2)}, \lambda_{l'}, \rho_{m'}^{(2)}, x_{w'}^{(2)}; t_{i+1}-\big)p\big(y_{j_0}|x_{w'}^{(2)}, \rho_{m'}^{(2)}\big)} \\ &\qquad\times\left(\sum_{j',k',l',m',w'} q_\varepsilon^{(2)}\big(\mu_{j'}^{(2)}, \sigma_{k'}^{(2)}, \lambda_{l'}, \rho_{m'}^{(2)}, x_{w'}^{(2)}; t_{i+1}-\big)\right), \end{aligned} \tag{4.9}$$

where the sums go over all the lattices in the discretized state spaces.

Equations (4.8) and (4.9) and the two corresponding equations for Model 1 compose the recursive algorithm we employ to calculate the approximate conditional measures at time t_{i+1} from those at time t_i to time $t_{i+1}-$, and then to time t_{i+1}. At time t_{i+1}, the Bayes factor

$$B_{21}(t_{i+1}) \approx q_\varepsilon^{(2)}(1, t_{i+1}) = \sum_{j',k',l'm',w'} q_\varepsilon^{(2)}\big(\mu_{j'}^{(2)}, \sigma_{k'}^{(2)}, \lambda_{l'}, \rho_{m'}^{(2)}, x_{w'}^{(2)}; t_{i+1}\big),$$

where the sum goes over all the lattices in the discretized state space.

4.2.1. *Choosing priors*

Finally, we choose a reasonable prior for each model. We assume the independence between $X(0)$ and the parameters of a model. Set $P\{X(0) = Y(t_1)\} = 1$ where $Y(t_1)$ is the first trade price of a data set because they are very close. If there is no special information of the parameters available, we may simply assign uniform distributions to the discretized spaces of the parameters to obtain the prior at $t = 0$. For example, the prior for Model 2 in this case would be

$$\begin{aligned} &p\big(\mu_j^{(2)}, \sigma_k^{(2)}, \lambda_l, \rho_m^{(2)}, x_w^{(2)}; 0\big) \\ &\quad= \begin{cases} \dfrac{1}{\big(1+n_\mu^{(2)}\big)\big(1+n_\sigma^{(2)}\big)\big(1+n_\lambda^{(2)}\big)\big(1+n_\rho^{(2)}\big)} & \text{if } x_w^{(2)} = Y(t_1) \\ 0 & \text{otherwise} \end{cases}. \end{aligned}$$

The statistical and computational concerns for a prior on a parameter have two aspects: suitable range and mesh size. Usually, the marginal posterior of a parameter obtained from a large data set is concentrated on a small area around the true value. Then, the uniform prior set on the small area is sufficient, because the posterior outside is of very small probability. After having a suitable range, we may choose a suitable mesh size, which ideally produces a posterior with a unique model and bell-shaped distribution as shown in Table 5.1 of [12]. Therefore, we need to rely on the posterior obtained from the Bayes estimation via filtering to choose the suitable range and mesh size for the prior of each parameter.

4.2.2. *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 (3.8) by Euler scheme, whose convergence is well-known. The first, more important one, is the approximation of the evolution Eqs. (3.4) and (3.5) by the evolution Eqs. (3.8) and (3.9) of the approximate model. Since $(\theta_\epsilon^{(k)}, X_\epsilon^{(k)}) \Rightarrow (\theta^{(k)}, X^{(k)})$ by construction, Theorem 3.2 guarantees the weak convergence of the evolution equations of the approximate models to those of the assumed models.

4.3. *Simulation and real data examples*

The recursive algorithm for computing the Bayes factors is fast enough to generate real-time Bayes factors. The algorithm is extensively tested and validated on simulated data. One simulation example is provided to demonstrate the effectiveness of Bayes factors for model selection. Then, the recursive algorithm is applied to two months of transaction prices of Microsoft and the Bayes factor is extremely in favor of Model 2, which has the stochastic volatility feature.

4.3.1. *Simulation study*

In the following simulation example, the parameter values for Model 2 are selected as: $\mu = 4.5 \times 10^{-8}$, corresponding to the annualized expected return 27.38% with annualized factor 260 days and each day with 6.5 business hours; and $\lambda = 3.75 \times 10^{-4}$, which means one change of volatility every $1/3.75 \times 10^{-4} = 2666.67$ seconds on average. The range of volatility is $[0.00004, 0.0004]$, corresponding to the annualized range of $[9.866\%, 98.66\%]$. Since $a(t)$ has no impact in estimation and noise, the trading intensity is assumed to be constant: $a(t) = 0.9$ for all $t > 0$ (i.e., one trade every $1/0.9 = 1.11$ seconds on average). For the parameters of noise, we let $\rho = 0.2$, $\alpha = 0.4$, and $\beta = 0.2$. Using these parameters, 90,000 observations of Model 2 are simulated.

Based on the above data set, we would like to conduct the model selection problem between Models 1 and 2 and demonstrate that the recursive algorithm and the Bayesian model selection approach works. We follow the two-step procedure.

Step 1 is to choose proper priors (including ranges and mesh sizes) for each model. We apply, respectively, Bayes estimation via filtering of Model 1 in [12] and that of Model 2 in [13] to the simulated data. Based on the marginal posterior of each parameter, we choose suitable ranges and mesh sizes so that each marginal posterior has a unique model and bell-shaped distribution. Step 2 is to apply the recursive algorithm constructed in Sec. 4.2 to compute the approximate Bayes factors. Finally, we interpret the Bayes factors accordingly.

An important observation is that the first volatility change is a turning point. Before σ changes, the simulated data of Model 2 coincides with Model 1. However, after σ changes, the simulated data diverge from Model 1. So, we expect that the Bayes factor of Models 2 over 1, $B_{21}(t)$, will be close to 1 before the first change of σ, and $B_{21}(t)$ will increase substantially after that. More precisely, $B_{21}(t)$ is expected to be larger than 12 (the benchmark for strong evidence against Model 1) and then larger than 150 (the benchmark for decisive evidence against Model 1) shortly after the first changing of σ. This is exactly what happened as shown in Figs. 1 and 2. Figure 1 presents how $B_{21}(t)$ evolves with time for the first 2550 data. The vertical dot line is where the volatility changes (after the 2166th datum). The lower horizontal dot line is 1 and the higher horizontal dot line is 12. Prior to the vertical dot line, the Bayes factor evolves close to 1. Indeed,the Bayes factor right before the changing of volatility (at the 2166th datum) is actually 0.9358. Soon after the vertical dot line, the Bayes factor increases dramatically while fluctuating widely, and then exceeds the benchmark, 12, at the 2529th datum. Figure 2, which is only about one fifth of the time length of Fig. 1, further shows the dramatically increasing trend of B_{21} during the second σ. The first passage above 150 is at the 2609th datum. Although it retraces to 150, it later increases sharply so that B_{21} reaches 1103.70 at the end of the second σ (the 2676th datum). At the end of the third σ, $B_{21} = 1.134 \times 10^{10}$. More information is shown in Table 2. Finally, at

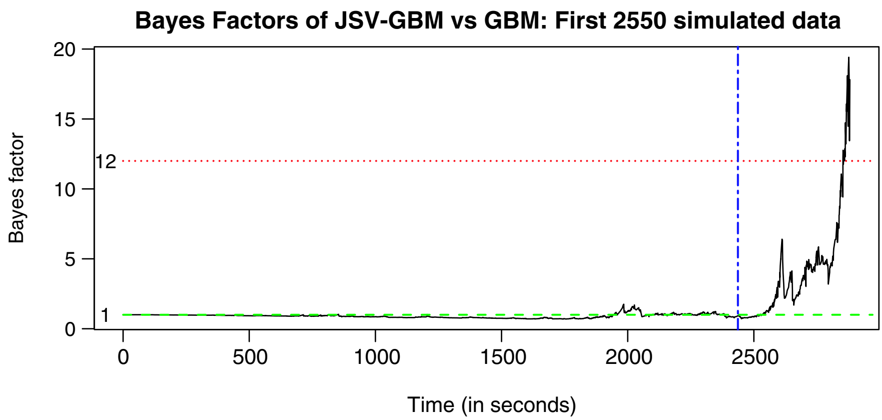

Fig. 1. Bayes factors of JSV–GBM vs GBM: First 2,550 simulated data.

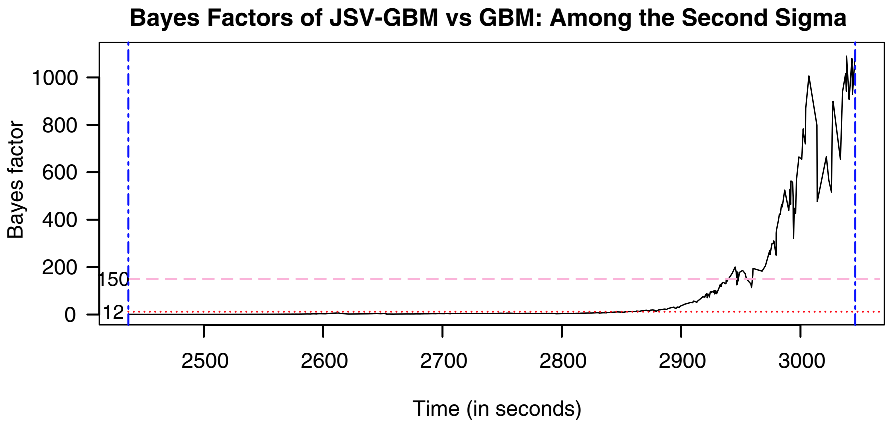

Fig. 2. Bayes factors of JSV–GBM vs GBM: Among the second sigma.

Table 2. Bayes factors for a simulated data.

Position before σ changes	2166	2676	7790	8113	8870	90000
Bayes Factor: B_{21}	0.9358	1103.70	1.134e + 10	1.255e + 10	8.197e + 21	1.089e + 194

the end of the 90,000 data, B_{21} reaches 1.089×10^{194}. All these demonstrate that the Bayes factor computed is sensitive to distinguish Model 2 from Model 1.

4.3.2. *An 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 data are extracted from the Trade and Quote (TAQ) database distributed by NYSE. We apply standard procedures to filter the data except we keep all those trades with zero time duration, because they can be handled by the recursive algorithm. The final sample has 49,937 observations.

Based on the relative frequencies of the fractional parts of the price for Microsoft, we may use the method of relative frequencies to estimate $\alpha = 0.2414$, and $\beta = 0.3502$. We follow the two-step procedure as for the simulated data to obtain the Bayes factor, $B_{12}(t)$. Table 3 consists of the precise statistics for each quarter day's trade-by-trade Bayes factors, B_{12}, in the first day, January 3, 1994 of Microsoft. The decisive benchmark, 150, which is exceeded easily. At the end of the first day, the Bayes factor (at the 826th datum) is 52669.72, which is much larger than 150. This clearly demonstrates that there are changes of σ within one trading day and is consistent with the empirical finding of higher volatility in opening and closing

Table 3. Summary statistics for BF_{21} of the first day in MSFT data.

Position	No. of Data	Min.	Median	Mean	Max.
1st Quarter Day	375	0.9133	48.19	104.20	931.30
2nd Quarter Day	164	28.64	69.40	660.00	11280.00
3rd Quarter Day	130	2178	7472	67060	584400
4th Quarter Day	287	24250	41360	75680	297800

periods of the market sessions. At the end of the second day, the Bayes factor (of the 1677th datum) is 8.261×10^{10}. At the 44478th datum (about the 36.5th day), $BF_{21} = 8.442 \times 10^{305}$. Soon after, the Bayes factor exceeds the computer's limit. This provides overwhelmingly decisive evidence for the even more frequent volatility changes in the micromovement of price.

5. Conclusion

In this paper, we investigate the model selection problems for a general class of micromovement models of asset price and develop Bayesian model selection via filtering in two steps. We first derive the evolution system of SDEs for the Bayes factors and then develop a general approach to construct consistent recursive algorithm for computing the Bayes factors. We provide an example to demonstrate the construction of recursive algorithm and the effectiveness of Bayes factor in the model selection.

The application of Bayesian model selection via filtering is not just mere to show more volatility changes in intraday data, but rather to provide a general, powerful tool to test related market microstructure theories, represented by the micromovement models. For examples, we may test whether NASDAQ has less trading noise after a market reform as argued in [1], test whether information affects trading intensity as argued by [3] and tested by [4], test whether there is relationship between transaction times and limit order arrival times as in [11], and test whether there is a structure break in transacting periods as in [15].

The Bayesian model selection via filtering is computationally intensive. To improve efficiency, especially when the number of parameters of model is large, we will extend the recent developments in particle filtering to the filtering problem with counting process observations in future work.

Appendix A

Proof of Lemma 3.1. Let $T^{(1)}$, $T^{(2)}$ be the semigroups with weak generators $A^{(1)}$, $A^{(2)}$; $\{\tau_k\}_{k=1}^{\infty}$ be the jump times of Y; $\tau_0 = 0$; and (q_1, q_2) be finite measure processes solving for $i = 1, 2$, $j = 3 - i$

$$q_i(f_i, t) = q_i(f_i, t_k) + \int_{\tau_k}^{t} q_i(A^{(i)} f_i, s) - q_i(a_i f_i, s) + \frac{q_i(f_i, s) q_j(a_j, s)}{q_j(1, s)} ds, \qquad \text{(A.1)}$$

for all $f_i \in \mathcal{D}(A^{(i)})$, $t \in [\tau_k, \tau_{k+1})$. Then, by Assumption 2.5 and (A.1) we have that

$$e^{-C(t-\tau_k)} q_j(1,\tau_k) \leq q_j(1,t) \leq e^{C(t-\tau_k)} q_j(1,\tau_k)\,, \tag{A.2}$$

for all $j = 1, 2$; $t \in [\tau_k, \tau_{k+1})$. Now, we define for $i = 1, 2$, $j = 3 - i$

$$\chi_i(t,u,f_i) = q_i\big(T^{(i)}_{t-u} f_i, u\big) + \int_u^t \frac{q_i\big(T^{(i)}_{t-s} f_i, s\big) q_j(a_j, s)}{q_j(1,s)} - q_i\big(a_i T^{(i)}_{t-s} f_i, s\big) ds,$$

for all $f_i \in \mathcal{D}(A^{(i)})$, $u \leq t \in [\tau_k, \tau_{k+1})$. Then, using the fact that $T^i_s f_i \in \mathcal{D}(A^{(i)})$ for all $s \geq 0, i = 1, 2$ as well as Leibniz's rule, one has that

$$\frac{d}{du}\chi_i(t,u,f_i) = \frac{d}{du}\big\{q_i\big(T^i_{t-u} f_i, u\big)\big\} + q_i\big(a_i T^{(i)}_{t-u} f_i, u\big) - \frac{q_i\big(T^{(i)}_{t-u} f_i, u\big) q_j(a_j, u)}{q_j(1,u)} = 0,$$

where $j = 3 - i$, so substituting $u = t, \tau_k$ into χ_i, one finds that

$$q_i(f_i, t) = q_i\big(T^i_{t-\tau_k} f_i, \tau_k\big) + \int_{\tau_k}^t \frac{q_i\big(T^{(i)}_{t-s} f_i, s\big) q_j(a_j, s)}{q_j(1,s)} - q_i\big(a_i T^{(i)}_{t-s} f_i, s\big) ds. \tag{A.3}$$

Now, suppose that $\big\{\big(q^j_1, q^j_2\big)\big\}^2_{j=1}$ are two processes solving (A.1) such that $\big(q^1_1(\cdot,\tau_k),$ $q^1_2(\cdot,\tau_k)\big) = \big(q^2_1(\cdot,\tau_k), q^2_2(\cdot,\tau_k)\big)$. Then, using (A.3) for each pair, one finds that

$$\begin{aligned}
&\big|q^2_1(f_1,t) - q^1_1(f_1,t)\big| + \big|q^2_2(f_2,t) - q^1_2(f_2,t)\big| \\
&\leq \int_{\tau_k}^t \left| \frac{q^2_1\big(T^{(1)}_{t-s} f_1, s\big) q^2_2(a_2,s)}{q^2_2(1,s)} - \frac{q^1_1\big(T^{(1)}_{t-s} f_1, s\big) q^1_2(a_2,s)}{q^1_2(1,s)} \right| ds \\
&\quad \times \int_{\tau_k}^t \left| \frac{q^2_2\big(T^{(2)}_{t-s} f_2, s\big) q^2_1(a_1,s)}{q^2_1(1,s)} - \frac{q^1_2\big(T^{(2)}_{t-s} f_2, s\big) q^1_1(a_1,s)}{q^1_1(1,s)} \right| ds \\
&\quad + \int_{\tau_k}^t \big|q^1_1\big(a_1 T^{(1)}_{t-s} f_1, s\big) - q^2_1\big(a_1 T^{(1)}_{t-s} f_1, s\big)\big| + \big|q^1_2\big(a_2 T^{(2)}_{t-s} f_2, s\big) \\
&\quad - q^2_2\big(a_2 T^{(2)}_{t-s} f_2, s\big)\big| ds,
\end{aligned} \tag{A.4}$$

for all $t \in [\tau_k, \tau_{k+1})$, $f_i \in \mathcal{D}(A^{(i)}), i = 1, 2$. However, Assumption 2.5 and (A.2) give

$$\begin{aligned}
&\left| \frac{q^2_1\Big(T^{(1)}_{t-s} f_1, s\Big) q^2_2(a_2,s)}{q^2_2(1,s)} - \frac{q^1_1\Big(T^{(1)}_{t-s} f_1, s\Big) q^1_2(a_2,s)}{q^1_2(1,s)} \right| + \left| q^1_1\Big(a_1 T^{(1)}_{t-s} f_1, s\Big) - q^2_1\Big(a_1 T^{(1)}_{t-s} f_1, s\Big) \right| \\
&\leq 2C \sup\nolimits_{\{f \in \bar{C}(E), \|f\|_\infty \leq 1\}} \Big| q^2_1(f,s) - q^1_1(f,s) \Big| \\
&\quad + 2C \sup\nolimits_{\{f \in \bar{C}(E), \|f\|_\infty \leq 1\}} \Big| q^2_2(f,s) - q^1_2(f,s) \Big|,
\end{aligned} \tag{A.5}$$

for $s \in [\tau_k, \tau_{k+1})$, $f_1 \in \mathcal{D}(A^{(1)})$ with $\|f_1\|_\infty \leq 1$. Using (A.2) and the compact containment condition, one has increasing compact sets K_m satisfying $q^2_1(K_m, t) \wedge q^1_1(K_m, t) \geq 1 - \frac{1}{m}$ for all $t \in [\tau_k, \tau_{k+1})$ so Assumption 2.5, (A.4) and (A.5), and

Stone–Weierstrass yield

$$\begin{aligned}
&\sup_{\{f_1\in\bar{C}(E),\|f_1\|_\infty\le 1\}}\left|\int_{K_m} f_1\left[dq_1^2(t)-dq_1^1(t)\right]\right| \\
&\quad+\sup_{\{f_2\in\bar{C}(E),\|f_2\|_\infty\le 1\}}\left|\int_{K_m} f_2\left[dq_2^2(t)-dq_2^1(t)\right]\right| \\
&\le \frac{2}{m}+\sup_{\{f_1\in\mathcal{D}(A^{(1)}),\|f_1\|_\infty\le 1\}}\left|q_1^2(f_1,t)-q_1^1(f_1,t)\right| \\
&\quad+\sup_{\{f_2\in\mathcal{D}(A^{(2)}),\|f_2\|_\infty\le 1\}}\left|q_2^2(f_2,t)-q_2^1(f_2,t)\right| \\
&\le \frac{2}{m}+\frac{8C}{m}(t-\tau_k)+4C\int_{\tau_k}^t \sup_{\{f_1\in\bar{C}(E),\|f\|_\infty\le 1\}}\left|\int_{K_m} f_1[dq_1^2(s)-dq_1^1(s)]\right| \\
&\quad+\sup_{\{f_2\in\bar{C}(E),\|f_2\|_\infty\le 1\}}\left|\int_{K_m} f_2\left[dq_2^2(s)-dq_2^1(s)\right]\right| ds.
\end{aligned}$$

Hence, using Gronwall's inequality, and letting $m\to\infty$, we find that

$$\begin{aligned}
&\sup_{\{f_1\in\bar{C}(E),\|f_1\|_\infty\le 1\}}\left|\int_E f_1\left[dq_1^2(t)-dq_1^1(t)\right]\right|\Bigg| \\
&\quad+\sup_{\{f_2\in\bar{C}(E),\|f_2\|_\infty\le 1\}}\left|\int_E f_2\left[dq_2^2(t)-dq_2^1(t)\right]\right|\Bigg| = 0.
\end{aligned}$$

Uniqueness on $[0,\infty)$ derives from induction and the updating equations.

Acknowledgments

We are grateful to Thomas G. Kurtz for his inspiration and insightful, productive conversations. Yong would like to thank Wei Pan for introducing the papers on Bayes factor and thank Rama Cont, Tyrone Duncan, Robert Elliott, Yaozhong Hu, Hongwei Long, David Nualart, Bozenna Pasik-Duncan, Philip Protter, Xiaotong Shen, Richard Stockbridge, Jie Xiong, George Yin, Thaleia Zariphopoulou, Qing Zhang, Ziyu Zheng and an anonymous referee for helpful comments. We also thank participants at the Workshop on Filtering Theory and Applications (2002) held at University of Alberta, the IMS Annual Meeting/The Fourth International Probability Symposium (2002) at Banff, Canada, the Fourth Winemiller Statistics Symposium (2003) at University of Missouri — Columbia, Canadian Mathematical Society Summer 2003 Meeting at University of Alberta, and seminar participants at University of Alberta, University of Kansas, University of Missouri at Kansas City, University of Texas at Austin and Universities of Wisconsin at Madison and at Milwaukee for helpful comments. The work is partially supported by Faculty Research Grant of the University of Missouri at Kansas City as well as an NSERC discovery grant and the MITACS Canadian network of centers of excellence.

References

[1] M. Barclay, W. Christie, J. Harris, E. Kandel and P. H. Schultz, The effects of market reform on the trading costs and depths of NASDAQ stocks, *Journal of Finance* **54**(1) (1999) 1–34.

[2] J. Cvitanic, R. Liptser and B. Rozovskii, A filtering approach to tracking volatility from prices observed at random times, working paper, Southern California University (2003).

[3] D. Easley and M. O'Hara, Time and the process of security price adjustment, *Journal of Finance* **47** (1992) 577–605.

[4] R. Engle, The econometrics of ultra-high-frequency data, *Econometrica* **68** (2000) 1–22.

[5] S. Ghosal, J.-K. Ghosh and T. Samanta, On convergence of posterior distributions, *Annals of Statistics* **23** (1995) 2145–2152.

[6] R. Frey and W. Runggaldier, A nonlinear filtering approach to volatility estimation with a view towards high frequency data, *International Journal of Theoretical and Applied Finance* **4** (2001) 199–210.

[7] H. Jeffreys, *Theory of Probability* (Oxford University, London, 1961).

[8] R. E. Kass and A. E. Raftery, Bayes factors and model uncertainty, *Journal of the American Statistical Association* **90** (1995) 773–795.

[9] M. Kouritzin and Y. Zeng, Weak 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) (2003) 231–239.

[10] T. Kurtz and P. Protter, Weak limit theorems for stochastic integrals and stochastic differential equations, *Annals of Probability* **19** (1991) 1035–1070.

[11] R. Russell, Econometric modeling of multivariate irregularly-spaced high-frequency data, working paper, University of Chicago (1999).

[12] Y. Zeng, A partially-observed model for micro-movement of asset prices with Bayes estimation via filtering, *Mathematical Finance* **13** (2003) 411–444.

[13] Y. Zeng, Estimating stochastic volatility via filtering for the micro-movement of asset prices, *IEEE Transactions on Automatic Control* **49** (2004) 338–348.

[14] Y. Zeng, Bayesian inference via filtering for a class of counting process: Application to the micromovement of asset price, to appear in *Statistical Inference for Stochastic Process.*

[15] M.-Y. Zhang, J.-R. Russell and R.-S. Tsay, A nonlinear autoregressive conditional duration model with applications to financial transaction data, *Journal of Econometric* **104** (2001) 179–207.