Estimating Stochastic Volatility via Filtering for the Micromovement of Asset Prices

Yong Zeng

Abstract—**Under the general framework of a previous paper, a unified approach via filtering is developed to estimate stochastic volatility for micromovement models. The key feature of the models is that they can be transformed as filtering problems with counting process observations. In order to obtain trade-by-trade, real-time Bayes estimates of stochastic volatility, the Markov chain approximation method is applied to the filtering equation to construct a consistent recursive algorithm, which computes the joint posterior. To illustrate the approach, a recursive algorithm is constructed in detail for a jumping stochastic volatility micromovement model. Simulation results show that the Bayes estimates for stochastic volatilities capture the movement of volatility. Trade-by-trade stochastic volatility estimates for a Microsoft transaction data set are obtained and they provide strong affirmative evidence that volatility changes even more dramatically at trade-by-trade level.**

Index Terms—**Counting process, filtering, high-frequency data, Markov chain approximation, stochastic volatility.**

I. Introduction

STOCHASTIC volatility is well documented for asset prices in both macromovement and micromovement [2]. Macromovement refers to daily, weekly, and monthly closing price behavior while micromovement refers to transactional (trade-by-trade) price behavior. There is a strong connection as well as striking distinctions between the macro and micromovements. The strong connection is observed through the identity of the overall shapes of both, because the macromovement is an equally spaced time series drawn from the micromovement data. Their striking distinctions are mainly due to financial noise. In macromovement, the impact of noise is small and is usually neglected. In micromovement, however, the impact of noise is substantial and noise must be modeled explicitly. If the noise is ignored, then the impact of noise is transferred to volatility, and the volatility estimates are substantially inflated. This is documented by [8], [1], and [4] for discrete noise and further in [9] and [18] for discrete plus other types of noise.

Economically, the asset price is distinguished from its intrinsic value and this distinction is also noise. Noise, as contrasted with information, is well-documented in the market microstructure literature. Three important types of noise have been identified and extensively studied: discrete, clustering and nonclustering. First, intraday prices move discretely (tick by tick), resulting in "discrete noise." Second, because prices do not distribute evenly on all ticks, but gather more on some ticks such as integers and halves, "price clustering" is obtained. [10] confirms that this phenomenon is remarkably persistent through time, across assets, and across market structures. Third, the "nonclustering noise" includes other unspecified noise, and the outliers in prices are one of the evidence for the existence of nonclustering noise.

In [18], a novel, economically well-grounded, partially observed micromovement model for asset price is proposed to bridge the gap between the macro and micro movements caused by noise. The most prominent feature of the proposed model is that it can be formulated as a filtering problem with counting process observations. This connects the model to the filtering literature, which has found great success in engineering and networking. Under this framework, the observables are the whole sample paths of the counting processes, which contain the complete information of price and trading time. Then, the continuous-time likelihoods and posterior, built upon the sample paths, not only exist, but also are uniquely characterized by the unnormalized, Duncan–Mortensen–Zakai (DMZ)-like filtering equation, and the normalized, Kushner–Stratonovich (KS) (or Fujisaki–Kallianpur–Kunita)-like filtering equations respectively.

Transaction (or tick) data are discrete in value, irregularly spaced in time and extremely large in size. Despite recent advances in statistics and econometrics, obtaining "reliable" parameter estimates for even simple, nonstochastic volatility, micromovement models is extremely challenging. Contrasted with [5], [18] develops continuous-time Bayes estimation via filtering with efficient algorithms for the parameter estimation of the micromovement model. That represents a significant advance in the estimation for micromovement models, also because the continuous-time likelihoods and posterior are utilized as the foundation for statistical inference. This foundation is informationally better than those provided by the discrete-time likelihoods and posterior, which merely make use of a discrete-time subset of the sample paths. In [18], however, only the parameters of a simple model with GBM as value process is estimated.

In this paper, first, a class of stochastic volatility micromovement models is developed from the macromovement models by incorporating the three types of noise mentioned. A new, jumping stochastic volatility (JSV) micromovement model, stemming from geometric Brownian motion (GBM), is proposed and studied (later, it is called the JSV-GBM micromovement model). Second, a unified approach, Bayes parameter estimation via filtering, is developed for the mi-

Manuscript received October 15, 2002; revised November 15, 2003. Recommended by Guest Editor B. Pasik-Duncan. This work supported in part by a Faculty Research Grant of the University of Missouri at Kansas City.

The author is with Department of Mathematics and Statistics, University of Missouri at Kansas City, Kansas City, MO 64110 USA (e-mail: zeng@mendota.umkc.edu).

Digital Object Identifier 10.1109/TAC.2004.824478

cromovement models, especially for estimating stochastic volatility. Stochastic volatility models are more realistic and more interesting but more difficult to estimate than the simple model with GBM as value process, where the parameters, the signal of interest, are fixed. In stochastic volatility model, estimation becomes a "real" filtering problem: the stochastic volatility, the signal of interest, changes over time and the stock prices are the observations corrupted by discrete types of noise.

The JSV-GBM model is employed to demonstrate the effectiveness of estimating stochastic volatility using Bayes estimation via filtering. To illustrate the approach, JSV-GBMs consistent recursive algorithm, which approximates the normalized filtering equation and calculates the joint posterior, is constructed in detail. Simulation results show that the Bayes estimates for stochastic volatilities are close to their true volatilities, and are able to capture the movement of volatility. Trade-by-trade volatility estimates for an actual transaction data set are computed and they confirm that the volatility changes even more dramatically in micromovement.

The rest of the article proceeds as follows. Section II presents the class micromovement models with stochastic volatility. Section III develops the unified approach, Bayes estimation via filtering, for the proposed models and the recursive algorithm of JSV-GBM is constructed in detail as an illustration of the approach. Section IV presents simulation and estimation results based on actual data. Concluding remarks are offered in Section V.

II. Stochastic Volatility Micromovement Models

The micromovement model is built upon the macromovement model, or, economically, the price is built upon the intrinsic value by incorporating noise. Suppose that the intrinsic value process X 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 $1/M$. The combination of (X, Y) provides a natural partially observed framework for the micromovement process. So, we start with stochastic volatility models for value process.

A. Stochastic Volatility Models for Value Process

Suppose $\theta(t)$ is a vector of parameters in the model. (X, Y) is augmented to (θ, X, Y) in preparation for parameter estimation, and the following assumption is invoked on (θ, X) as in [18].

Assumption 2.1: (θ, X) is the solution of a martingale problem for a generator $\mathbf{A}$ such that

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

is a $\mathcal{F}_t^{\theta,X}$-martingale, where $\mathcal{F}_t^{\theta,X}$ is the σ-algebra generated by $(\theta(s), X(s))_{0 \le s \le t}$.

Remark 2.1: This is a very general assumption including many stochastic volatility models and even more. Three examples are given later.

Example 2.1: Hull and White's [13] model in SDE form is

$$\frac{dX(t)}{X(t)} = \mu dt + \sigma(t) dW(t)$$
$$\frac{dV(t)}{V(t)} = \nu dt + \kappa dB(t)$$

where $V(t) = \sigma^2(t)$ and $W(t)$ and $B(t)$ are Weiner processes with correlation coefficient ρ. Its generator is

$$\mathbf{A} f(v, x) = \mu x \frac{\partial f}{\partial x}(v, x) + \frac{1}{2} v x^2 \frac{\partial^2 f}{\partial x^2}(v, x) + \nu v \frac{\partial f}{\partial v}(v, x) + \frac{1}{2}\kappa^2 v^2 \frac{\partial^2 f}{\partial v^2}(v, x) + \rho \kappa x v^{3/2} \frac{\partial^2 f}{\partial x \partial v}(v, x).$$

The parameters of this model are $(\mu, \sigma(t), \nu, \kappa, \rho)$. If ν is allowed to be a nice function that can depend on $V(t)$, then this model can include another stochastic volatility model—the limiting diffusion model of AR(1)EARCH model [16].

Example 2.2: Heston's [12] model in SDE form is

$$\frac{dX(t)}{X(t)} = \mu dt + \sigma(t) dW(t)$$
$$dV(t) = (\nu - \alpha V(t))dt + \kappa \sigma(t) dB(t)$$

where $V(t), W(t)$, and $B(t)$ are defined in Example 2.1. Its generator is

$$\mathbf{A} f(v, x) = \mu x \frac{\partial f}{\partial x}(v, x) + \frac{1}{2} v \frac{\partial^2 f}{\partial x^2}(v, x) + (\nu - \alpha v) \frac{\partial f}{\partial v}(v, x) + \frac{1}{2}\kappa^2 v^2 \frac{\partial^2 f}{\partial v^2}(v, x) + \rho \kappa v x \frac{\partial^2 f}{\partial x \partial v}(v, x).$$

This model has parameters $(\mu, \sigma(t), \nu, \alpha, \kappa, \rho)$.

The following example presents a new JSV-GBM model for value process. Its micromovement version is developed in Section II-B and is further specified in Section III-B.

Example 2.3: (JSV-GBM) A jumping stochastic volatility model in SDE form is

$$\frac{dX(t)}{X(t)} = \mu dt + \sigma(t) dW(t)$$
$$d\sigma(t) = (J_{N(t)} - \sigma(t-))dN(t)$$

where $N(t)$ is a Poisson process with intensity λ, and $\{J_i\}$ is a sequence of i.i.d. random variables with a density $g(z)$, and is assumed to be independent of $W(t)$ and $N(t)$. When the ith Poisson event happens at time t_i, the volatility changes from $\sigma(t_{i-1}-)$, which is J_{i-1}, to J_i, because $\sigma(t_i) = (J_i - \sigma(t_{i-1}-)) + \sigma(t_{i-1}-)$. Then, the volatility remains the same until the next Poisson event occurs. Its generator is

$$\mathbf{A} f(\sigma, x) = \mu x \frac{\partial f}{\partial x}(\sigma, x) + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 f}{\partial x^2}(\sigma, x) + \lambda \int_{-\infty}^{\infty} (f(z, x) - f(\sigma, x)) g(z)\, dz.$$

The parameters of this model are (μ, σ, λ) and may include parameters in $g(z)$ if they exist. This model is closely related to the asset price models with Markov modulated volatilities studied by [6].

After specifying the value process $X(t)$, there are two equivalent methods to build the price process from $X(t)$. 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 important for parameter estimation.

B. Building Price Process by Construction

Since prices can only be observed at irregularly spaced trading times, there are two steps in constructing Y from X. First, determine trading times $t_1, t_2, \dots, t_i, \dots,$ which are assumed to be modeled by a conditional Poisson process. The rate of trading activity is described by an intensity function, denoted by $a(\theta(t), X(t), t)$. Second, $Y(t_i)$, the price at time t_i, is constructed from $X(t_i)$ via a random function $F(\cdot)$. That is, $Y(t_i) = F(X(t_i))$, where $y = F(x)$ is a random transformation with the transition probability $p(y \mid x)$.

Under this construction, the observable price is produced from the value process by combining noises when a trade occurs. The model has a desirable structure as in [11]: Information affects $X(t)$, the value of an asset, and has a permanent influence on the price, while noise affects $F(x)$, the random transition function, and only has a transient impact on price.

This article focuses on estimating this class of micromovement stochastic volatility models: the value process, $X(t)$, is a continuous-time stochastic volatility model which can be formulated as a martingale problem, the trading times $t_1, t_2, \dots, t_i, \dots$ are driven by a conditional Poisson process, and the price at time t_i is $Y(t_i) = F(X(t_i))$. The advantage of these models is that they not only have stochastic volatility, but also can capture the micromovement features created by noise. Later, a simple $F(x)$ is built to accommodate the three types of noise that are well-documented in the financial literature: discrete noise, clustering noise, and nonclustering noise.

To simplify notation, fix i and 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 defined as the nonclustering noise. Instead of directly specifying $p(y \mid x)$ of $F(x)$, we move the value x to the price y in three steps.

Step 1) Incorporate discrete noise by rounding off x to its closest tick, $R[x, (1/M)]$. Without other noises, trades should occur at this tick, which is closest to the stock value.

Step 2) Incorporate nonclustering noise by adding $V; y' = R[x+V, (1/M)]$, where V is the nonclustering 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

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

Step 3) Incorporate clustering noise by biasing y'. After rounding the value process and adding nonclustering noise, the fractional part of y' should still be approximately uniformly distributed. Here, I bias y' through a random biasing function $b(\cdot)$ to reflect price clustering. $\{b_i(\cdot)\}$ are assumed independent of $\{y'_i\}$ and serially independent given the sequence $\{y'_i\}$.

To be consistent with the data analyzed in Section IV, I construct a simple random biasing function only for the tick of $1/8$ dollar. It can be generalized to other ticks such as $1/100$ dollar easily. Although, since 2000, the trading tick has been converted to decimal system, but this does not change the validity of the proposed model.

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 eighths 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 even eighths, then y stays on y' with probability one; if the fractional part of y' is 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. The detail of $b(\cdot)$ is presented in Appendix A.

In summary, the construction 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 financial noise, is specified. The transition probability $p(y \mid x)$ of F can be computed and the explicit $p(y \mid x)$ is given in Appendix A.

C. Filtering With Counting Process Observations

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

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

where $Y_k(t) = N_k(\int_0^t \lambda_k(\theta(s), X(s), s)\, ds)$ is the counting process recording the cumulative number of trades that have occurred at the kth price level (denoted by y_k) up to time t.

Under this representation, $(\theta(t), X(t))$ becomes the signal process, which cannot be observed directly, and $\vec{Y}(t)$ becomes the observation process, which is corrupted by noise, modeled by $p(y \mid x)$. Hence, $(\theta, X, \vec{Y})$ is framed as a filtering problem with counting process observations.

Four mild assumptions are invoked so that the general construction is equivalent in distribution to the counting process observations in (1). The equivalence ensures that Bayes estimation based on the latter specification can be applied to the former.

Assumption 2.2: N_k's are unit Poisson processes under the physical measure P.

Assumption 2.3: $(\theta, X), N_1, N_2, \dots, N_n$ are independent under the physical measure P.

Assumption 2.4: The intensity, $\lambda_k(\theta, x, t) = a(\theta, x, t)p(y_k \mid x)$, where $a(\theta, x, t)$ is the total intensity at time t and

$p(y_k \mid x)$ is the transition probability from x to y_k, the kth price level.

Assumption 2.5: The total intensity, $a(\theta, x, t)$, is uniformly bounded above, i.e., there exist a constant, C, such that $0 \leq a(\theta, x, t) \leq C$ for all θ, x, and t.

Remark 2.2: These four assumptions imply that there exists a reference measure Q and that after a suitable change of measure to $Q, (\theta, X), Y_1, \ldots, Y_n$ will become independent, and $Y_1, Y_2, \ldots, Y_n$ will become unit Poisson processes. These two assumptions are general since a large class of counting processes can be transformed into this setup by the technique of change of measure (see [3, p. 165]).

Remark 2.3: Assumption 2.4 imposes a desirable structure for the intensities of the model. It means that the total trading intensity determines the overall rate of trade occurrence at time t and $p(y_k \mid x)$ determines the proportional intensity of trade at the price level, y_k, when the value is x. The structure of intensities guarantees the equivalence of the above two approaches.

III. Bayes Estimation via Filtering

Bayes estimate, which is the posterior mean, is the least mean square error (MSE) estimate. The normalized filtering equation characterizes how the posterior evolves. The core of the Bayesian estimation via filtering is to construct an algorithm to compute the conditional distribution, which becomes a posterior after a prior is assigned. The algorithm, which is based on the filtering equation, is naturally recursive with every trade. One basic requirement for the recursive algorithm is consistency, namely, the conditional distribution computed by the recursive algorithm converges to the true one assumed in the model. This is guaranteed by a theorem on the convergence of conditional expectation.

In this section, we first review two main results: the filtering equation and the theorem on the convergence of conditional expectation. Then, we further prepare JSV-GBM as an illustration for the approach. Finally, we construct a consistent recursive algorithm for JSV-GBM.

A. Review

The following two results are from [18].

1) Filtering Equation: Before the filtering equation is presented, some terms are defined later. Let $\mathcal{F}_t^{\vec{Y}} = \sigma\{(\vec{Y}(s)) \mid 0 \leq s \leq t\}$ be the σ-algebra generated by the observed sample path of $\vec{Y}$. $\mathcal{F}_t^{\vec{Y}}$ is all the available information up to time t.

Definition 3.1: Let π_t be the conditional distribution of $(\theta(t), X(t))$ given $\mathcal{F}_t^{\vec{Y}}$.

Remark 3.1: Assuming priors on $(\theta(0), X(0)), \pi_t$ becomes the joint posterior of $(\theta(t), X(t))$.

Definition 3.2: Let $\pi(f,t) = E^P[f(\theta(t), X(t)) \mid \mathcal{F}_t^{\vec{Y}}] = \int f(\theta, x)\pi_t(d\theta, dx)$ be the conditional expectation of $(\theta(t), X(t))$ given $\mathcal{F}_t^{\vec{Y}}$.

Theorem 3.1: Suppose that (θ, X) satisfies Assumption 2.1, and $\vec{Y}$ is defined in (1)with Assumptions 2.2–2.5. Then, for every $t > 0$ and every f in the domain of generator $\mathbf{A}$, $\pi(f,t)$ is the unique solution of the SDE, the normalized filtering equation,

$$\begin{aligned}\pi(f,t) = \pi(f,0) + \int_0^t [\pi(\mathbf{A}f, s) - \pi(fa, s) \\ + \pi(f,s)\pi(a,s)]\, ds \\ + \sum_{k=1}^n \int_0^t \left[\frac{\pi(fap_k, s-)}{\pi(ap_k, s-)} - \pi(f, s-)\right] dY_k(s). \quad (2)\end{aligned}$$

Remark 3.2: When the trading intensity is deterministic, $a(\theta(t), X(t), t) = a(t)$, the normalized filtering equation is simplified as

$$\begin{aligned}\pi(f,t) = \pi(f,0) + \int_0^t \pi(\mathbf{A}f, s)\, ds \\ + \sum_{k=1}^n \int_0^t \left[\frac{\pi(fp_k, s-)}{\pi(p_k, s-)} - \pi(f, s-)\right] dY_k(s). \quad (3)\end{aligned}$$

Note that $a(t)$ disappears in (3). This reduces the computation greatly in the Bayesian parameter estimation. Hence, this convenient case is assumed in JSV-GBM. The tradeoff is that the relationship between trading intensity and other parameters (such as volatility) is excluded.

Let the trading times be $t_1, t_2, \ldots$, then (3) 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 updating when a trade occurs. The propagation equation has no random component and is written as

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

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

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

$$\pi(f, t_{i+1}) = \frac{\pi(fp_k, t_{i+1}-)}{\pi(p_k, t_{i+1}-)}.$$

It is random because the price level is random.

2) Convergence Theorem on Conditional Expectation: Suppose the state space of (θ, X) is discretized with ϵ_i as the length between lattices in the ith component of θ and ϵ_x as that of X. Let $\epsilon = (\epsilon_1, \ldots, \epsilon_p)$. Then, $(\theta_\epsilon, X_{\epsilon_x})$, an approximation for (θ, X), can be constructed. Define

$$\vec{Y}_\varepsilon(t) = \begin{pmatrix} N_1\left(\int_0^t \lambda_1(\theta_\epsilon(s), X_{\epsilon_x}(s), s)\, ds\right) \\ N_2\left(\int_0^t \lambda_2(\theta_\epsilon(s), X_{\epsilon_x}(s), s)\, ds\right) \\ \vdots \\ N_n\left(\int_0^t \lambda_n(\theta_\epsilon(s), X_{\epsilon_x}(s), s)\, ds\right) \end{pmatrix} \quad (4)$$

where $\varepsilon = \max(\epsilon_x, |\epsilon|)$, and define $\mathcal{F}_t^{\vec{Y}_\varepsilon} = \sigma(\vec{Y}_\varepsilon(s), 0 \leq s \leq t)$. We use the notation, $X_\epsilon \Rightarrow X$, to mean X_ϵ converges weakly (or in probability measure) to X in the Skorohod topology as $\varepsilon \to 0$.

Theorem 3.2: Suppose that $(\theta, X, \vec{Y})$ is on the probability space $(\Omega, \mathcal{F}, P)$ with Assumptions 2.1–2.5. $\vec{Y}_\varepsilon$ is defined by (4). Suppose that $(\theta_\epsilon, X_{\epsilon_x}, \vec{Y}_\varepsilon)$ is on $(\Omega_\varepsilon, \mathcal{F}_\varepsilon, P_\varepsilon)$, and Assumptions

2.1–2.5 also hold for $(\theta_\epsilon, X_{\epsilon_x}, \vec{Y}_\varepsilon)$. If $(\theta_\epsilon, X_{\epsilon_x}) \Rightarrow (\theta, X)$ as $\varepsilon = \max\{\epsilon_x, |\epsilon|\} \to 0$, then

i) $\vec{Y}_\varepsilon \Rightarrow \vec{Y}$ as $\varepsilon \to 0$;

ii) $E^{P_\varepsilon}[F(\theta_\epsilon(t), X_{\epsilon_x}(t)) \,|\, \mathcal{F}_t^{\vec{Y}_\varepsilon}] \Rightarrow E^P[F(\theta(t), X(t)) \,|\, \mathcal{F}_t^{\vec{Y}}]$ as $\varepsilon \to 0$ for every bounded continuous function F.

Remark 3.3: This theorem provides a recipe for constructing a consistent recursive algorithm. When we take "F" as an appropriate indicator function, $E^{P_\varepsilon}[F(\theta_\epsilon(t), X_{\epsilon_x}(t)) \,|\, \mathcal{F}_t^{\vec{Y}_\varepsilon}]$ becomes the conditional probability mass function for $(\theta_\epsilon, X_{\epsilon_x})$. Theorem 3.2, then, implies that the conditional probability mass function is an approximation to the conditional distribution of (θ, X). The recursive algorithm to be constructed is to compute such a conditional probability mass function, which becomes a posterior after a prior is assigned.

B. Further Preparation of JSV-GBM

The value process is the JSV-GBM model in Example 2.3. This model builds upon GBM as the models of [13] and [12] do. Their difference is in the volatility dynamics. In [13] and [12], the path of stochastic volatility is continuous. However, in JSV-GBM, the volatility jumps according to a Poisson process.

To reduce computation, we further assume that the distribution for the i.i.d. sequence of jumps, $\{J_i\}$, is uniformly distributed on a range, $[\alpha_\sigma, \beta_\sigma]$. This implies that every time when a volatility changes, the new volatility is randomly drawn from a uniform distribution on $[\alpha_\sigma, \beta_\sigma]$. Accommodating other parameters, the generator of JSV-GBM becomes

$$\begin{aligned}\mathbf{A}f(\mu,\sigma,\lambda,\rho,x) &= \mu x \frac{\partial f}{\partial x}(\mu,\sigma,\lambda,\rho,x) + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 f}{\partial x^2}(\mu,\sigma,\lambda,\rho,x) \\ &\quad + \lambda \int_{\alpha_\sigma}^{\beta_\sigma} (f(\mu,z,\lambda,\rho,x) - f(\mu,\sigma,\lambda,\rho,x)) \\ &\quad \times \frac{1}{\beta_\sigma - \alpha_\sigma}\, dz. \end{aligned} \tag{5}$$

There are two reasons to choose this model for illustration. First, the model is built upon GBM, the standard model in much of mathematical finance, with stochastic volatility. Second, the stochastic volatility jumps and the generator of the model involves both diffusion and jump generators, the two major types of generator. Then, the issue of how to build the approximate Markov chain and further recursive algorithms for both diffusion and jump processes is illustrated in one example.

There are six parameters in the model: $(\mu, \sigma(t), \lambda, \rho, \alpha, \beta)$. The two clustering parameters (α, β) can be estimated by the method of relative frequency. The estimation of stochastic volatility as well as the other parameters is done by Bayes estimation via filtering.

C. Construction of the Recursive Algorithm

For the nonlinear filtering problem, Kushner [15, Ch. 12] develops approximation methods based on replacing the signal process by a finite state Markov chain that approximates the signal equally spaced in time. Although trades occur irregularly spaced in time, the same idea can be applied here.

Theorem 3.1 provides the optimum filter and Theorem 3.2 provides a recipe for constructing a consistent recursive algorithm as an approximate optimum filter. The recursive algorithm in this section is constructed for (3), when the trading intensity is deterministic with the generator in (5).

There are three steps in the process of deriving the recursive algorithm.

Step 1: Construct a Markov chain $(\theta_\varepsilon(t), X_{\varepsilon_x}(t))$ as an approximation to $(\theta(t), X(t))$. Here, $\theta(t) = (\mu, \sigma(t), \lambda, \rho)$, all the parameters to be estimated by filtering.

First, we discretize the parameter spaces of $\mu, \sigma, \lambda, \rho$, and X. 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 $\mu, \sigma, \lambda, \rho$, and X respectively. For example, the discretization for μ is, $\mu : [\alpha_\mu, \beta_\mu] \to \{\alpha_\mu, \alpha_\mu + \epsilon_\mu, \alpha_\mu + 2\epsilon_\mu, \ldots, \alpha_\mu + j\epsilon_\mu, \ldots, \alpha_\mu + n_\mu\epsilon_\mu\}$ where $\alpha_\mu + n_\mu\epsilon_\mu = \beta_\mu$ and the number of lattices is $n_\mu + 1$. Define $\mu_j = \alpha_\mu + j\epsilon_\mu$, the jth lattice in the latticized parameter space of μ. Similarly, define $\sigma_k = \sigma_k(t) = \alpha_\sigma + k\epsilon_\sigma$, $\lambda_l = \alpha_\lambda + l\epsilon_\lambda$, $\rho_m = \alpha_\rho + m\epsilon_\rho$,, and $x_w = x_w(t) = \alpha_x + w\epsilon_x$. 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, and the Markov chain is a natural approximation for the stochastic process $(\sigma(t), X(t))$.

Second, it is observed that the construction of an approximate Markov chain can be transformed to the construction of a Markov chain generator, $\mathbf{A}_\varepsilon$, such that as $\varepsilon \to 0$. $\mathbf{A}_\varepsilon \to \mathbf{A}$, the one given by (5). The diffusion generator involves first- and second-order differentiation and the jump generator involves integration. The finite difference approximation is applied for differentiation and the rectangle approximation for integration. One constructs $\mathbf{A}_\varepsilon$ as follows:

$$\begin{aligned}&\mathbf{A}_\varepsilon f(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w) \\ &= \mu_j x_w \left(\frac{f(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w+\epsilon_x)}{2\epsilon_x} \right. \\ &\quad \left. - \frac{f(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w-\epsilon_x)}{2\epsilon_x} \right) \\ &\quad + \frac{1}{2}\sigma_k^2 x_w^2 \left(\frac{f(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w+\epsilon_x)}{\epsilon_x^2} \right. \\ &\quad \left. + \frac{f(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w-\epsilon_x) - 2f(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w)}{\epsilon_x^2} \right) \\ &\quad + \lambda_l \sum_{i=0}^{n_\sigma} (f(\mu_j,\sigma_i,\lambda_l,\rho_m,x_w) \\ &\quad - f(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w)) \frac{1}{n_\sigma + 1} \\ &= a(\mu_j,\sigma_k,x_w)(f(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w+\epsilon_x) \\ &\quad - f(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w)) \\ &\quad + b(\mu_j,\sigma_k,x_w)(f(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w-\epsilon_x) \\ &\quad - f(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w)) \\ &\quad + \lambda_l(\bar{f}(\mu_j,\lambda_l,\rho_m,x_w) - f(\mu_j,\sigma_k,\lambda_l,\rho_m,x_w)) \end{aligned} \tag{6}$$

where

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

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

and

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

Remark 3.4: $a(\mu_j, \sigma_k, x_w)$ is the birth rate and $b(\mu_j, \sigma_k, x_w)$ is the death rate, which should always be nonnegative. If for some values of x, μ, and σ in their ranges, one of the rates becomes negative, then we always make the negative rate positive by making ϵ_x small. And $\bar{f}(\mu_j, \lambda_l, \rho_m, x_w)$ is the mean of f on σ.

Clearly, $\mathbf{A}_\varepsilon \to \mathbf{A}$ as $\varepsilon \to 0$. The Markov chain based upon $\mathbf{A}_\varepsilon$ is the approximate model of $(\theta(t), X(t))$. Then, $\vec{Y}_\varepsilon$, defined by Equation (4), is obtained.

Remark 3.5: The counting process observations can be viewed as $\vec{Y}(t)$ defined by (1) or $\vec{Y}_\varepsilon(t)$ be (4) depending upon whether the driving process is $(\theta(t), X(t))$ or $(\theta_\epsilon(t), X_{\epsilon_x}(t))$. By Theorem 3.2, when ε is small, the recursive algorithm computes the posterior for the approximate model $(\theta_\epsilon, X_{\epsilon_x}, \vec{Y}_\varepsilon)$, which is close to the posterior of the true model (θ, X, Y).

Step 2: Obtain the filtering equation for the approximate model. When (θ, X) is replaced by $(\theta_\epsilon, X_{\epsilon_x})$, $\mathbf{A}$ by $\mathbf{A}_\varepsilon$, $\vec{Y}$ by $\vec{Y}_\varepsilon$, and there also exists a probability measure P_ε to replace P, then Assumptions 2.1–2.5 also hold for $(\theta_\epsilon, X_{\epsilon_x}, \vec{Y}_\varepsilon)$. Let $(\mu_\epsilon, \sigma_\epsilon, \lambda_\epsilon, \rho_\epsilon, X_\epsilon)$ denote the discretized signal of $(\mu_{\epsilon_\mu}, \sigma_{\epsilon_\sigma}, \lambda_{\epsilon_\lambda}, \rho_{\epsilon_\rho}, X_{\epsilon_x}(t))$. To present the filtering equation for the approximate model, the discretized approximations of π_t and $\pi(f, t)$ in Definitions 3.1 and 3.2 are defined as follows.

Definition 3.3: Let $\pi_{\varepsilon,t}$ be the conditional probability mass of $(\mu_\epsilon, \sigma_\epsilon, \lambda_\epsilon, \rho_\epsilon, X_\epsilon(t))$ given $\mathcal{F}_t^{\vec{Y}_\varepsilon}$.

Definition 3.4: Let

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

where $(\mu, \sigma, \lambda, \rho, x)$ covers all lattices in the discretized state space.

Applying Theorem 3.1, we obtain the similar filtering equation for the approximate model

$$\begin{aligned}\pi_\varepsilon(f, t) = \pi_\varepsilon(f, 0) &+ \int_0^t \pi_\varepsilon(\mathbf{A}_\varepsilon f, s)\, ds \\ &+ \sum_{k=1}^n \int_0^t \left[\frac{\pi_\varepsilon(fp_k, s-)}{\pi_\varepsilon(p_k, s-)} - \pi_\varepsilon(f, s-)\right] dY_{\varepsilon,k}(s) \quad (7)\end{aligned}$$

where $Y_{\varepsilon,k}$ is the kth component of $\vec{Y}_\varepsilon$.

Similarly, the previous filtering equation can be separated into the propagation equation

$$\pi_\varepsilon(f, t_{i+1}-) = \pi_\varepsilon(f, t_i) + \int_{t_i}^{t_{i+1}-} \pi_\varepsilon(\mathbf{A}_\varepsilon f, s)\, ds. \quad (8)$$

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

$$\pi_\varepsilon(f, t_{i+1}) = \frac{\pi_\varepsilon(fp_k, t_{i+1}-)}{\pi_\varepsilon(p_k, t_{i+1}-)}. \quad (9)$$

Step 3: Convert (8) and (9) to the recursive algorithm. First, we define the approximate posterior that the recursive algorithm computes.

Definition 3.5: The posterior of the approximate model, $(\mu_\epsilon, \sigma_\epsilon(t), \lambda_\epsilon, \rho_\epsilon, X_\epsilon(t), Y_\varepsilon(t))$, at time t is denoted by

$$\begin{aligned}&p_\varepsilon(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w; t) \\ &\quad = P\Big\{\mu_\epsilon = \mu_j, \sigma_\epsilon(t) = \sigma_k, \lambda_\epsilon = \lambda_l, \\ &\qquad \rho_\epsilon = \rho_m, X_{\epsilon_x}(t) = x_w \,\Big|\, \mathcal{F}_t^{\vec{Y}_\varepsilon}\Big\}.\end{aligned}$$

The core of the conversion is to take f as the following indicator function:

$$\begin{aligned}&\mathbf{I}_{\{\mu_\epsilon = \mu_j, \sigma_\epsilon(t) = \sigma_k, \lambda_\epsilon = \lambda_l, \rho_\epsilon = \rho_m, X_\epsilon(t) = x_w\}} \\ &\quad \times (\mu_\epsilon, \sigma_\epsilon(t), \lambda_\epsilon, \rho_\epsilon, X_\epsilon(t)) \\ &\quad \stackrel{\text{def}}{=} \mathbf{I}(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w). \quad (10)\end{aligned}$$

Then, the following facts emerge:

$$\begin{aligned}&\pi_\varepsilon(\mathbf{I}(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w), t) \\ &\quad = p_\varepsilon(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w; t)\end{aligned}$$

and

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

Along with three similar results, $\pi_\varepsilon(\mathbf{A}_\varepsilon \mathbf{I}, t)$ in (8) becomes explicit and the propagation (8) becomes

$$\begin{aligned}&p_\varepsilon(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w; t_{i+1}-) \\ &\quad = p_\varepsilon(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w; t_i) \\ &\qquad + \int_{t_i}^{t_{i+1}-} (a(\mu_j, \sigma_k, x_{w-1}) \\ &\qquad \times p_\varepsilon(\mu_j, \sigma_k, \lambda_l, \rho_m, x_{w-1}; t) \\ &\qquad - (a(\mu_j, \sigma_k, x_w) + b(\mu_j, \sigma_k, x_{w+1})) \\ &\qquad \times p_\varepsilon(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w; t) \\ &\qquad + b(\mu_j, \sigma_k, x_{w+1})p_\varepsilon(\mu_j, \sigma_k, \lambda_l, \rho_m, x_{w+1}; t) \\ &\qquad + \lambda_l(\bar{p}_\varepsilon(\mu_j, \lambda_l, \rho_m, x_w; t) \\ &\qquad - p_\varepsilon(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w; t)))\, dt. \quad (11)\end{aligned}$$

If a trade at kth price level occurs at time t_{i+1}, the updating (9) can be written as (12), shown at the bottom of the page, where the summation is over the latticized spaces of $\mu, \sigma, \lambda, \rho$ and $X(t_{i+1}-)$. Note that $p(y_k \,|\, x_w, \rho_m)$, the transition probability from x_w to y_k, is specified in (15) in Appendix A.

Next, we make (11) a recursive algorithm. Equation (11) is deterministic and an Euler scheme is employed for approximation. After excluding the probability-zero event that two or more jumps occur at the same time, there are two possible cases for

$$p_\varepsilon(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w; t_{i+1}) = \frac{p_\varepsilon(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w; t_{i+1}-)p(y_k \,|\, x_w, \rho_m)}{\sum_{j',k',l',m',w'} p_\varepsilon(\mu_{j'}, \sigma_{k'}, \lambda_{l'}, \rho_{m'}, x_{w'}; t_{i+1}-)p(y_k \,|\, x_{w'}, \rho_{m'})} \quad (12)$$

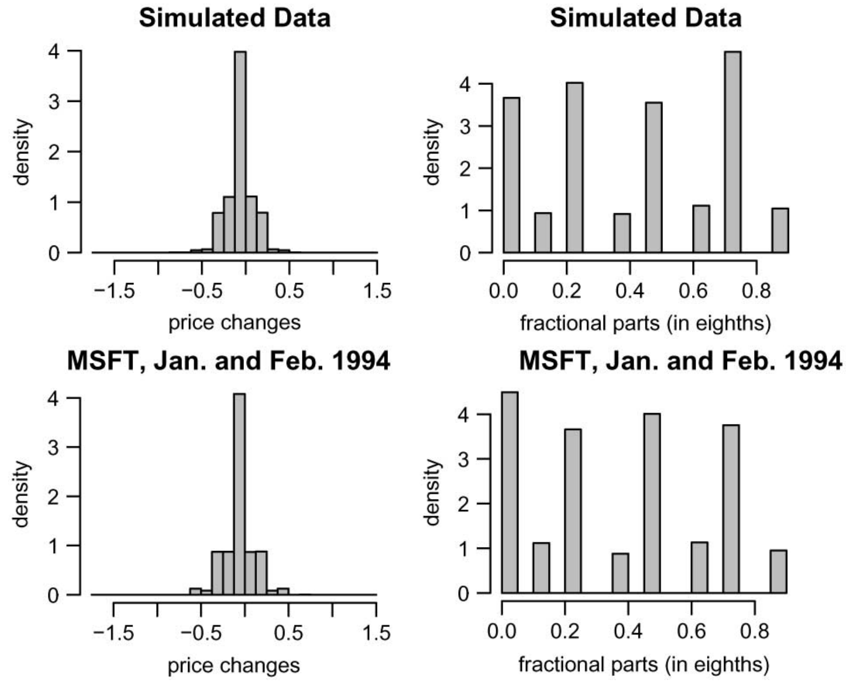

Fig. 1. Histograms of price changes and of fractional parts: Simulated data and Microsoft data.

the intertrading time. Case 1, if $t_{i+1} - t_i \leq LL$, the length controller, then we can approximate $p_\varepsilon(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w; t_{i+1}-)$ as

$$\begin{aligned} p_\varepsilon(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w; t_{i+1}-) \\ \approx p_\varepsilon(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w; t_i) \\ + [a(\mu_j, \sigma_k, x_{w-1}) p_\varepsilon(\mu_j, \sigma_k, \lambda_l, \rho_m, x_{w-1}; t_i) \\ - (a(\mu_j, \sigma_k, x_w) + b(\mu_j, \sigma_k, x_w)) \\ \times p_\varepsilon(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w; t_i) \\ + b(\mu_j, \sigma_k, x_{w+1}) p_\varepsilon(\mu_j, \sigma_k, \lambda_l, \rho_m, x_{w+1}; t_i) \\ + \lambda_l(\bar{p}_\varepsilon(\mu_j, \lambda_l, \rho_m, x_w; t) \\ - p_\varepsilon(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w; t))](t_{i+1} - t_i). \end{aligned} \quad (13)$$

Case 2, if $t_{i+1} - t_i > LL$, then we can choose a fine 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 approximate $p(\mu_j, \sigma_k, \rho_m, x_l; t_{i+1}-)$ by applying repeatedly the recursive algorithm given by (13) from $t_{i,0}$ to $t_{i,1}$, then $t_{i,2}, \ldots$, until $t_{i,n} = t_{i+1}$.

Equations (12) and (13) consist of the recursive algorithm we employ to calculate the approximate posterior at time t_{i+1} for $(\mu, \sigma(t_{i+1}), \lambda, \rho, X(t_{i+1}))$ based on the posterior at time t_i.

Finally, we choose a reasonable prior. We assume the independence between $X(0)$ and $(\mu, \sigma(0), \lambda, \rho)$. 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 $(\mu, \sigma(0), \lambda, \rho)$ available, we may simply assign uniform distributions to the latticized state–space of $(\mu, \sigma(0), \lambda, \rho)$ and obtain the prior at $t = 0$ as

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

TABLE I
BAYES ESTIMATES FOR A SIMULATED DATA

Parameter	True Value	Bayes Est.	S.E.
μ	4.5e-8	5.555e-7	3.874e-7
λ	3.75e-4	4.733e-4	7.149e-5
ρ	0.20	0.2018	.0013

1) Consistency of the Recursive Algorithm: There are two approximations in the construction of the recursive algorithm. The first is to approach the time integral in the propagation (11) by an Euler scheme, whose convergence is well-known. The second is more important. It is the approximation of the filtering equation (3) (the optimum filter) by the filtering equation (7) of the approximate model (the approximate optimum filter). Since $(\vec{\theta}_{\vec{\varepsilon}}, X_{\epsilon_x}) \Rightarrow (\vec{\theta}, X)$ by construction, Theorem 3.2 guarantees the convergence of the filtering equation (7) to filtering equation (3) in the sense of weak convergence in the Skorohod topology.

IV. SIMULATION AND REAL DATA EXAMPLES

The consistent (or robust) recursive algorithm for computing the joint posterior and Bayes estimates, the margin posterior means, is extensively tested and validated on simulated data. One simulation example is provided to demonstrate that the Bayes estimates for stochastic volatility are able to capture the moving of volatility quickly and are close to the true values. Then, the recursive algorithm is applied to two months of transaction prices of Microsoft and the stochastic volatility estimates can be obtained in real-time.

A. Simulation Study

Extensive simulation studies are done and one of them is reported as follows.

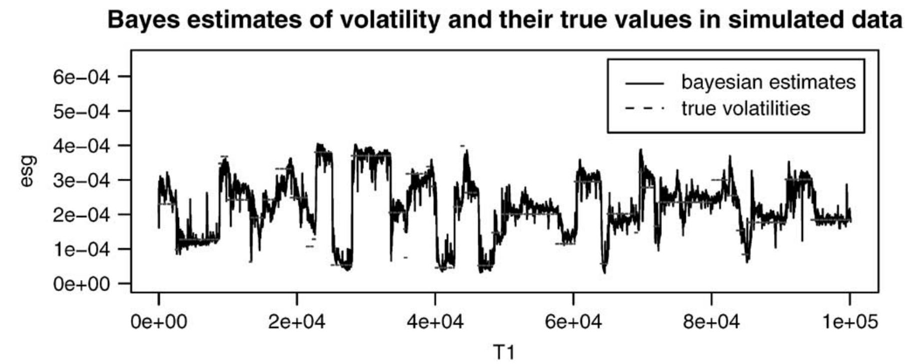

Fig. 2. Bayes estimates of volatility for 90 000 simulated data.

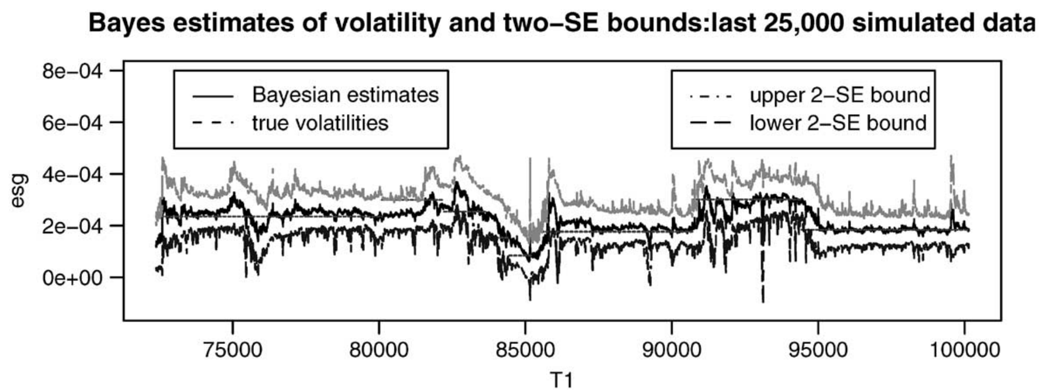

Fig. 3. Bayes estimates of volatility with two-SE bounds for the last 25 000 simulated data.

1) Simulated Data and Its Micromovement Features: In the following example, these parameters are picked for the jumping stochastic volatility models: $\mu = 4.5 \times 10^{-8}$, 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 in $1/3.75 \times 10^{-4} = 2666.67$ seconds on average. The range of volatility is [0.000 04, 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 in about $1/0.9 = 1.11$ s). For the parameters of noise, in order to show the model can produce the micromovement features of actual transaction price data, I choose the parameters close to those of the Microsoft data set, which is discussed in Section IV-B. Let $\rho = 0.2$, $\alpha = 0.4$, and $\beta = .2$. Using these parameters, 90 000 observations are simulated.

Fig. 1 has two pair of density histograms of price changes and of the fractional parts of price. The upper pair are produced from a simulated data and the lower pair from the actual Microsoft data. Their similarity shows that by incorporating discrete, non-clustering, and clustering noises, the proposed model possesses the micromovement features of the actual prices.

2) Bayes Estimates for the Simulated Data: 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.

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 studied in [18]. Hence, only the final Bayes estimates, their SE, and true values are presented in Table I. The true values are close to the Bayes estimates and all within two SE bounds.

Estimating stochastic volatility is the focus of this paper. Fig. 2 shows how the Bayesian estimates of volatility evolve in comparison of the time-varying true values of volatility, $\sigma(t)$, for all 90 000 data. Overall, we see that the Bayes estimates of volatility are close to their true values. As the true volatility changes, the Bayes estimates catch up with the movement quickly. Fig. 3 presents them for about the last 25 000 data. We observe that most of the true values of stochastic volatility are within the two SE bounds.

TABLE II
BAYES ESTIMATES FOR TRANSACTION DATA OF MSFT, JANUARY AND FEBRUARY 1994

Model	μ	λ	ρ
GBM			
with	12.89%	5176.79	0.2070
JSV	(24.91%)	(392.72)	(.0024)
GBM			
only	23.01%	NA	0.2226
	(83.27%)	NA	(0.0017)

B. An Application to Real Data

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

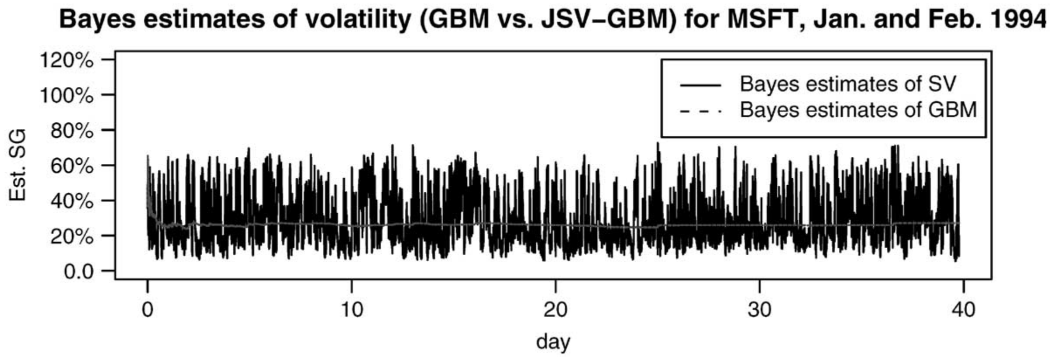

Fig. 4. Bayes estimates of volatility (GBM versus JSV-GBM) for MSFT data

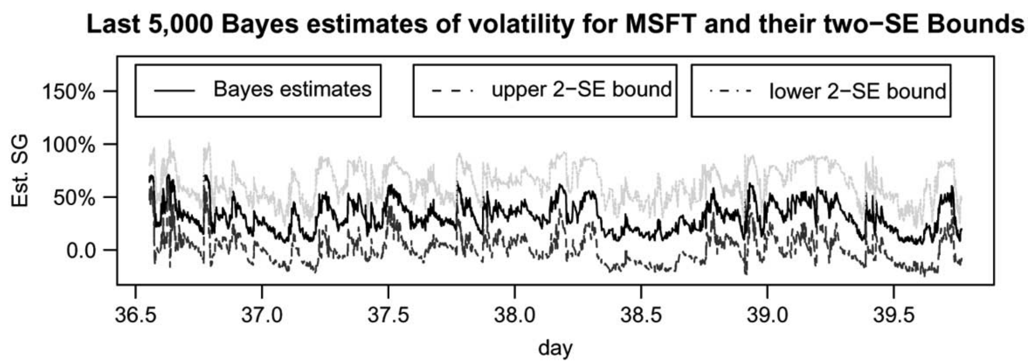

Fig. 5. Bayes estimates of volatility with two-SE bounds for the last 5000 MSFT data.

1) The Data: The data are extracted from the trade and quote (TAQ) database distributed by NYSE. We apply standard procedures to filter the data but with one important exception. Previous studies cannot handle multiple trades at a given point in time and they exclude those trades with zero time duration. Autoregressive conditional duration model proposed by [7] is such an example. The present method can handle such cases. 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, a variant of the method of moments, 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.

2) Bayes Estimates for Volatility: JSV-GBM Versus GBM: The recursive algorithm for JSV-GBM is applied to the transaction data of Microsoft. For comparison, the recursive algorithm for GBM is applied to the same data set to obtain Bayes estimates.

Table II presents the final annualized (with an annual factor of 260) Bayes estimates and their SE bounds for μ, λ, ρ for JSV-GBM, and for GBM. For JSV-GBM, the jump intensity, λ, is 5176.79, which means there are 5176.79 changes in volatility annually, or about 20 times daily. It is observed that intraday volatility is "U-shaped" [17]: higher volatility in opening and closing of the market. So volatility changes at least twice a day. The large λ is consistent with the observation and indicates the volatility changes even more frequently. The parameter for nonclustering noise, ρ, is reduced significantly in JSV-GBM, probably because more price variation is explained in stochastic volatility.

Fig. 4 shows how the Bayesian estimates of volatility for JSV-GBM and GBM evolve and demonstrate how different the volatility estimates are for the two models. In GBM, 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 Bayesian model selection via filtering based on Bayes factor developed in [14]. To avoid crowdedness in picture, Fig. 5 presents only the last 5 000 Bayesian estimates of volatility for JSV-GBM and their estimated two-SE bounds. We can see the volatility estimates vary greatly. Sometimes it moves continuously and sometimes it jumps just as the model suggests. Overall, the smallest volatility estimate is 5.424% annually, and the largest is 72.58%.

V. Conclusion

A unified Bayesian estimation via filtering approach is developed for estimating stochastic volatility for a class of micromovement models, which capture the impact of noise at

$$p(y\,|\,x) = \begin{cases} (1-\alpha-\beta)(1-\rho), & \text{if } r(y)=3 \text{ and } D=0 \\ \frac{1}{2}(1-\alpha-\beta)(1-\rho)\rho^D, & \text{if } r(y)=3 \text{ and } D\geq 0 \\ (1-\rho)(1+\alpha\rho), & \text{if } r(y)=2 \text{ and } D=0 \\ \frac{1}{2}(1-\rho)[\rho+\alpha(2+\rho^2)], & \text{if } r(y)=2 \text{ and } D=1 \\ \frac{1}{2}(1-\rho)\rho^{D-1}[\rho+\alpha(1+\rho^2)], & \text{if } r(y)=2 \text{ and } D\geq 2 \\ (1-\rho)(1+\beta\rho), & \text{if } r(y)=1 \text{ and } D=0 \\ \frac{1}{2}(1-\rho)[\rho+\beta(2+\rho^2)], & \text{if } r(y)=1 \text{ and } D=1 \\ \frac{1}{2}(1-\rho)\rho^{D-1}[\rho+\beta(1+\rho^2)], & \text{if } r(y)=1 \text{ and } D\geq 2 \end{cases}. \tag{15}$$

the microlevel. The class of models has an important feature in that it can be formulated as a filtering problem with counting process observations. Under this formulation, the whole sample paths are observable, and the complete tick data information is used in Bayes parameter estimation via filtering. A consistent recursive algorithm is developed to compute the Bayes estimates for the parameters in the model, especially, the stochastic volatility. Simulation studies show that Bayes estimates for time-invariant parameters are consistent, and Bayes estimates for stochastic volatility are close to their true values and are able to capture the movement of volatility quickly. The recursive algorithm is fast and feasible for large data sets and it has the recursive feature allowing quick and easy update. The recursive algorithm is applied to Microsoft's transaction data and we obtain Bayes estimates and provide strong affirmative evidence that volatility changes even more dramatically in trade-by-trade level. The model and its Bayes estimation via filtering equation can be extended to jump-diffusion process for the value process, and other kinds of noise according to the sample characteristics of data. The models and the Bayes estimation can be applied to other asset markets such as exchange rates and commodity prices. It can also apply to assess the quality of security market, and to compare information flows and noises in different periods and different markets.

APPENDIX A

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

$$r(y) = \begin{cases} 3, & \text{if the fractional part of y is odd eighth} \\ 2, & \text{if the fractional part of y is odd quarter} \\ 1, & \text{if the fractional part of y is a half or zero} \end{cases}. \tag{14}$$

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 $p(y\,|\,x) = \sum_{x'} p(y\,|\,x')p(x'\,|\,x)$ where $p(y'\,|\,x) = P\{V = 8(y' - R[x,(1/8)])\}$. Suppose $D = 8|y - R[x,(1/8)]|$. Then, $p(y\,|\,x)$ can be calculated as (15), shown at the top of the page.

ACKNOWLEDGMENT

The author would like to thank T. G. Kurtz for his inspiration and the insightful and productive conversations. An important part of this work was done when the author visited Prof. M. Kouritzin at the University of Alberta, Canada, during the summer of 2002; the author would like to thank him for the hospitality and productive conversations. He also would like to thank Guest Editor B. Pasik-Duncan and the two referees for detailed comments and suggestions on an earlier version, improving the exposition of this manuscript.

REFERENCES

[1] C. Ball, "Estimation bias induced by discrete security prices," *J. Finance*, vol. 43, pp. 841–865, 1988.
[2] T. Bollerslev, R. Chou, and K. Kroner, "ARCH modeling in finance: A review of the theory and empirical evidencs," *J. Econometrics*, vol. 52, pp. 5–59, 1992.
[3] P. Bremaud, *Point Processes and Queues: Martingale Dynamics*. New York: Springer-Verlag, 1981.
[4] D. C. Cho and E. W. Frees, "Estimating the volatility of discrete stock prices," *J. Finance*, vol. 43, pp. 451–466, 1988.
[5] P. Del Moral, J. Jacod, and P. Protter, "The Monte-Carlo method for filtering with discrete time observations," *Probab. Theory Related Fields*, vol. 120, pp. 346–368, 2001.
[6] R. J. Elliott, W. P. Malcolm, and A. H. Tsoi, "Robust parameter estimation for asset price models with Markov modulated volatilities," *J. Econ. Dyna. Control*, vol. 27, pp. 1391–1409, 2003.
[7] R. Engle and J. Russell, "Autoregressive conditional duration: A new model for irregularly spaced transaction data," *Econometrica*, vol. 66, pp. 1127–1162, 1998.
[8] G. Gottlieb and A. Kalay, "Implications of the discreteness of observed stock prices," *J. Finance*, vol. 40, pp. 135–153, 1985.
[9] L. Harris, "Estimation of stock price variances and serial covariances from discrete observations," *J. Finan. Quantitat. Anal.*, vol. 25, pp. 291–306, 1990.
[10] L. Harris, "Stock price clustering and discreteness," *Rev. Finan. Stud.*, vol. 4, pp. 389–415, 1991.
[11] *Handbook of Statistics*, vol. 14, G. S. Maddala and C. R. Rao, Eds., North-Holland, Amsterdam, The Netherlands, 1996, pp. 647–692.
[12] S. Heston, "A closed-form solution for options with stochastic volatility with applications to bond and currency options," *Rev. Finan. Stud.*, vol. 6, pp. 327–343, 1993.
[13] J. Hull and A. White, "The pricing of options on assets with stochastic volatilities," *J. Finance*, vol. 42, pp. 281–300, 1987.
[14] M. Kouritzin and Y. Zeng, Bayesian model selection via filtering for a class of micro-movement models of asset prices, Univ. Missouri at Kansas City, Kansas City, MO, 2003.
[15] H. J. Kushner and P. G. Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time*. New York: Springer-Verlag, 1994.
[16] D. B. Nelson, "ARCH models as diffusion approximations," *J. Econometrics*, vol. 45, pp. 7–39, 1990.
[17] M. O'Hara, *Market Microstructure Theory*. Oxford, U.K.: Blackwell, 1995.
[18] Y. Zeng, "A partially observed model for micromovement of asset prices with Bayes estimation via filtering," *Math. Finance*, vol. 13, pp. 411–444, 2003.

Yong Zeng received the B.S. degree in statistics from Fudan University, Shanghai, China, in 1990, the M.S. degree in statistics from the University of Georgia, Athens, in 1994, and the Ph.D. degree in statistics from the University of Wisconsin at Madison, in 1999.

Subsequently, he joined the Department of Mathematics and Statistics, the University of Missouri at Kansas City, as an Assistant Professor. He visited the University of Alberta, Canada, in summers 2002 and 2003, and participated in the Summer Probability Internship at the University of Wisconsin at Madison in 2003. His research interests include marked point process modeling of and Bayesian inference (estimation and model selection) for transaction data, filtering with marked point process observations, and modeling the term structure of interest rates.