

Estimating Stochastic Volatility via Filtering for the Micro-movement of Asset Prices *

Yong Zeng[†]

Assistant Professor

Department of Mathematics and Statistics,

University of Missouri at Kansas City

November 4, 2002

*I am grateful to Thomas G. Kurtz for his inspiration and the insightful and productive conversations. The major part of the paper was done when the author visited Professor Michael Kouritzin in University of Alberta during the summer of 2002. I thank Mike for the hospitality and many productive conversations. I thank participants at the Workshop on Filtering Theory and Applications (2002) held at University of Alberta, the IMS Annual Meeting/4th International Probability Symposium (2002) held at Banff, Canada, and seminar participants at University of Alberta, University of Kansas, and University of Missouri at Kansas City for helpful comments. This work is partially supported by Faculty Research Grant of the University of Missouri at Kansas City.

[†]Department of Mathematics and Statistics, University of Missouri at Kansas City, Kansas City, MO 64110, USA. Tel No: (816) 235 5850 and e-mail: zeng@mendota.umkc.edu

Estimating Stochastic Volatility via Filtering for the Micro-movement of Asset Prices

ABSTRACT

Under the general framework of Zeng (2002), a unified approach via filtering is developed to estimate stochastic volatility for micro-movement models. The key feature of the models is that they can be transformed as filtering problems with counting process observation. In order to obtain trade-by-trade, real-time Bayes estimates of stochastic volatility, the Markov chain approximation method is applied to filtering equation to construct consistent recursive algorithms, which compute the joint posterior. To illustrate the approach, the recursive algorithm is constructed in detail for a jumping stochastic volatility micro-movement 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.

KEY WORDS: Filtering, counting process, Markov chain approximation, Stochastic Volatility, High-frequency data

1 Introduction

Stochastic volatility is well documented for asset prices in both macro-movement and micro-movement (Bollerslev, Chou & Kroner (1992)). Macro-movement refers to daily, weekly, and monthly closing price behavior while micro-movement refers to transactional (trade-by-trade) price behavior. There are the strong connection as well as striking distinctions between the macro- and micro-movements. The strong connection is observed through the identity of the overall shapes of both, because the macro-movement is an equally-spaced time series drawn from the micro-movement data. Their striking distinction are mainly due to financial noise. In macro-movement, the impact of noise is small and is usually neglected. In micro-movement, 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 Gottlieb & Kalay (1985), Ball (1988) and Cho & Frees (1988) for discrete noise and further in Harris (1990) and Zeng (1999 and 2002) for discrete plus other types of noise.

Economically, asset price is distinguished from its intrinsic value and their distinction is also noise. Noise, as contrasted with information, is well-documented in the market microstructure literature. Black (1986) noted that noise renders our observations of value imperfect by creating a wedge between price and intrinsic value, and provides an impetus for trading in financial markets. Hasbrouck (1996) pointed out that information has a long-term, permanent impact while noise has a short-term, transitory impact on price. Three important types of noise have been identified and extensively studied: discrete, clustering and non-clustering. First, intraday prices move discretely (tick by tick), resulting in “discrete noise”. Second, because prices do not distributed evenly on all ticks, but gather more on some ticks such as integers and halves, “price clustering” is obtained. Harris (1991) confirms that this phenomenon is remarkably persistent through time, across assets,

and across market structures. Third, the “non-clustering noise” includes other unspecified noise, and the outliers in prices are one of the evidence for the existence of non-clustering noise.

In Zeng (2002), a novel, economically well-grounded, partially-observed micro-movement 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 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, non-stochastic volatility, micro-movement models is extremely challenging. Zeng (2002) develops continuous-time Bayes estimation via filtering with efficient algorithms for the parameter estimation of the micro-movement model. That represents a significant advance in the estimation for micro-movement 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 Zeng (2002), however, only the parameters of a simple model with GBM as value process is estimated.

In this paper, first, a class of stochastic volatility micro-movement models is developed from the macro-movement models by incorporating the three types of noise summarized. A new, jump-

ing stochastic volatility (JSV) micro-movement model, stemmed from geometric Brownian motion (GBM), is proposed and studied (later, it is called JSV-GBM micro-movement model). Second, a unified approach, Bayes parameter estimation via filtering, is developed for the micro-movement 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.

JSV-GBM model is employed to demonstrate the effectiveness of estimating stochastic volatility using Bayes estimation via filtering. To illustrate the approach, JSV-GBM’s 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 micro-movement.

The rest of the article proceeds as follows. Section 2 presents the class micro-movement models with stochastic volatility. Section 3 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 4 presents simulation and estimation results based on actual data. Concluding remarks and future works are offered in Section 5.

2 The Stochastic Volatility Micro-movement Models

The micro-movement model is built upon the macro-movement 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 $\frac{1}{M}$. The combination of (X, Y) provides a natural partially-observed framework for the micro-movement process. So, we start with stochastic volatility models for value process.

2.1 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 Zeng (2002).

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 $(\theta(s), X(s))_{0 \leq s \leq t}$.

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

Example 2.1 Hull & White (1987)'s model in SDE form is

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

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

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

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 (Nelson (1990)).

Example 2.2 Heston (1993)'s model in SDE form is

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

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

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

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

The following example presents a new JSV-GBM model for value process. Its micro-movement version is developed in Section 2.2 and further in Section 3.2.

Example 2.3 A jumping stochastic volatility model in SDE form is

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

where $N(t)$ is a Poisson process with intensity λ , and $\{J_i\}$ is a sequence of i.i.d. random variables with density $g(z)$, and is assumed to be independent of $W(t)$ and $N(t)$. When the i th 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 exist. This model is closely related to the asset price models with Markov modulated volatilities studied by Elliot, Malcolm & Tsoi (2002).

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.

2.2 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 function with the transition probability $p(y|x)$.

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

This article focuses on estimating this class of micro-movement 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 has stochastic volatility, but also can capture the micro-movement features created by noise.

Below, a simple $F(x)$ is built to accommodate the three types of noise that are well-documented in financial literature: discrete noise, clustering noise, and non-clustering noise.

To simplify notation, fix i and set $x = X(t_i)$, $y = Y(t_i)$, and $y' = Y'(t_i) = R[X(t_i) + V_i, \frac{1}{M}]$, where V_i is defined as the non-clustering noise. Instead of directly specifying $p(y|x)$ 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, $R[x, \frac{1}{M}]$. Without other noises, trades should occur at this tick, which is closest to the stock value.

Step 2: Incorporate non-clustering noise by adding V ; $y' = R[x + V, \frac{1}{M}]$, where V is the non-clustering noise of trade i at time t_i .

We assume $\{V_i\}$, are independent of the value process, and they are *i.i.d.* with a doubly geometric distribution:

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

We pick the doubly geometric distribution because it is uni-modal, symmetric and bell shaped with the implication that the trading price at a tick closer to the stock value is more likely to occur and trading prices with the same distance to the stock value have the same chances.

Non-clustering noise is intended to explain three discreteness-related sample characteristics of transactional data that cannot be explained by the rounding model. First, the non-clustering noise considerably increases the probability of the successive price changes that are more than a tick. Second, it allows the prices of trades occurring within the same second to differ and the difference can be two or more ticks. Finally, it produces outliers.

Step 3: Incorporate clustering noise by biasing y' . After rounding the value process and adding non-clustering 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 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. 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

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

In this way, the random function of $F(x)$, which models the impact of financial noise, is specified. The rounding function takes care of price discreteness, V_i takes care of non-clustering noise and the random biasing function b_i takes care of clustering noise. The transition probability $p(y|x)$ of F can be computed and the explicit $p(y|x)$ is given in Appendix A.

2.3 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:

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

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 k th 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. Hence, (θ, 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 to the counting process observations in Equation (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 .

Remark 2.2 These two assumptions imply that there exists a reference measure Q and that after a suitable change of measure to Q , $(\theta, X), Y_1, \dots, Y_n$ will become independent, and Y_1, Y_2, \dots, 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 Bremaud (1981), page 165).

Assumption 2.4 *The intensity, $\lambda_k(\theta, x, t) = a(\theta, x, t)p(y_k|x)$, where $a(\theta, x, t)$ is the total intensity at time t and $p(y_k|x)$ is the transition probability from x to y_k , the k th price level.*

Remark 2.3 This assumption 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 $p(y_k|x)$ 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, because the conditional finite dimensional distributions of the price Y given $\{X(s) : 0 < s < t\}$ can be shown to be identical in these two setups.

A final technical assumption ensures the uniqueness of the filtering equation.

Assumption 2.5 *The total intensity, $a(\theta, x, t)$, is uniformly bounded from below and above; i.e., there exist constants, C_1 and C_2 , such that $C_1 < a(\theta, x, t) \leq C_2$ for all $t > 0$.*

3 Bayesian Estimation via Filtering

Bayes estimate, which is the posterior mean, is the least Mean Square Errors (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 consistent, namely, the conditional distribution computed by the recursive algorithm converges to the true one determined by the filtering equation. 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.

3.1 Review

The following two results are from Zeng (2002).

3.1.1 Filtering Equation

Before the filtering equation is presented, some terms are defined below. Let $\mathcal{F}_t^{\vec{Y}} = \sigma\{(\vec{Y}(s)) | 0 \leq s \leq t\}$ 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 $(\theta(0), X(0))$, π_t becomes the joint posterior of $(\theta(t), X(t))$.

Definition 3.2 Let $\pi(f, t) = E^P[f(\theta(t), X(t)) | \mathcal{F}_t^{\vec{Y}}] = \int f(\theta, x) \pi_t(d\theta, dx)$ 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 Equation (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,

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

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

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

Note that $a(t)$ disappears in Equation (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, \dots , then Equation (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 k th 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.

3.1.2 A Convergence Theorem on Conditional Expectation

Suppose the state space of (θ, X) is discretized with ϵ_i as the length between lattices in the i th component of θ and ϵ_x as that of X . Let $\epsilon = (\epsilon_1, \dots, \epsilon_p)$. Then, $(\theta_\epsilon, X_{\epsilon_x})$, an approximation for

(θ, X) , can be constructed. Define

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

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

Theorem 3.2 *Suppose that (θ, X, \vec{Y}) is on the probability space (Ω, \mathcal{F}, P) with Assumptions 2.1 - 2.5. \vec{Y}_ε is defined by (4). Suppose that $(\theta_\varepsilon, X_{\varepsilon_x}, \vec{Y}_\varepsilon)$ is on $(\Omega_\varepsilon, \mathcal{F}_\varepsilon, P_\varepsilon)$, and Assumptions 2.1 - 2.5 also hold for $(\theta_\varepsilon, X_{\varepsilon_x}, \vec{Y}_\varepsilon)$. If $(\theta_\varepsilon, X_{\varepsilon_x}) \Rightarrow (\theta, X)$ as $\varepsilon = \max\{\varepsilon_x, |\varepsilon|\} \rightarrow 0$, then*

(i) $\vec{Y}_\varepsilon \Rightarrow \vec{Y}$ as $\varepsilon \rightarrow 0$; and

(ii) $E^{P_\varepsilon}[F(\theta_\varepsilon(t), X_{\varepsilon_x}(t))|\mathcal{F}_t^{\vec{Y}_\varepsilon}] \Rightarrow E^P[F(\theta(t), X(t))|\mathcal{F}_t^{\vec{Y}}]$ as $\varepsilon \rightarrow 0$ for function F in the domain of the generator **A**.

Remark 3.3 This theorem provides a recipe for constructing a consistent recursive algorithm.

It guarantees that as long as $(\theta_\varepsilon, X_{\varepsilon_x})$ is an approximation of (θ, X) , Y_ε is an approximation of \vec{Y} and the conditional expectation of $(\theta_\varepsilon, X_{\varepsilon_x})$ is an approximation of (θ, X) given the observed sample paths of the counting processes. When we take ‘‘F’’ as an appropriate indicator function, $E^{P_\varepsilon}[F(\theta_\varepsilon(t), X_{\varepsilon_x}(t))|\mathcal{F}_t^{\vec{Y}_\varepsilon}]$ becomes the conditional probability mass function for $(\theta_\varepsilon, X_{\varepsilon_x})$. 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.

3.2 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 Hull & White (1987) and Heston (1993) do. Their difference is in the volatility dynamics. In Hull & White (1987) and Heston (1993), 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 becomes

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

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, 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.

3.3 Construction of the Recursive Algorithm

For the nonlinear filtering problem, Kushner (See Chapter 12 of Kushner & Dupuis (1994)) 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 Equation (3), which is the case when the trading intensity is deterministic.

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, $\mu : [\alpha_\mu, \beta_\mu] \rightarrow \{\alpha_\mu, \alpha_\mu + \epsilon_\mu, \alpha_\mu + 2\epsilon_\mu, \dots, \alpha_\mu + j\epsilon_\mu, \dots, \alpha_\mu + n_\mu\epsilon_\mu\}$ where $\alpha_\mu + n_\mu\epsilon_\mu = \beta_\mu$ and the number of lattices is $n_\mu + 1$. Define $\mu_j = \alpha_\mu + j\epsilon_\mu$, the j th lattice in the latticized parameter space of μ . Similarly, define $\sigma_k = \alpha_\sigma + k\epsilon_\sigma, \lambda_l = \alpha_\lambda + l\epsilon_\lambda, \rho_m = \alpha_\rho + m\epsilon_\rho$, 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. But, Markov chain is a natural approximation for the stochastic process $(\sigma(t), X(t))$.

Second, it is observe that the construction of an approximate Markov chain can be transformed to construct a Markov chain generator, \mathbf{A}_ε , such that as $\varepsilon \rightarrow 0$. $\mathbf{A}_\varepsilon \rightarrow \mathbf{A}$, the one given by Equation (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}_ε as follows:

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

where

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

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

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

for the approximate model $(\theta_\epsilon, X_{\epsilon_x}, \vec{Y}_\epsilon)$, 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}_ϵ , \vec{Y} by \vec{Y}_ϵ , and there also exists a probability measure P_ϵ to replace P , then Assumptions 2.1 - 2.5 also hold for $(\theta_\epsilon, X_{\epsilon_x}, \vec{Y}_\epsilon)$. 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 $\pi_{\epsilon,t}$ be the conditional probability mass of $(\mu_\epsilon, \sigma_\epsilon, \lambda_\epsilon, \rho_\epsilon, X_\epsilon(t))$ given $\mathcal{F}_t^{\vec{Y}_\epsilon}$.

Definition 3.4 Let

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

where $(\mu, \sigma, \lambda, \rho, x)$ covers all lattices in the discretized state space, be the conditional expectation of $f(\mu_\epsilon, \sigma_\epsilon, \lambda_\epsilon, \rho_\epsilon, X_\epsilon(t))$ given $\mathcal{F}_t^{\vec{Y}_\epsilon}$.

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

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

where $Y_{\epsilon,k}$ is the k -th component of \vec{Y}_ϵ .

Similarly, the above filtering equation can be separated into the propagation equation:

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

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

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

Step 3: Convert Equations (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, $(\mu_\varepsilon, \sigma_\varepsilon(t), \lambda_\varepsilon, \rho_\varepsilon, X_\varepsilon(t), Y_\varepsilon(t))$, at time t is denoted by*

$$p_\varepsilon(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w; t) = P\{\mu_\varepsilon = \mu_j, \sigma_\varepsilon(t) = \sigma_k, \lambda_\varepsilon = \lambda_l, \rho_\varepsilon = \rho_m, X_{\varepsilon_x}(t) = x_w \mid \mathcal{F}_t^{\bar{Y}_\varepsilon}\}.$$

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

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

Then, the following facts emerge:

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

and

$$\pi_\varepsilon(a(\mu_\varepsilon, \sigma_\varepsilon, X_\varepsilon)\mathbf{I}(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w + \varepsilon_x)) = a(\mu_j, \sigma_k, x_{w-1})p_\varepsilon(\mu_j, \sigma_k, \lambda_l, \rho_m, x_{w-1}; t).$$

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

$$\begin{aligned}\pi_\varepsilon(\mathbf{A}_\varepsilon \mathbf{I}, t) = & a(\mu_j, \sigma_k, x_{w-1})p_\varepsilon(\mu_j, \sigma_k, \lambda_l, \rho_m, x_{w-1}; t) \\ & + b(\mu_j, \sigma_k, x_{w+1})p_\varepsilon(\mu_j, \sigma_k, \lambda_l, \rho_m, x_{w+1}; t) \\ & - (a(\mu_j, \sigma_k, x_w) + b(\mu_j, \sigma_k, x_{w+1}))p_\varepsilon(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w; t) \\ & + \lambda_l(\bar{p}(\mu_j, \lambda_l, \rho_m, x_w; t) - p(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w; t)),\end{aligned}$$

where

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

Therefore, the propagation equation in Equation (8) becomes

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

If a trade at k th price level occurs at time t_{i+1} , the updating Equation (9) can be written as,

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

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 Equation (15) in Appendix A.

Next, we make Equation (11) a recursive algorithm. Equation (11) is deterministic and 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 the inter-trading time. Case 1, if $t_{i+1} - t_i \leq LL$, the length controller, then we can approximate $p(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w; t_{i+1}-)$ as

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

Case 2, if $t_{i+1} - t_i > LL$, then we can choose a fine partition $\{t_{i,0} = t_i, t_{i,1}, \dots, t_{i,n} = t_{i+1}\}$ of $[t_i, t_{i+1}]$ such that $\max_j |t_{i,j+1} - t_{i,j}| < LL$ and then approximate $p(\mu_j, \sigma_k, \rho_m, \nu_l; t_{i+1}-)$ by applying repeatedly the recursive algorithm given by Equation (13) from $t_{i,0}$ to $t_{i,1}$, then $t_{i,2}, \dots$, 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,

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

3.3.1 Consistency of the Recursive Algorithm

There are two approximations in the construction of the recursive algorithm. The first is to approach the integral in the propagation equation (11) by 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 $(\vec{\theta}_\varepsilon, v_{\varepsilon_v}) \Rightarrow (\vec{\theta}, v)$ 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. By taking f as the indicator (defined in Equation (10)) in Equation (7), one obtains the recursive algorithm to compute $p_\varepsilon(\mu_j, \sigma_k, \rho_m, v_l; t)$. According to Theorem 3.2, $p_\varepsilon(\mu_j, \sigma_k, \rho_m, v_l; t)$ converges to an expression related to the true posterior (see below). In order to identify this expression, we first define some neighborhoods.

Define the neighborhood of μ_j as $N_{\mu_j} = \{\mu_j - 0.5\varepsilon_\mu \leq \mu < \mu_j + 0.5\varepsilon_\mu\}$ and similarly for the neighborhoods of $\sigma_k, \lambda_l, \rho_m$ and x_w .

Definition 3.6 *The true posterior in the neighborhood of $(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w)$ is defined as*

$$p(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w; t) = E^P \left[\mathbf{I}_{\{\mu \in N_{\mu_j}, \sigma(t) \in N_{\sigma_k}, \lambda \in N_{\lambda_l}, \rho \in N_{\rho_m}, X(t) \in N_{x_w}\}}(\mu, \sigma(t), \lambda, \rho, X(t)) | \mathcal{F}_t^{\vec{Y}} \right].$$

Note that the indicator function defined in the above $p(\mu_j, \sigma_k, \lambda_l, \rho_m, x_w; t)$ becomes the indicator function defined in Equation (10) in the approximate model. Therefore, Theorem 3.2 implies $p_\varepsilon(\mu_j, \sigma_k, \rho_m, v_l; t) \rightarrow p(\mu_j, \sigma_k, \rho_m, v_l; t)$ as $\varepsilon \rightarrow 0$.

4 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 in detail to two months of transaction prices of Microsoft and the stochastic volatility estimates can be obtained in real-time.

4.1 Simulation Study

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

4.1.1 Simulated Data and its Micro-movement 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 averagely. 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 in about $1/0.9 = 1.11$ seconds). For the parameters of noise, in order to show the model can produce the micro-movement features of actual transaction price data, I choose the parameters close to those of the Microsoft data set, which is discussed in Section 4.2. Let $\rho = 0.2$, $\alpha = 0.4$, and $\beta = .2$. Using these parameters, 90,000 observations are simulated.

Figures 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 micro-movement features of the actual prices.

4.1.2 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, 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 in Zeng (2002). Hence, only the final Bayes estimates, their SE, and true values are presented in Table 1. The true values are close to the Bayes estimates and all within two SE bounds.

Estimating stochastic volatility is the focus of this paper. Figure 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. Figure 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.

4.2 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.

4.2.1 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 Engle & Russell (1998) is such an example. The present method can handle such cases and we keep all zero durations. 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 method of moments, to estimate $\alpha = .2414$, and $\beta = .3502$. 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 seconds with a standard error 30.01 seconds.

4.2.2 Bayes Estimates for Volatility: JSV-GBM vs. 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 2 presents the final annualized (with annual factor 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 has “U-shaped” (O’Hara (1995)): 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 frequent. The parameter for non-clustering noise, ρ , is reduced significantly in JSV-GBM, probably because more price variation is explained in

stochastic volatility.

Figure 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 model selection approach of Bayes Factor via filtering developed in Zeng & Kouritzin (2002). To avoid crowdedness in picture, Figure 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 varies 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%.

5 Conclusions

A unified Bayesian estimation via filtering approach is developed for estimating stochastic volatility for a class of micro-movement models, which capture the impact of noise in micro-level. 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.

A Appendix A

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

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

The biasing rules specify the transition probabilities from x' to y , $p(y|x')$ as below: $p(x'|x') = 1$, if $r(x') = 1$ or 2 ; $p(x'|x') = 1 - \alpha - \beta$, if $r(x') = 3$; $p(x' + \frac{1}{8}|x') = \alpha$, if $x' - \text{floor}(x') = \frac{1}{8}$ or $\frac{5}{8}$; $p(x' - \frac{1}{8}|x') = \alpha$, if $x' - \text{floor}(x') = \frac{3}{8}$ or $\frac{7}{8}$; $p(x' - \frac{1}{8}|x') = \beta$, if $x' - \text{floor}(x') = \frac{1}{8}$ or $\frac{5}{8}$; $p(x' + \frac{1}{8}|x') = \beta$, if $x' - \text{floor}(x') = \frac{3}{8}$ or $\frac{7}{8}$. Note that the $\text{floor}(x)$ function gives the integer part of x .

Then, $p(y|x)$, the transition probability can be computed through $p(y|x) = \sum_{x'} p(y|x')p(x'|x)$ where $p(y|x')$ s are just defined above and $p(x'|x) = P\{U = 8(x' - R[x, \frac{1}{8}])\}$. Suppose $D = 8|p -$

$R[x, \frac{1}{8}]$. Then, $p(y|x)$ is

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

References

- Ball, C. (1988), ‘Estimation bias induced by discrete security prices’, *Journal of Finance* **43**, 841–865.
- Black, F. (1986), ‘Noise’, *Journal of Finance* **41**, 529–543.
- Bollerslev, T., Chou, R. & Kroner, K. (1992), ‘Arch modeling in finance: A review of the theory and empirical evidences’, *Journal of Econometrics* **52**, 5 – 59.
- Bremaud, P. (1981), *Point Processes and Queues: Martingale Dynamics*, Springer-Verlag, New York.
- Cho, D. & Frees, E. W. (1988), ‘Estimating the volatility of discrete stock prices’, *Journal of Finance* **43**, 451–466.
- Elliot, R., Malcolm, W. & Tsoi, A. H. (2002), ‘Robust parameter estimation for asset price models with markove modulated volatilities’. To appear in *Journal of Economic Dynamics and Control*.
- Engle, R. & Russell, J. (1998), ‘Autoregressive conditional duration: A new model for irregularly spaced transaction data’, *Econometrica* **66**, 1127–1162.
- Gottlieb, G. & Kalay, A. (1985), ‘Implications of the discreteness of observed stock prices’, *Journal of Finance* **40**, 135–153.
- Harris, L. (1990), ‘Estimation of stock price variances and serial covariances from discrete observations’, *Journal of Financial and Quantitative Analysis* **25**, 291–306.
- Harris, L. (1991), ‘Stock price clustering and discreteness’, *Review of Financial Studies* **4**, 389–415.

- Hasbrouck, J. (1996), Modeling market microstructure time series, in G. Maddala & C. Rao, eds, 'Handbook of Statistics', Vol. 14, Elsevier, pp. 647–692.
- Heston, S. (1993), 'A closed-form solution for options with stochastic volatility with applications to bond and currency options', *The Review of Financial Studies* **6**, 327–343.
- Hull, J. & White, A. (1987), 'The pricing of options on assets with stochastic volatilities', *Journal of Finance* **42**, 281–300.
- Kushner, H. & Dupuis, P. (1994), *Numerical Methods for Stochastic Control Problems in Continuous Time*, Springer-Verlag, New York.
- Nelson, D. B. (1990), 'Arch models as diffusion approximations', *Journal of Econometrics* **45**, 7–39.
- O'Hara, M. (1995), *Market Microstructure Theory*, Blackwell, Oxford.
- Zeng, Y. (2002), 'A partially-observed model for micro-movement of asset prices with bayes estimation via filtering'. to appear in *Mathematical Finance*.
- Zeng, Y. & Kouritzin, M. (2002), 'Bayesian model selection via filtering: Application to partially-observed micro-movement models of asset prices'.

Table 1: Bayes Estimates for a Simulated Data

Parameter	True Values	Bayes Est.	Standard Errors
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

Table 2: Bayes Estimates for Transaction Data of MSFT, Jan. and Feb. 1994

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

Bayes estimates of μ and λ are annualized with annual factor 260.
Standard errors are in parentheses