Mathematical Finance, Vol. 13, No. 3 (July 2003), 411–444

A PARTIALLY OBSERVED MODEL FOR MICROMOVEMENT OF ASSET PRICES WITH BAYES ESTIMATION VIA FILTERING

YONG ZENG

Department of Mathematics and Statistics, University of Missouri at Kansas City

A general micromovement model that describes transactional price behavior is proposed. The model ties the sample characteristics of micromovement and macromovement in a consistent manner. An important feature of the model is that it can be transformed to a filtering problem with counting process observations. Consequently, the complete information of price and trading time is captured and then utilized in Bayes estimation via filtering for the parameters. The filtering equations are derived. A theorem on the convergence of conditional expectation of the model is proved. A consistent recursive algorithm is constructed via the Markov chain approximation method to compute the approximate posterior and then the Bayes estimates. A simplified model and its recursive algorithm are presented in detail. Simulations show that the computed Bayes estimates converge to their true values. The algorithm is applied to one month of intraday transaction prices for Microsoft and the Bayes estimates are obtained.

KEY WORDS: high-frequency data, filtering, Bayes estimation, counting process, conditional distribution, Markov chain approximation, price discreteness, price clustering

1. INTRODUCTION

A large body of literature in financial economics and statistics is devoted to the modeling of asset price behavior. Existing models can be classified into two broad categories: macro- and micromovement models. Macromovement refers to daily, weekly, and monthly closing price behavior and micromovement refers to transactional (trade-by-trade) price behavior. In early studies, daily, weekly, and monthly asset closing prices were explored and popular models, such as geometric Brownian motion (GBM), diffusion, and jump diffusion models, were developed. In the late 1980s, transaction, or intraday, data became available and micromovement models emerged. To model the micromovement, one needs

This paper is a revision of the author's doctoral dissertation, Zeng (1999), at Department of Statistics, the University of Wisconsin at Madison, completed under the supervision of Thomas G. Kurtz. I am grateful to Tom for his intuition, inspiration, guidance, and encouragement. I thank Philip Protter for introducing me to the tick-by-tick data, thank Cyrus Ramezani and an associate editor for many detailed comments that improved the exposition, and thank Jie Chen, Chin-Shan Chuang, Robert Elliott, Larry Harris, Michael Kosorok, Yi Lin, Mark Ready, Chris Rogers, Boris Rozovskii, Ruey Tsay, and Kam-Wah Tsui for helpful comments. I thank participants at the Workshop on Stochastic Theory and Control (2001) held at University of Kansas, the Quantitative Risk Management in Finance Conference (2000) held at Carnegie Mellon University, the Symposium on Inference for Stochastic Processes (2000) held at the University of Georgia, and seminar participants at the University of Kansas, University of Missouri at Kansas City, University of Southern California, and University of Wisconsin at Madison for comments. A more detailed version of the paper is available from author upon request.

Manuscript received October 2001; final revision received May 2002.

Address correspondence to the author at the Department of Mathematics and Statistics, University of Missouri at Kansas City , Kansas City, MO 64110, USA; e-mail: zeng@mendota.umkc.edu.

to understand both the strong connection between the macro- and micromovements and their striking distinctions. As for the connection, because the macromovement is an equally spaced time series drawn from the micromovement data, the overall shape of the two are the same. This suggests that the models for the macro- and the micromovements should be closely related.

According to market microstructure theory, financial noise emanates from two sources. First, noise is the result of "noise trading" (Bagehot 1971; Black 1986) induced, for example, by transitory liquidity needs of traders and investors, and by other traders' mistreating "fully discounted information" as information. Second, noise reflects the impact of the trading mechanism by which prices are set in the market. This includes, for example, price discreteness,[1] price clustering,[2] the random arrival of buy or sell orders to the market (Mendelson 1982), the transitory state of the market maker's inventory positions (Amihud and Mendelson 1987; Hasbrouck 1988), the transient component of the price response to a block trade (Chan and Lakonishok 1993), and delayed price discovery (Cohen et al. 1980).

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 integer and half ticks, "price clustering" is obtained. Harris (1991) confirmed that this phenomenon is remarkably persistent through time, across assets, and across market structures. Third, the distribution of price changes[3] and the outliers in prices show the existence of "nonclustering noise," which includes other unspecified noise.

Noise creates the major distinction between the macro- and micromovements. Black (1986) noted that noise renders our observations of value imperfect by creating a wedge between price and intrinsic value, and it provides an impetus for trading in financial markets. Hasbrouck (1996) pointed out that information has a long-term, permanent impact whereas noise has a short-term, transitory impact on price. In the low-frequency macroprice data, the impact of noise is assumed to be small, hence the price process is assumed to be indistinguishable from its value. However, in the high-frequency data, the impact of noise cannot be neglected, and therefore the price process must be distinguished from its value process and noise must be modeled explicitly.

In this paper, a partially observed model for the micromovement is proposed. The aim is to bridge the gap between the macro- and micromovements caused by noise. The

[1] The impact of price discreteness on parameter estimation was studied by Gottlieb and Kalay (1985), Cho and Frees (1988), and Ball (1988).

[2] Price clustering was described first in the academic literature by Osborne (1962) and then by Niederhoffer (1965, 1966). There are a number of theories that explain price clustering. Niederhoffer (1965) argued that it is a consequence of limit and stop orders tending to be placed on specialists' books at even eighths. From an economic perspective, clustering is often characterized as arising when market participants agree (perhaps implicitly) to a price increment that is coarser than the technically mandated minimum. Harris (1991) suggested that such a convention arises from traders seeking to minimize negotiation costs, to avoid extended rounds of bargaining over amounts of diminishing importance. Christie and Schultz (1994) observed excessive clustering in NASDAQ and suspected collusion among NASDAQ dealers. Their papers led to market reforms in NASDAQ that reduced clustering and trading costs (see Barclay et al. 1999). Godek (1996) provided a theory of preference trading and Grossman et al. (1997) presented a competitive theory of clustering that emphasizes the effect of uncertainty, the size of transactions, volatility, and the informational and transactional roles of quotations on the degree of clustering.

[3] See the last pair of plots in Figure 5.1.

model maintains the strong connection between the macro- and micromovements and distinguishes intrinsic value from price. The value process is formulated by a martingale problem and is assumed to be partially observed through prices, which are discrete in state space. Prices are observed at the unevenly spaced trading times driven by a conditional Poisson process. The price process is constructed from the value process by including the three types of noise. The most prominent feature of the proposed model is that prices can be viewed as a collection of counting processes of price levels. Consequently, the model can be formulated as a filtering problem with counting process observations, and filtering equations can be derived. Under this representation, the whole sample paths of counting processes are observable and the complete information of prices and trading times can be utilized in parameter estimation.

Intraday data are discrete in value, irregularly spaced in time, and extremely large in size. Despite recent advances in econometrics, obtaining reliable parameter estimates for even simple micromovement models is extremely challenging. The Bayes estimation via filtering introduced in this paper represents a significant advance in real-time estimation. One key contribution is that a theorem on the convergence of conditional expectation of the model is proved. The theorem suggests that the Markov chain approximation method may be applied to construct a consistent recursive algorithm to compute the approximate posterior and then the Bayes estimates.

To demonstrate the usefulness of the proposed structure, a simplified version of the model and its recursive algorithm are constructed. Simulation shows that the simplified model conforms to the sample characteristics of the micromovement of stock prices and the computed Bayes estimates converge to their true values. Moreover, the recursive algorithm is sufficiently fast to provide real-time estimates. Finally, the tested procedure is applied to one month of transaction prices for Microsoft to obtain the Bayes estimates.

In Section 2 the model is developed in two equivalent ways, one by construction and the other by formulating it into a filtering problem with counting process observations. The former approach is intuitive, but it is the latter framework that provides the foundation for the statistical analysis utilizing complete information. In Section 3 we study the likelihoods and the posterior of the proposed model and derive the filtering equations. Bayes parameter estimation via filtering is introduced in Section 4. In Section 5 a simplified version of the model and its recursive algorithm are constructed in detail. The consistency of the Bayes parameter estimates for the simplified model is proved. Simulation and estimation results based on real data are also presented. Summary and concluding remarks are contained in Section 6.

2. THE MODEL

The model proposed is predicated on the simple intuition that the price is formed from an intrinsic value process by incorporating the noises that arise from the trading activity. Throughout, it is assumed that all stochastic processes are right continuous with left limits (cadlag).

Suppose that the value process X cannot be observed directly, but can be partially observed through the price process, Y. Suppose that X lives in a continuous state space and 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 micromovement process.

Prices can only be observed at irregularly spaced trading times, which are modeled by a conditional Poisson process. The rate of trading activity is described by an intensity function. Easley and O'Hara (1992) observed that trading intensity may depend on volatility. Our framework is more general in that we allow the intensity to depend on the parameters of the model, intrinsic value, and trading time; that is, the intensity function is $a(\theta(t), X(t), t)$, where θ is a vector of parameters in the model.

For the purpose of allowing time-dependent parameters such as stochastic volatility, and in preparation for parameter estimation, we augment the partially observed model (X, Y) to (θ, X, Y). Assume (θ, X, Y) is defined in a complete probability space $(\Omega, \mathcal{F}, P)$ with a filtration $\{\mathcal{F}_t\}_{0\le t\le\infty}$. Since the macromovement models are appropriate for the value processes, we invoke a mild assumption on (θ, X) so that all relevant stochastic processes are included.

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

When θ is a time-invariant vector of parameters, Assumption 2.1 can be simplified as follows.

ASSUMPTION 2.1$'$. *X is the solution of a martingale problem for a generator* $\mathbf{A}_\theta$ *such that*

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

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

The martingale problem and the generator approach (Ethier and Kurtz 1986) provide a powerful tool for the characterization of Markov processes. Their advantages are shown in Section 3 where the filtering equations are derived, and in Section 4 where the recursive algorithm is developed. Three examples illustrates the generality of Assumption 2.1.

EXAMPLE 2.1. The GBM model in SDE form is

$$\frac{dX(t)}{X(t)} = \mu\, dt + \sigma\, dW(t), \tag{2.1}$$

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

$$\mathbf{A}_1 f(x) = \mathbf{A}_{\theta,1} f(x) = \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 f}{\partial x^2}(x) + \mu x \frac{\partial f}{\partial x}(x),$$

where θ includes μ and σ.

EXAMPLE 2.2. Stochastic volatility model in SDE form is

$$dX(t) = c\sigma^2(t)\, dt + \sqrt{V(t)}\, dW(t)$$
$$dV(t) = (w - \beta V(t))\, dt + \alpha V(t)\, dB(t),$$

where $V(t) = \sigma^2(t)$, and $W(t)$ and $B(t)$ are independent standard Brownian motions. The generator for this model is

$$\mathbf{A}_\theta f(x, v) = \frac{1}{2} v \frac{\partial^2 f}{\partial x^2}(x, v) + cv \frac{\partial f}{\partial x}(x, v) + \frac{1}{2}\alpha^2 v^2 \frac{\partial^2 f}{\partial v^2}(x, v) + (w - \theta v)\frac{\partial f}{\partial v}(x, v),$$

where θ includes c, w, β, and α. This is the limiting diffusion model of GARCH(1,1) (Nelson 1990).

EXAMPLE 2.3. The jump-GBM model (Merton 1976) in SDE form is

$$\frac{dX(t)}{X(t)} = \mu\, dt + \sigma\, dW(t) + Y_{N(t)} dN(t),$$

where $N(t)$ is a Poisson process independent of $W(t)$ and it has intensity λ. The jump-magnitudes $\{Y_i\}$ are i.i.d. random variables with continuous density $q_Y(y)$. Its generator is

$$\mathbf{A}_\theta f(x) = \mathbf{A}_1 f(x) + \lambda \int [f(x(1+z)) - f(x)] q_Y(z)\, dz,$$

where $\mathbf{A}_1$ is defined in Example 2.1 and θ includes μ, σ and parameters of $q_Y(y)$.

As the above examples show, the additive property of generator is useful in producing more complicated models from simpler formulations. Furthermore, the affine jump-diffusion models, recently popularized by Duffie, Pan, and Singleton (2000), are also special cases of Assumption 2.1.

There are two equivalent methods to build the model proposed in this paper from the value process. The first constructs Y from X by incorporating noises. The second formulates (X, Y) as a filtering problem with counting process observations. The former approach is intuitive and is related to a *state space model*; the latter approach is useful for statistical analysis and is related to the *hidden Markov model* presented in Elliott, Lakhdar, and Moore (1995).

2.1. Construction of Y from X

There are three general steps in constructing Y from X. First, specify the value process $X(t)$. Next, determine trading times $t_1, t_2, \ldots, t_i, \ldots$, which may be driven by a conditional Poisson process with an intensity $a(X(t), \theta(t), t)$. Finally, $Y(t_i)$, the price at time t_i, is determined by

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

where $y = F(x)$ is a random transition function 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. Information affects $X(t)$, the value of an asset, and has a permanent influence on the price, whereas noise affects $F(x)$, the random transition function, and only has a transitory impact on price. This formulation is similar to the structural model proposed by Hasbrouck (1996) in that $X(t)$ is the permanent component and $F(x)$ is the transient component.

Our framework subsumes two related models. One is the rounding model where $Y(t_i) = R[X(t_i), \frac{1}{M}]$ with probability one, where $F(x)$ is deterministic. It was studied by Gottlieb and Kalay (1985), Cho and Frees (1988), and Ball (1988). The second is the ask-bid model developed by Harris (1990). Suppose that c is one half of the bid and ask spread. Then, $Y(t_i) = R[X(t_i) \pm c, \frac{1}{M}]$, each with probability one half.

In the rest of this section, we focus on building $F(x)$ to accommodate the three types of noise that are well-documented in financial literature: discrete noise, clustering noise, and nonclustering noise.[4]

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 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, $R[x, \frac{1}{M}]$. Without other noises, trades should occur at this tick, which is closest to the stock value.

Step 2. Incorporate nonclustering noise by adding V; $y' = R[x + V, \frac{1}{M}]$, where V is the nonclustering noise of trade i at time t_i. V can depend on x, but the sequences $\{V_i\}$ are assumed to be serially independent given the value sequence $\{x_i\}$.

Nonclustering noise is intended to explain three discreteness-related sample characteristics of transactional data that cannot be explained by the rounding model. First, the nonclustering 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 nonclustering noise, the fractional part of y' should still be approximately uniformly distributed. Here, we bias y' through a random biasing function $b(\cdot)$ to reflect price clustering. Similar to the nonclustering noise, the biasing function $b(\cdot)$ can also depend on x, but the sequence of $\{b_i(\cdot)\}$ is assumed to be independent given the value sequence $\{y_i'\}$. The random biasing function is similar to the clustering control variable proposed in Hasbrouck (1999).

In summary, the construction is

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

where the rounding function takes care of price discreteness, V_i takes care of nonclustering noise, and the random biasing function b_i takes care of clustering noise. The transition probability $p(y \mid x)$ of F can be computed given the specifications of V and $b(\cdot)$ by the following formula:

$$p(y \mid x) = \sum_{y'} p(y \mid y')p(y' \mid x), \tag{2.2}$$

where $p(y' \mid x)$ is the transition probability from the value x to y', and $p(y \mid y')$ is the transition probability from y' to the price y.

Alternatively, the model can also be framed as a filtering problem with counting process observations. This is important for statistical analysis because under this framework we are able to derive the filtering equations, which characterize the likelihoods and the posterior of the model.

[4] Although stocks have been traded using dollars and cents at the NYSE since 2000, these three types of noise are still significant.

2.2. Counting Process Observations

In Section 2.1, 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:

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

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. Hence, $(\theta, X, \vec{Y})$ is framed as a filtering problem with counting process observations.

We invoke three mild assumptions so that the model in Section 2.1 is equivalent to the counting process observations in equation (2.3). The equivalence ensures that the statistical analysis based on the latter specification can be applied to the former.

ASSUMPTION 2.2. N_k's *are unit Poisson processes under measure* P.

REMARK 2.1. The random time change implies that $Y_k(t) = N_k(\int_0^t \lambda_k(\theta(s-), X(s-), s-)\,ds)$ is a counting process with intensity $\int_0^t \lambda_k(\theta(s-), X(s-), s-)\,ds$, and $Y_k(t) - \int_0^t \lambda_k(\theta(s-), X(s-), s-)\,ds$ is a martingale (see Ethier and Kurtz 1986, Chap. 6).

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

REMARK 2.2. This assumption does not imply that $(\theta, X), Y_1, Y_2, \ldots, Y_n$ are independent under P, because Y depends on (θ, X) through the intensities; however, it does 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. The reference measure Q plays an important role in deriving the filtering equation in Section 3 and in proving the consistency of the recursive algorithm in Section 4.

REMARK 2.3. 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, p. 165).

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.*

REMARK 2.4. This assumption imposes a desirable structure for the intensities of the model. It means that the total trading intensity $a(\theta(t), X(t), t)$ determines the overall rate of trade occurrence at time t and $p(y_k \mid x)$ determines the proportional intensity of trade at the price level, y_k, when the value is x.

The point process literature refers to Y_k (the record of trades occurring at the kth price level) as a *thinned point process* of the total trading process $\vec{Y}$. A heuristic argument to obtain the intensity of $Y_k(t)$ is the following:

$$P\left\{ Y_k(t+h) - Y_k(t) > 0 \,\middle|\, \mathcal{F}_t^{\theta, X} \right\} = a(\theta(t), X(t), t) p(y_{k \mid X(t)}).$$

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 Sections 2.1 and 2.2. It also allows us to generalize the finite price levels to the countably infinite price levels to accommodate the GBM case.

A final assumption ensures the uniqueness of the filtering equation and the consistency of the Bayesian estimates for the simplified model discussed below.

ASSUMPTION 2.5. *The total intensity, $a(\theta, x, t)$, is uniformly bounded from below and above; that is, there exist constants C_1 and C_2 such that* $0 < C_{1a}(\theta, x, t) \leq C_2$ *for all* $t > 0$.

3. FOUNDATIONS FOR STATISTICAL INFERENCE

In this section, we study the likelihood, the marginal likelihood, and the posterior of the proposed model, and establish their relationships with the unnormalized and normalized filtering equations.

3.1. The Likelihoods

The observable data are a sequence $\{(t_i, Y(t_i))\}$, representing trading times and prices. Standard econometric approaches (for a brief survey and new development, see Ait-Sahalia 2002) would consider the likelihood of the form

$$f((y_{t_1}, t_1), (y_{t_2}, t_2), \ldots, (y_{t_n}, t_n); \theta),$$

which, in terms of measure theory, is the Radon-Nikodym derivative with respect to the dominating measure of an *n-product* Lebesgue measure at times $t_1, t_2, \ldots, t_n$. The likelihood implies that trades *occur* at the time $t_1, t_2, \ldots, t_n$. This likelihood, in terms of stochastic process, is a finite-dimensional distribution of the process. It contains only the information of a discrete-time subset of a sample path where the trades occur. It is called a discrete-time likelihood because it is the likelihood given a discrete-time subset of the sample path.

Correspondingly, there is a continuous-time likelihood, which is the likelihood given the whole sample path. Just as the whole sample path contains more information than any discrete-time subset, the continuous-time likelihood contains more information about the process than any discrete-time counterpart. The continuous-time likelihood is the Radon-Nikodym derivative with respect to the dominating measure such as the Wiener measure, the Poisson measure, or their product. In most settings, the whole sample path may not be observable. Even when it is observable, the continuous-time likelihood may not exist or may not be computable. However, when the whole sample path is observable and the continuous-time likelihood is available, it provides a better informational base than any discrete-time likelihood for the inference about the stochastic processes.

One key contribution of this paper is that when the proposed model is framed as a filtering problem with counting process observations, the whole sample paths of the counting processes are observable. The observed counting processes imply that trades *occur and only occur* at the time $t_1, t_2, \ldots, t_n$ and the corresponding continuous-time likelihood has the same implication. Under this representation, the complete information—that is, trades occur and only occur at the time $t_1, t_2, \ldots, t_n$—is used in the continuous-time likelihood for statistical inference.

Before we present the continuous-time likelihood of the model, we specify the probability space $(\Omega, \mathcal{F}, P)$ for $(\theta, X, \vec{Y})$, and concisely define some terminology. Let $D_{R^n}[0,\infty)$ be the space of sample paths of the right continuous functions with left limits in the state space R^n. $D_{R^n}[0,\infty)$ is embedded with the Skorohod topology. Let $\mathcal{B}(D_{R^n}[0,\infty))$ be the Borel σ-algebra of $D_{R^n}[0,\infty))$. Then, $\Omega = D_{R^p}[0,\infty) \times D_R[0,\infty) \times D_{R^n}[0,\infty)$, $\mathcal{F} = \mathcal{B}(D_{R^p}[0,\infty)) \times \mathcal{B}(D_R[0,\infty)) \times \mathcal{B}(D_{R^n}[0,\infty))$, and $P = P_{\theta,x} \times P_{y\,|\,\theta,x}$, where $P_{\theta,x}$ is the probability measure on $D_{R^{p+1}}[0,\infty)$ for (θ, X) such that $M_f(t)$ in Assumption 2.1 is a $\mathcal{F}_t^{\theta,X}$-martingale, and $P_{y\,|\,\theta,x}$ is the conditional probability measure on $D_{R^n}[0,\infty)$ for $\vec{Y}$ given (θ, X). Under this structure, $\vec{Y}$ depends on (θ, X).

Assumptions 2.2 and 2.3 imply that there exists a reference measure Q such that P is absolutely continuous with respect to Q and $Q = P_{\theta,x} \times Q_{\vec{y}}$. Under the reference measure Q, the counting processes are n independent unit Poisson processes, and $\vec{Y}$ is independent of (θ, X).

The Radon-Nikodym derivative (see Bremaud 1981, pp. 165–167) of the model is

$$
\begin{aligned}
L(t) &= \frac{dP}{dQ}(t) = \frac{dP_{\theta,x}}{dP_{\theta,x}}(t) \times \frac{dP_{y\,|\,\theta,x}}{dQ_{\vec{y}}}(t) = \frac{dP_{y\,|\,\theta,x}}{dQ_{\vec{y}}}(t) \\
&= \prod_{k=1}^{n} \exp\left\{ \int_0^t \log \lambda_k(\theta(s-), X(s-), s-)\, dY_k(s) - \int_0^t [\lambda_k(\theta(s), X(s), s) - 1]\, ds \right\}.
\end{aligned}
$$

In SDE form,

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

where $L(t)$ is the joint likelihood of $(\theta, X, \vec{Y})$. However, X and the stochastic components of θ (such as stochastic volatility) cannot be observed, and then $L(t)$ is not computable. What is needed for the statistical analysis is the likelihood of $\vec{Y}$ alone.

REMARK 3.1. The following analogy suggests a way to obtain the likelihood of Y from $L(t)$. Suppose $f(\theta, x, y)$ is the joint density for real random variables (θ, X, Y) with respect to a Lebesgue measure Q'. Then the marginal density of Y can be obtained by integrating on (θ, X), or in terms of conditional expectation as:

$$
f_Y(y) = \int\!\!\int f(\theta, x, y)\, d\theta\, dx = E^{Q'}[f(\theta, X, Y) \mid Y = y].
$$

To present the likelihood of Y concisely, two definitions are required.

DEFINITION 3.1. Let $\mathcal{F}_t^{\vec{Y}} = \sigma\{(\vec{Y}(s)) | 0 \le s \le t\}$ 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 3.2. Let $\phi(f, t) = E^Q[f(\theta(t), X(t))L(t) \mid \mathcal{F}_t^{\vec{Y}}]$ be the conditional expectation of fL given $\mathcal{F}_t^{\vec{Y}}$.

From a frequentist's viewpoint, assuming that $\theta(0)$ is fixed, the likelihood of Y is $E^Q[L(t) \mid \mathcal{F}^{\vec{Y}}] = \phi(1, t)$. From a Bayesian's viewpoint, assuming a prior on $\theta(0)$, the marginal (or integrated) likelihood of Y is also $E^Q[L(t) \mid \mathcal{F}^{\vec{Y}}] = \phi(1, t)$.

3.1.1. Maximum Likelihood versus Bayesian Inference. Maximum likelihood (ML) and Bayesian methods are two major approaches in statistical inference. Under regular conditions on the likelihood functions, ML estimates are consistent, asymptotically normal, and asymptotically efficient. Moreover, there is a generalized likelihood ratio (GLR) test for hypothesis testing. Similarly, under regular conditions, Bayes estimates have the same asymptotic properties as MLE and both estimates are asymptotically equivalent (Ghosal, Ghosh, and Samanta 1995). Also, Bayesian hypothesis testing and model selection procedure are available via Bayes factor (Kass and Raftery 1995).

However, the likelihood of the proposed model is irregular. This implies that the desirable properties of ML estimates of the proposed model are unknown and the GLR test is not applicable. Hence, the ML approach is less preferable, although the likelihood function can be computed using the technique developed in Section 4.

In this paper, we choose the Bayesian approach for two additional reasons. First, Bayes estimates are the least mean square error estimates, and the approximate posterior computed by the recursive algorithm developed in Section 4 provides more information than ML estimates. Second, even though the GLR test fails for the model, the Bayesian hypothesis testing procedure via Bayes factor still works and the Bayes factor, which is the ratio of the marginal likelihoods, is computable.

3.2. The Posterior for Bayes Estimation

Two additional terms are defined before the posterior of (θ, X) is presented.

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

DEFINITION 3.4. 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 $f(\theta(t), X(t))$ given $\mathcal{F}_t^{\vec{Y}}$.

The Bayes Formula (see Bremaud 1981, p. 171) provides the relationship between $\phi(f, t)$ and $\pi(f, t)$:

$$E^P\left[f(\theta(t), X(t)) \,\middle|\, \mathcal{F}_t^{\vec{Y}}\right] = \frac{E^Q\left[f(\theta(t), X(t))L(t) \,\middle|\, \mathcal{F}_t^{\vec{Y}}\right]}{E^Q\left[L(t) \,\middle|\, \mathcal{F}_t^{\vec{Y}}\right]},$$

or, equivalently,

$$\pi(f, t) = \frac{\phi(f, t)}{\phi(1, t)}.$$

That is, $\pi(f, t)$ is obtained by normalizing $\phi(f, t)$ with $\phi(1, t)$. Hence, the equation governing the evolution of $\phi(f, t)$ is called the *unnormalized filtering equation*, and that of

$\pi(f, t)$ is called the *normalized equation*. The Bayes Formula implies that $\phi(f, t)$ can determine $\pi(f, t)$, but the converse is not usually true.

Again, from a Bayesian's viewpoint, assuming priors on $\theta(0)$, π_t becomes the joint posterior of $(\theta(t), X(t))$. The normalized filtering equation then determines how π_t evolves. Together with the Markov chain approximation method in Section 4, the normalized filtering equation provides an effective way to compute the approximate joint posterior and then the Bayes estimates for the parameters.

3.3. Filtering Equations

The filtering equations provide an effective means to characterize $\phi(f, t)$ and $\pi(f, t)$. Note that $\phi(f, t)$, which determines the likelihood or marginal likelihood, is characterized by the unnormalized filtering equation. And $\pi(f, t)$, which determines the posterior, is characterized by the normalized filtering equation, and is the optimum filter in the sense of least mean square error (Kallianpur 1980). We summarize the filtering equations in the following theorem.

THEOREM 3.1. *Suppose that* (θ, X) *satisfies Assumption 2.1 and that* $\vec{Y}$ *is the counting process observations defined in equation (2.3) with Assumptions 2.2–2.5. Then, for every* $t > 0$ *and every* f *in the domain of generator* $\mathbf{A}$, $\phi(f, t)$ *is the unique solution of the SDE, the unnormalized filtering equation,*

(3.1)

$$\phi(f, t) = \phi(f, 0) + \int_0^t \phi(\mathbf{A}f - (a - n)f, s)\, ds + \sum_{k=1}^{n} \int_0^t \phi((ap_k - 1)f, s-)\, dY_k(s),$$

where $a = a(\theta(t), X(t), t)$, *is the trading intensity, and* $p_k = p(y_k \mid x)$ *is the transition probability from* x *to* y_k.

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

$$\begin{aligned}\pi(f, t) = \pi(f, 0) &+ \int_0^t [\pi(\mathbf{A}f, s) - \pi(fa, s) + \pi(f, s)\pi(a, s)]\, ds \\ &+ \sum_{k=1}^{n} \int_0^t \left[\frac{\pi(fap_k, s-)}{\pi(ap_k, s-)} - \pi(f, s-)\right] dY_k(s).\end{aligned} \tag{3.2}$$

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

(3.3)

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

Proof: The proof of Theorem 3.1 is given in Appendix A.

Note that $a(t)$ disappears in equation (3.3). This reduces the computation greatly in the Bayesian parameter estimation, hence this convenient case is studied in detail. The trade-off 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 equation (3.3) can be written in two parts: the *propagation equation*, which describes the evolution without trades, and the *updating equation*, which describes 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. Moreover, when $\mathbf{A}$ is a diffusion generator, the propagation equation is a *parabolic*-type partial differential equation.

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. The preceding two equations provide the foundation for Bayes estimation via filtering.

4. BAYES ESTIMATION VIA FILTERING

Through the normalized filtering equation, we are able to compute the continuous-time posterior and the Bayes estimate, which is the posterior mean. The core of Bayesian estimation via filtering is to construct an algorithm to compute the approximate conditional distribution, which becomes the approximate posterior after a prior is assigned. The algorithm based on a filtering equation is naturally recursive with every trade. The recursiveness is a desirable feature because a recursive algorithm handles a datum at a time, so it requires less memory, makes real-time updating implementable, and can handle large data sets.

One basic requirement for the recursive algorithm is consistency: The conditional distribution computed by the recursive algorithm must converge to the true one. The following theorem proves the convergence of the conditional expectation. The theorem is important in that it provides the theoretical foundation for consistency. Careful examination of the theorem provides a recipe for constructing a consistent recursive algorithm.

4.1. 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 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

$$\vec{Y}_\varepsilon(t) = \begin{pmatrix} N_1\big(\int_0^t \lambda_1(\theta_\epsilon(s-), X_{\epsilon_x}(s-), s-)\, ds\big) \\ N_2\big(\int_0^t \lambda_2(\theta_\epsilon(s-), X_{\epsilon_x}(s-), s-)\, ds\big) \\ \vdots \\ N_n\big(\int_0^t \lambda_n(\theta_\epsilon(s-), X_{\epsilon_x}(s-), s-)\, ds\big) \end{pmatrix}, \tag{4.1}$$

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

THEOREM 4.1. *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 (4.1). 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* $\varepsilon \to 0$; *and*
(ii) $E^{P_\varepsilon}[F(\theta_\epsilon(t), X_{\epsilon_x}(t)) \mid \mathcal{F}_t^{\vec{Y}_\varepsilon}] \Rightarrow E^P[F(\theta(t), X(t)) \mid \mathcal{F}_t^{\vec{Y}}]$ as $\varepsilon \to 0$ *for function F in the domain of the generator* $\mathbf{A}$.

Proof: The proof relies on three related theorems.

1. *Continuous Mapping Theorem.* See Corollary 1.9 in Chapter 3 of Ethier and Kurtz (1986).
2. *Kurtz and Protter's Theorem on the Convergence of Stochastic Integrals.* See Theorem 2.2 of Kurtz and Protter (1991) and their Example 3.3.
3. *Goggin's Theorem on Convergence of Conditional Expectation*. See Theorem 2.1 of Goggin (1994).

Since $(\theta_\epsilon, X_{\epsilon_x}) \Rightarrow (\theta, X)$, the Continuous Mapping Theorem implies that $\vec{Y}_\varepsilon \Rightarrow \vec{Y}$. Then we have $(\theta_\epsilon, X_{\epsilon_x}, \vec{Y}_\varepsilon) \Rightarrow (\theta, X, \vec{Y})$. Assumptions 2.2 and 2.3 for $(\theta_\epsilon, X_{\epsilon_x}, \vec{Y}_\varepsilon)$ imply that there exists a reference measure Q_ε such that, under Q_ε, $Y_{\varepsilon,k}$, the kth component of $\vec{Y}_\varepsilon$, for $k = 1, 2, \ldots, n$ are independent unit Poisson processes, and they are independent of $(\theta_\epsilon, X_{\epsilon_x})$. The Radon-Nikodym derivative $dP_\varepsilon/dQ_\varepsilon$ is

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

Kurtz and Protter's Theorem implies that

$$\int_0^t \log \lambda_k(\theta_\epsilon(s-), X_{\epsilon_x}(s-), s-)\, dY_{\varepsilon,k}(s) \Rightarrow \int_0^t \log \lambda_k(\theta(s-), X(s-), s-)\, dY_k(s)$$

and

$$\int_0^t [\lambda_k(\theta_\epsilon(s), X_{\epsilon_x}(s), s) - 1]\, ds \Rightarrow \int_0^t [\lambda_k(\theta(s), X(s), s) - 1]\, ds.$$

The Continuous Mapping Theorem again implies $L_\varepsilon(t) \Rightarrow L(t)$. Therefore, we have the triplet $((\theta_\epsilon, X_{\epsilon_x}), \vec{Y}_\varepsilon, L_\varepsilon) \Rightarrow ((\theta, X), \vec{Y}, L)$. Goggin's Theorem implies that Theorem 4.1 follows. □

REMARK 4.1. This theorem guarantees that as long as $(\theta_\epsilon, X_{\epsilon_x})$ is an approximation of (θ, X), Y_ε is an approximation of $\vec{Y}$ and the conditional expectation of $(\theta_\epsilon, X_{\epsilon_x})$ is an approximation of (θ, X) given the observed sample paths of the counting processes. When we take "F" as an appropriate indicator function, $E^{P_\varepsilon}[F(\theta_\epsilon(t), X_{\epsilon_x}(t)) \mid \mathcal{F}_t^{\vec{Y}_\varepsilon}]$ becomes the conditional probability mass function for $(\theta_\epsilon, X_{\epsilon_x})$. Theorem 4.1, then, implies that the conditional probability mass function is an approximation of the conditional distribution of (θ, X).

Next, a recursive algorithm is constructed to compute such a conditional probability mass function.

4.2. Construction of a Consistent Recursive Algorithm

For the nonlinear filtering problem, the equally spaced in time Markov chain approximation method is well suited for the computation of the approximation to an optimum filter (see Chapter 12 of Kushner and Dupuis 1994). Although trades occur irregularly spaced in time, we still apply the same idea of Markov chain approximation here.

Theorem 4.1 provides a recipe for constructing a consistent recursive algorithm. The Markov chain approximation approach is used to construct $(\theta_\epsilon, X_{\epsilon_x})$ and then the recursive algorithm. For simplicity, the recursive algorithm in this section is constructed for equation (3.3), which is the case when the trading intensity is deterministic. It is straightforward to generalize the recursive algorithm to the filtering equation (3.2), where trading intensity depends on other variables.

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

Step 1. Prepare the appropriate state processes in the appropriate probability space with the appropriate filtering equation. First, separate the parameters into two types according to whether they are time (in-) dependent. That is, set the p-dimensional parameter vector, $\theta(t) = (\xi, \eta(t))$, where p_1-dimension ξ does not depends on time but p_2-dimension $\eta(t)$ does.

Next, set the state processes to be $(\xi, \eta(t), X(t))$ and assume $(\xi, \eta, X, \vec{Y})$ lives in a complete probability space $(\Omega, \mathcal{F}, P)$ with the Skorohod topology. Define $\Omega = \Xi \times D_{R^{p_2+1}}[0, \infty) \times D_{R^n}[0, \infty)$, where Ξ is a compact subset in $R^{p_1+p_2+1}$ as the parameter space for $\xi \times \eta(0) \times X(0)$. Define $\mathcal{F} = \mathcal{B}(\Xi) \times \mathcal{B}(D_{R^{p_2+1}}[0, \infty)) \times \mathcal{B}(D_{R^n}[0, \infty))$ where $\mathcal{B}(\Xi)$ is a Borel σ-algebra on Ξ. Define $P = P_{\xi,\eta(0),X(0)} \times P_{\eta,X|\xi,\eta(0),X(0)} \times P_{\vec{y}|\xi,\eta,X}$, where $P_{\xi,\eta(0),X(0)}$ is a prior on Ξ, $P_{\eta,X|\xi,\eta(0),X(0)}$ is the conditional probability measure such that under the joint probability measure $P_{\theta,X} = P_{\xi,\eta,X} = P_{\xi,\eta(0),X(0)} \times P_{\eta,X|\xi,\eta(0),X(0)}$, $M_f(t)$ in Assumption 2.1 is a $\mathcal{F}_t^{X,\theta}$-martingale, and $P_{\vec{y}|\xi,\eta,X}$ is the conditional probability measure of the counting vector process $\vec{Y}$ given ξ, η, X.

For the new state processes and the new probability space, we redefine the corresponding new conditional distribution and its expectation.

DEFINITION 4.1. π_t is the conditional distribution of $(\xi, \eta(t), X(t))$ given $\mathcal{F}_t^{\vec{Y}}$.

DEFINITION 4.2. $\pi(f, t) = E^P[f(\xi, \eta(t), X(t)) \mid \mathcal{F}_t^{\vec{Y}}] = \int f(\xi, \eta, x)\pi_t(d\xi, d\eta, dx)$.

By Theorem 3.1, we obtain the new filtering equation for $\pi(f, t)$ defined above in the same form as equation (3.3) when trading intensity is $a(t)$. This is the conceptual simplicity we gain when we treat the unknown parameters of interest as the unobserved state process, which is an important idea in Bayesian analysis for latent variables.

Step 2. Construct the Markov chain approximation to (ξ, η, X). First, discretize the state space of (ξ, η, X) with mesh sizes $(\epsilon_\xi, \epsilon_\eta, \epsilon_x)$ and set $\varepsilon = \max(|\epsilon_\xi|, |\epsilon_\eta|, \epsilon_x)$. The discretized state space of Ξ is a natural approximate for Ξ. But for the stochastic process

$(\eta(t), X(t))$, its natural approximation is a Markov chain. Second, observe that the construction of an approximate Markov chain can be transformed to construct a Markov chain generator, $\mathbf{A}_\varepsilon$, such that $\mathbf{A}_\varepsilon \to \mathbf{A}$ as $\varepsilon \to 0$. The generators for asset price models can be classified as diffusion and jump generators. A diffusion generator involves first- and second-order differentiation. The finite difference type approximation can be applied to the diffusion generator in constructing a birth and death generator, which is a simple Markov chain generator. (This is illustrated in Section 5.2 where we deal with a simplified model.) The jump generator involves integration, and the usual rectangle approximation for integral can be employed to construct a Markov chain generator. Following these two rules, we are able to construct the approximate Markov chain generators for examples in Section 2. Two examples are given below. For clarification, assume that x and ξ are in the discretized spaces in the following examples.

EXAMPLE 4.1. Set $\xi = (\mu, \sigma, \rho)$, where ρ is a vector of parameters for different types of noise. The approximate birth and death generator for Example 2.1 is

$$\mathbf{A}_{\varepsilon,1} f(\xi, x) = a(\xi, x)(f(\xi, x + \epsilon_x) - f(\xi, x)) + b(\xi, x)(f(\xi, x - \epsilon_x) - f(\xi, x)),$$

where

$$a(\xi, x) = \frac{1}{2}\left(\frac{\sigma^2 x^2}{\epsilon_x^2} + \frac{\mu x}{\epsilon_x}\right), \qquad b(\xi, x) = \frac{1}{2}\left(\frac{\sigma^2 x^2}{\epsilon_x^2} - \frac{\mu x}{\epsilon_x}\right).$$

EXAMPLE 4.2. Set $\xi = (\mu, \sigma, \lambda, \rho)$, where ρ is a vector of parameters for different types of noise and jump magnitudes. The approximate Markov chain generator for Example 2.3 is

$$\mathbf{A}_\varepsilon f(\xi, x) = \mathbf{A}_{\varepsilon,1} f(\xi, x) + \lambda \sum_z [f(\xi, z) - f(\xi, x)]\hat{q}_Y(z),$$

where $\mathbf{A}_{\varepsilon,1}$ is defined in Example 4.1, the summation is over all lattices of x, and $\hat{q}_Y(z) = P\{Y \in [(z - 0.5\epsilon_x)/x - 1, (z + 0.5\epsilon_x)/x - 1)\}$.

The approximate Markov chain generator for Example 2.2 can be obtained similarly as shown in Section 5.2. All $\mathbf{A}_\varepsilon$ converges to $\mathbf{A}$ as $\varepsilon \to 0$.

Based on $\mathbf{A}_\varepsilon$, the Markov chain, $(\theta_\epsilon(t), X_{\epsilon_x}(t))$ can be constructed as the approximate model of $(\theta(t), X(t))$, where $\theta_\epsilon(t) = (\xi_{\epsilon_\xi}, \eta_{\epsilon_\eta}(t))$ and $\theta(t) = (\xi, \eta(t))$. Then, $\vec{Y}_\varepsilon$, defined by equation (4.1), is obtained.

REMARK 4.2. The counting process observations can be chosen to be $\vec{Y}(t)$ defined by equation (2.3) or $\vec{Y}_\varepsilon(t)$ depending on whether the driving process is $(\theta(t), X(t))$ or $(\theta_\epsilon(t), X_{\epsilon_x}(t))$. When we model the parameters and the stock value as $(\theta(t), X(t))$, the counting process observations of stock price are regarded as $\vec{Y}(t)$. When we intend to compute the posteriors of the parameters and the value process, we use $(\theta_\epsilon(t), X_{\epsilon_x}(t))$ to approach $(\theta(t), X(t))$ and the counting process observations of price are regarded as $\vec{Y}_\varepsilon(t)$.

REMARK 4.3. By Theorem 4.1, 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 3. 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 it is simple to check that Assumptions 2.1–2.5 also hold for $(\theta_\epsilon, X_{\epsilon_x}, \vec{Y}_\varepsilon)$. To present the filtering equation for the approximate model, the discretized approximations of π_t and $\pi(f, t)$ in Definitions 4.1 and 4.2 are defined as follows.

DEFINITION 4.3. Let $\pi_{\varepsilon,t}$ be the conditional probability mass function of $(\xi_{\epsilon_\xi}, \eta_{\epsilon_\eta}(t), X_{\epsilon_x}(t))$ given $\mathcal{F}_t^{\vec{Y}_\varepsilon}$.

DEFINITION 4.4.

$$\pi_\varepsilon(f, t) = E^{P_\varepsilon}\big[f(\xi_{\epsilon_\xi}, \eta_{\epsilon_\eta}(t), X_{\epsilon_x}(t)) \,\big|\, \mathcal{F}_t^{\vec{Y}_\varepsilon}\big] = \sum_{\xi,\eta,x} f(\xi, \eta, x)\pi_{\varepsilon,t}(\xi, \eta, x),$$

where (ξ, η, x) covers all lattices in the discretized state space.

Applying Theorem 3.1, we obtain the filtering equation of the approximate model in a similar form:

(4.2)

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

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

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

$$\pi_\varepsilon(f, t_{i+1}-) = \pi_\varepsilon(f, t_i) + \int_{t_i}^{t_{i+1}-} \pi_\varepsilon(\mathbf{A}_\varepsilon f, s)\, ds, \tag{4.3}$$

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

$$\pi_\varepsilon(f, t_{i+1}) = \frac{\pi_\varepsilon(fp_k, t_{i+1}-)}{\pi_\varepsilon(p_k, t_{i+1}-)}. \tag{4.4}$$

Step 4. Convert equations (4.3) and (4.4) to the recursive algorithm. First, assume all the vectors in the discretized state space of ξ are indexed and represented as $\{\xi_j, j \in \mathcal{J}\}$, those of η at time t as $\{\eta_m, m \in \mathcal{M}\}$ and those of X at time t as $\{x_l, l \in \mathcal{L}\}$. Now, we define the approximate posterior that the recursive algorithm computes.

DEFINITION 4.5. The posterior of the approximate model, $(\xi_{\epsilon_\xi}, \eta_{\epsilon_\eta}, X_{\epsilon_x}, Y_\varepsilon)$, at time t is denoted by

$$p_\varepsilon(\xi_j, \eta_m, x_l; t) = P\big\{\xi_{\epsilon_\xi} = \xi_j, \eta_{\epsilon_\eta}(t) = \eta_m, X_{\epsilon_x}(t) = x_l \,\big|\, \mathcal{F}_t^{\vec{Y}_\varepsilon}\big\}.$$

The core of the conversion is to take f as the following indicator function, in which ξ_ϵ denotes ξ_{ϵ_ξ}, $\eta_\epsilon(t)$ denotes $\eta_{\epsilon_\eta}(t)$, and $X_\epsilon(t)$ denotes $X_{\epsilon_x}(t)$.

$$\mathbf{I}_{\{\xi_\epsilon=\xi_j, \eta_\epsilon=\eta_m, X_\epsilon=x_l\}}(\xi_\epsilon, \eta_\epsilon, X_\epsilon) \stackrel{\text{def}}{=} \mathbf{I}(\xi_j, \eta_m, x_l). \tag{4.5}$$

Then the following facts emerge:

$$\pi_\varepsilon(\mathbf{I}(\xi_j, \eta_m, x_l), t) = p_\varepsilon(\xi_j, \eta_m, x_l; t),$$

$$\pi_\varepsilon(\mathbf{I}(\xi_j, \eta_m, x_l)p_k, t) = p_\varepsilon(\xi_j, \eta_m, x_l; t)p(y_k \mid x_l; \xi_j, \eta_m),$$

and

$$\pi_\varepsilon(p_k, t) = \sum_{j',m',l'} p_\varepsilon(\xi_{j'}, \eta_{m'}, x_{l'}; t) p(y_k \mid x_{l'}; \xi_{j'}, \eta_{m'}),$$

where the transition probability $p_k = p(y_k \mid x) = p(y_k \mid x; \xi, \eta)$.
Hence, the propagation equation (4.3) becomes

$$p_\varepsilon(\xi_j, \eta_m, x_l; t_{i+1}-) = p_\varepsilon(\xi_j, \eta_m, x_l; t_i) + \int_{t_i}^{t_{i+1}-} \pi_\varepsilon(\mathbf{A}_\varepsilon \mathbf{I}, s)\, ds, \tag{4.6}$$

where $\pi_\varepsilon(\mathbf{A}_\varepsilon\mathbf{I}, s)$ depends on $\mathbf{A}_\varepsilon$. Note that $\mathbf{A}_\varepsilon\mathbf{I}$ is a linear combination of some indicator functions. Therefore, $\pi_\varepsilon(\mathbf{A}_\varepsilon\mathbf{I}, s)$ is a linear combination of $p_\varepsilon(\cdot, \cdot, \cdot; s)$.

The updating equation (4.4) can be written as

$$p_\varepsilon(\xi_j, \eta_m, x_l; t_{i+1}) = \frac{p_\varepsilon(\xi_j, \eta_m, x_l; t_{i+1}-) p(y_k \mid x_l, \xi_j, \eta_m)}{\sum_{j',m',l'} p_\varepsilon(\xi_{j'}, \eta_{m'}, x_{l'}; t) p(y_k \mid x_{l'}; \xi_{j'}, \eta_{m'})} \tag{4.7}$$

if a trade at the kth price level occurs at time t_{i+1}.

Next, we convert equation (4.6) to a recursive algorithm. Equation (4.6) is deterministic and therefore one can employ the Euler scheme. After excluding the probability-zero event that two or more jumps occur at the same time, there are two possible cases for the trade waiting time. Case 1, if $t_{i+1} - t_i \le LL$, a length controller, then we can approximate $p_\varepsilon(\xi_j, \eta_m, x_l; t_{i+1}-)$ by the following recursive equation

$$p_\varepsilon(\xi_j, \eta_m, x_l; t_{i+1}-) \approx p_\varepsilon(\xi_j, \eta_m, x_l; t_i) + \pi_\varepsilon(\mathbf{A}_\varepsilon\mathbf{I}, t_i)(t_{i+1} - t_i). \tag{4.8}$$

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 approximate $p_\varepsilon(\xi_j, \eta_m, x_l; t_{i+1}-)$ by applying repeatedly the recursive algorithm given by equation (4.8) from $t_{i,0}$ to $t_{i,1}$, then $t_{i,2}, \ldots,$ until $t_{i,n} = t_{i+1}$. Note that $p_\varepsilon(\xi_j, \eta_m, x_l; t_{i,j}-) = p_\varepsilon(\xi_j, \eta_m, x_l; t_{i,j})$ for all j except $j = n$, because $p_\varepsilon(\xi_j, \eta_m, x_l; t)$ is continuous in t in $[t_i, t_{i+1}-)$.

Equations (4.7) and (4.8) constitute the recursive algorithm we employ to calculate the posterior.

REMARK 4.4. $p_\varepsilon(\xi_j, \eta_m, x_l; t_{i+1})$ only depends on $p_\varepsilon(\cdot, \cdot, \cdot; t_{i+1}-)$, which in turn depends on $p_\varepsilon(\cdot, \cdot, \cdot; t_i)$, all previously available information. Therefore, once the prior distribution at time $t = 0$ is specified, the recursive algorithm can calculate the posteriors based on available data.

The last step is to assign priors to ξ, $\eta(0)$, and $X(0)$. We assume independence among ξ, $\eta(0)$, and $X(0)$. If there is no prior information about $\theta(0)$ available, we simply assign uniform distributions to the discretized state space $(\xi, \eta(0))$ in a reasonable range and assume all components are independent. For the prior of $X(0)$, we have a better option. Since the best estimate of $X(0)$ is the first trading price, denoted by $Y(0)$, we let $X(0) = Y(0)$ with probability 1.

4.3. Consistency of the Recursive Algorithm

Reviewing the construction of the recursive algorithm, we observe that there are two approximations employed. The first is to approach the integral in the propagation equation (4.6). The Euler scheme is applied to approximate the integration equation and the convergence of the Euler scheme is well-known. The second is more important. It approximates the filtering equation (3.3) by the filtering equation (4.2) of the approximate model. The weak convergence of equation (4.2) to equation (3.3) in the Skorohod topology is guaranteed by Theorem 4.1. By taking f as the indicator (defined in equation (4.5)) in equation (4.2), one obtains the recursive algorithm to compute $p_\varepsilon(\xi_j, \eta_m, x_l; t)$. According to Theorem 4.1, $p_\varepsilon(\xi_j, \eta_m, x_l; t)$ converges to an expression related to the true posterior (see below). In order to identify this expression, we first define three neighborhoods.

Suppose ξ_j is a vector in the discretized parameter space for ξ, which is represented by $\{\xi_j, j \in \mathcal{J}\}$. $\xi_{d,j}$ (or ξ_d) is the dth component of the vector ξ_j (or ξ) and $\epsilon_{\xi,d}$ is the mesh size for the dth component of the discretized space for ξ. Define the neighborhood of ξ_j as

$$N(\xi_j) = \left\{(\xi_1, \xi_2, \ldots, \xi_{p_1}) : \xi_{d,j} - \tfrac{1}{2}\epsilon_{\xi,d} \le \xi_d < \xi_{d,j} + \tfrac{1}{2}\epsilon_{\xi,d}, d = 1, 2, \ldots, p_1\right\}.$$

Similarly, define the neighborhoods for η_m and x_l as $N(\eta_m)$ and $N(x_l)$. Now, the true posterior that the recursive algorithm aims to approximate is defined as follows.

DEFINITION 4.6.

$$p(\xi_j, \eta_m, x_l; t) = P\left\{\xi \in N_{\xi_j}, \eta(t) \in N_{\eta_m}, X(t) \in N_{x_l} \,\middle|\, \mathcal{F}_t^{\vec{Y}}\right\}.$$

When one takes f to be the following indicator function,

$$\mathbf{I}_{\{\xi \in N_{\xi_j}, \eta \in N_{\eta_m}, X \in N_{x_l}\}}(\xi, \eta, x),$$

in equation (3.3), Theorem 4.1 implies $p_\varepsilon(\xi_j, \eta_m, x_l; t) \to p(\xi_j, \eta_m, x_l; t)$, for every $t > 0$ as $\varepsilon \to 0$. This is because the indicator function just defined becomes the one in equation (4.5) in the approximate model.

5. A SIMPLIFIED MODEL WITH BAYES ESTIMATION

We construct a simplified version of the foregoing general model, and use outline the procedure for constructing a recursive algorithm. Simulation examples demonstrate that this simplified model conforms to the real data and the recursive algorithm works efficiently.

5.1. A Model with GBM as Value Process

In the simplified model, let GBM (equation (2.1)) be the intrinsic value process, where information is generated by $\sigma\, dW(t)$. The total trading intensity is deterministic and, as shown in Remark 3.2, it drops out of the normalized filtering equation. A deterministic intensity $a(t)$ fits the duration data better than the naive assumption that the trading intensity is time-invariant and it can explain the observation that trading activity is higher in the opening and the closing periods.

The nonclustering noise, $\{V_i\}$ is assumed to be a sequence of i.i.d. random variables and independent of the value process, X. The distribution of V should be unimodal, symmetric, and bell-shaped in order to conform to the desirable features 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 equal probabilities. A good candidate is the doubly geometric distribution, whose probability mass function is given by

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

Then, the transition probability is $p(y' \mid x) = p(x + v \mid x) = P_V\{V = y' - x\}$. To be consistent with the tick-by-tick data analyzed in Section 5.4, we set $M = 8$.

For the clustering noise, the sequence of biasing functions $\{b_i(\cdot)\}$ is assumed to be independent of the value process X and of the nonclustering noise $\{V_i\}$. A biasing function for the case of $M = 8$ is constructed next.[5]

Empirical findings of Harris (1991) show that integers of prices are most likely, halves are second most likely, odd quarters follow and have roughly the same probabilities, and odd eighths are least likely and also have roughly the same probabilities. To generate such clustering, a biasing function is constructed based on the following biasing rules: if the fractional part of y' is even eighths, then y stays on y' with probability 1; if the fractional part of y' is an odd eighth, then y stays on y' with probability $1 - \alpha - \beta - \gamma$, y moves to the closest odd quarter with probability α, to the closest half with probability β, and to the closest integer with probability γ.

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

$$r(y) = \begin{cases} 3 & \text{if the fractional part of y is } \frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8} \\ 2 & \text{if the fractional part of y is } \frac{1}{4}, \frac{3}{4} \\ 1 & \text{if the fractional part of y is } \frac{1}{2} \\ 0 & \text{if y is an integer.} \end{cases}$$

Relying on the simplified notation of Section 2.1, the set of biasing rules specifies the transition probability of y' to y, $p(y \mid y')$. Define $D = 8|y - R[x, \frac{1}{8}]|$ as the number of ticks of the difference between the price and the closest tick of the value, where y is a price level. Note that the nonclustering noise determines $p(y' \mid x) = P\{V = y' - R[x, \frac{1}{8}]\}$. Then the transition probability, $p(y \mid x)$ can be computed using the formula in equation (2.2).

For example, when $r(y) = 2$(y is an odd quarter), two cases emerge:

Case 1. $D = 0$ results in three subcases: $V = 0$, $V = 1$, and $V = -1$, and

$$p(y \mid x) = P\{V = 0\} + (P\{V = 1\} + P\{V = -1\})\alpha = P\{V = 0\} + 2\alpha P\{V = 1\}.$$

Case 2. For $D \geq 1$, similar results emerge.

Plugging in the probability mass function of V results in

$$p(y \mid x) = \begin{cases} (1-\rho)(1+\alpha\rho) & \text{if } r(y) = 2 \text{ and } D = 0 \\ \frac{1}{2}(1-\rho)[\rho + \alpha(2+\rho^2)] & \text{if } r(y) = 2 \text{ and } D = 1 \\ \frac{1}{2}(1-\rho)\rho^{D-1}(\rho + \alpha(1+\rho^2)) & \text{if } r(y) = 2 \text{ and } D \geq 2 \end{cases}. \tag{5.1}$$

The rest of $p(y \mid x)$ can be obtained similarly.

[5] Although $M = 100$ is now in the NYSE, the idea of constructing a biasing function is still valid.

The simplified model is similar to the structural model proposed by Hasbrouck (1996). It also induces the autocorrelation pattern consistent with extant empirical findings in financial economics. That is, the first autocorrelation of asset returns is close to zero and negative for interday data, but is near -0.5 for intraday data; second and higher order autocorrelations are zero for both data.

For the simplified model, there are six parameters, $(\mu, \sigma, \rho, \alpha, \beta, \gamma)$. The first two relate to the value process, ρ relates to the nonclustering noise, and the last three relate to the clustering noise.

The clustering parameters α, β, and γ are estimated by the method of relative frequency, which is a variant of the method of moments. This implies that the relative frequency estimates are consistent and asymptotically normal. Let f_i be the sample relative frequency for $r(y) = i$, for $i = 0, 1, 2, 3$. Assuming that the fractional parts of X are uniformly distributed, the method of relative frequency implies $f_0 = \frac{1}{8} + \frac{1}{2}\gamma$, $f_1 = \frac{1}{8} + \frac{1}{2}\beta$, $f_2 = \frac{1}{4} + \frac{1}{2}\alpha$, $f_3 = \frac{1}{2}(1 - \alpha - \beta - \gamma)$ with the unique solutions:

$$\hat{\alpha} = 2\big(f_2 - \tfrac{1}{4}\big), \qquad \hat{\beta} = 2\big(f_1 - \tfrac{1}{8}\big), \qquad \hat{\gamma} = 2\big(f_0 - \tfrac{1}{8}\big).$$

The parameters μ, σ, and ρ are then estimated by the Bayes estimation via filtering.

5.2. Bayes Estimation for the Simplified Model

In this section, the recursive algorithm for the simplified model is constructed in detail and the consistency of Bayes estimates is proved.

5.2.1. Recursive Algorithm. The steps in Section 4.2 are followed to construct a recursive algorithm.

For the preparation step, we define $\vec{\xi} = (\mu, \sigma, \rho)$ and the parameter space $\Xi = [\alpha_\mu, \beta_\mu] \times [\alpha_\sigma, \beta_\sigma] \times [\alpha_\rho, \beta_\rho] \times [\alpha_x, \beta_x]$. Because there are no time-depedent parameters, we only consider $(\vec{\xi}, X)$. The filtering equation is still equation (3.3) except the generator becomes

$$\mathbf{A} f(\mu, \sigma, \rho, x) = \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 f}{\partial x^2}(\mu, \sigma, \rho, x) + \mu x \frac{\partial f}{\partial x}(\mu, \sigma, \rho, x), \tag{5.2}$$

and the transition probability $p_k = p(y_k \mid x)$ is specified by equation (5.1).

In the second step, we discretize the parameter spaces of μ, σ, ρ and the state space of X. Suppose there are $n_\mu + 1$, $n_\sigma + 1$, $n_\rho + 1$, and $n_x + 1$ lattices in the discretized spaces of μ, σ, ρ, 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 $n_\mu + 1$. Let $\mu_j = \alpha_\mu + j\epsilon_\mu$ be the jth lattice in the discretized parameter space of μ. Similarly, define $\sigma_k = \alpha_\sigma + k\epsilon_\sigma$, $\rho_m = \alpha_\rho + m\epsilon_\rho$, and $x_l = X_l(t) = \alpha_x + l\epsilon_x$.

We apply the central difference approximation[6] to the differentials in equation (5.2) to construct the birth and death generator as follows:

[6] There exist one-sided difference approximations and central difference approximations and they are asymptotically identical.

(5.3)

$$
\begin{aligned}
&\mathbf{A}_\varepsilon f(\mu_j, \sigma_k, \rho_m, x_l) \\
&\quad = \mu_j x_l \left(\frac{f(\mu_j, \sigma_k, \rho_m, x_l + \epsilon_x) - f(\mu_j, \sigma_k, \rho_m, x_l - \epsilon_x)}{2\epsilon_x} \right) \\
&\qquad + \frac{1}{2}\sigma_k^2 x_l^2 \left(\frac{f(\mu_j, \sigma_k, \rho_m, x_l + \epsilon_x) + f(\mu_j, \sigma_k, \rho_m, x_l - \epsilon_x) - 2f(\mu_j, \sigma_k, \rho_m, x_l)}{\epsilon_x^2} \right) \\
&\quad = a(\mu_j, \sigma_k, x_l)(f(\mu_j, \sigma_k, \rho_m, x_l + \epsilon_x) - f(\mu_j, \sigma_k, \rho_m, x_l)) \\
&\qquad + b(\mu_j, \sigma_k, x_l)(f(\mu_j, \sigma_k, \rho_m, x_l - \epsilon_x) - f(\mu_j, \sigma_k, \rho_m, x_l)),
\end{aligned}
$$

where

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

REMARK 5.1. $a(\mu_j, \sigma_k, x_l)$ is the birth rate and $b(\mu_j, \sigma_k, x_l)$ is the death rate[7] for the birth and death process.

In the third step, let the generator $\mathbf{A}_\varepsilon$ be defined as equation (5.3).

In the fourth step, we assume $\mu_\epsilon = \mu_{\epsilon_\mu}$, $\sigma_\epsilon = \sigma_{\epsilon_\sigma}$, $\rho_\epsilon = \rho_{\epsilon_\rho}$, and $X_\epsilon = X_{\epsilon_x}$ in the indicator function. Then

$$
\mathbf{I}_{\{\mu_\epsilon = \mu_j, \sigma_\epsilon = \sigma_k, \rho_\epsilon = \rho_m, X_\epsilon(t) = x_l\}}(\mu_\epsilon, \sigma_\epsilon, \rho_\epsilon, X_\epsilon(t)) \stackrel{\text{def}}{=} \mathbf{I}(\mu_j, \sigma_k, \rho_m, x_l).
$$

The approximate posterior at $(\mu_j, \sigma_k, \rho_m, x_l)$ is defined as

$$
p(\mu_j, \sigma_k, \rho_m, x_l; t) = E^P\big[\mathbf{I}(\mu_j, \sigma_k, \rho_m, x_l) \,\big|\, \mathcal{F}_t^{\vec{Y}_\varepsilon}\big].
$$

To determine $\pi_\varepsilon(\mathbf{A}_\epsilon \mathbf{I}, t)$ in equation (4.6), we observe that

$$
E^P\big[a(\mu_\epsilon, \sigma_\epsilon, X_\epsilon(t))\mathbf{I}(\mu_j, \sigma_k, \rho_m, x_l + \epsilon_x) \,\big|\, \mathcal{F}_t^{\vec{Y}_\varepsilon}\big] = a(\mu_j, \sigma_k, x_{l-1}) p(\mu_j, \sigma_k, \rho_m, x_{l-1}; t)
$$

and

$$
E^P\big[b(\mu_\epsilon, \sigma_\epsilon, X_\epsilon(t))\mathbf{I}(\mu_j, \sigma_k, \rho_m, x_l - \epsilon_x) \,\big|\, \mathcal{F}_t^{\vec{Y}_\varepsilon}\big] = b(\mu_j, \sigma_k, x_{l+1}) p(\mu_j, \sigma_k, \rho_m, x_{l+1}; t).
$$

Then, along with two similar results, $\pi_\varepsilon(\mathbf{A}_\varepsilon \mathbf{I}, t)$ in equation (4.6) becomes explicit as

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

and the recursive equation (4.8) can be written as

$$
\begin{aligned}
(5.4) \quad p(\mu_j, \sigma_k, \rho_m, x_l; t_{i+1}-) \approx {} & p(\mu_j, \sigma_k, \rho_m, x_l; t_i) \\
& + [a(\mu_j, \sigma_k, x_{l-1}) p(\mu_j, \sigma_k, \rho_m, x_{l-1}; t_i) \\
& \quad - (a(\mu_j, \sigma_k, x_l) + b(\mu_j, \sigma_k, x_l)) p(\mu_j, \sigma_k, \rho_m, x_l; t_i) \\
& \quad + b(\mu_j, \sigma_k, x_{l+1}) p(\mu_j, \sigma_k, \rho_m, x_{l+1}; t_i)](t_{i+1} - t_i).
\end{aligned}
$$

[7] These rates 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.

When a trade at the k_0th price level, or a jump at Y_{k_0}, occurs at time t_{i+1}, the updating equation (4.7) turns out to be

$$(5.5)\qquad p(\mu_j, \sigma_k, \rho_m, x_l; t_{i+1}) = \frac{p(\mu_j, \sigma_k, \rho_m, x_l; t_{i+1}-) p(y_{k_0} \mid x_l; \rho_m)}{\sum_{j',k',m',l'} p(\mu_{j'}, \sigma_{k'}, \rho_{m'}, x_{l'}; t_{i+1}-) p(y_{k_0} \mid x'_l, \rho'_m)},$$

where the summation is over the discretized spaces of μ, σ, ρ and $X(t_{i+1}-)$; and $p(y_{k_0} \mid x_l, \rho_m)$ is the transition probability from x_l to y_{k_0}, which also depends on ρ_m. Then, $p(y_{k_0} \mid x_l, \rho_m)$ is specified by equation (5.1).

Similarly, equations (5.4) and (5.5) are the recursive algorithm we employ to calculate the posteriors at time t_{i+1} for $(\mu, \sigma, \rho, X(t_{i+1}))$ based on the posteriors at time t_i.

Last, we choose a reasonable prior. We assume the independence of $X(0)$ and (μ, σ, ρ), and set $P\{X(0) = Y(t_1)\} = 1$, where $Y(t_1)$ is the first trade price data. If there is no special prior information on (μ, σ, ρ) available, it is reasonable to assign uniform distributions to the discretized state space of μ, σ, ρ. Based on the preceding arguments, we obtain the prior at $t = 0$ as

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

5.2.2. Consistency of the Bayesian Estimates.

THEOREM 5.1. *Under the setup of Section 5.1, suppose that the clustering parameters α, β, γ are known, and (μ, σ, ρ) has a prior. Then $E[f(\mu, \sigma, \rho) \mid \mathcal{F}_t^{\vec{Y}}] \to f(\mu, \sigma, \rho)$ a.s. as $t \to \infty$ when $\mu - \frac{1}{2}\sigma^2 > 0$ for any bounded continuous function f.*

Proof: See Appendix B.

REMARK 5.2. The standard condition "$\mu - \frac{1}{2}\sigma^2 > 0$" rules out bankruptcy.

Theorem 5.1 together with the convergence of the recursive algorithm implies that the computed Bayes estimates will converge to their true values. This is confirmed by the simulation results presented next.

5.3. Simulation Study

Under standard simulation design, it can be shown that the Bayes estimates evaluated by the recursive algorithm converge to their true values. In order to demonstrate that the simplified model conforms to the sample characteristics of the real data, a special set of parameters is selected.

We first describe a one-month (March 1994) transaction data for Microsoft, which serves as the benchmark for the simulation study. The data are extracted from the Trade and Quote (TAQ) database distributed by NYSE. To filter the data we apply standard procedures[8] but with one important exception. Previous studies (e.g. Engle and Russell

[8] The market is open for trading between 9:30 a.m. to 4:00 p.m. and eligible trades must be time-stamped within this trading interval. All trades executed outside the regular trading hours are removed. Trades that are canceled later and wrongly recorded are marked in the TAQ data; these trades are also removed. Late reported trades, which are transactions that are reported to the tape at a time later than they occurred and when other trades occurred between the time of the transactions, are also marked in the TAQ data. To be conservative, these trades are also removed.

1998) cannot handle multiple trades at a given point in time even though the prices can be different. Consequently, those trades with zero time duration were excluded. The present method can handle such cases and therefore we keep all zero duration prices. The final sample is composed of 32,348 observations.

5.3.1. Micromovement Features of Simulated Data. For simulation, we choose the parameter values close to estimates from the Microsoft data. Let $\mu = 4.4 \times 10^{-8}$, $\sigma = 1.2 \times 10^{-4}$, $\rho = 0.20$, $\alpha = 0.225$, $\beta = 0.066$, and $\gamma = 0.3$. Since $a(t)$ has no impact in estimation and noise, we assume the trading intensity is constant: $a(t) = 0.06$ for all $t > 0$ (i.e., one trade in about $1/0.06 = 16.67$ seconds). Using these parameters, we simulated 32,500 observations.

Figure 5.1 shows the impact of the nonclustering noise and clustering noise on the distribution of price changes and the impact on the fractional parts of price. Figure 5.1 has four pairs of histograms for these measures. The first three pairs are produced from the simulated data of the rounding model, the rounding plus nonclustering noise model, and the simplified model (rounding plus nonclustering and clustering noise). The last pair is from the actual Microsoft data. The first three histograms in the left column show how price changes are concentrated on at most one tick, then spread out geometrically on both sides, and then move toward even eighths, which conforms to the histogram for Microsoft data. The first three histograms in the right column show that the rounding model and the rounding plus nonclustering model do not conform to the histogram for Microsoft. However, adding the clustering noise does produce price clustering phenomenon similar to the Microsoft histogram.

5.3.2. Bayes Estimates of the Simulated Data. A Fortran program for the recursive algorithm is constructed to calculate, at each trading time t_i, the joint posterior of (μ, σ, ρ, X), their marginal posteriors, their Bayes estimates, and their standard errors, respectively.

Figure 5.2 has three plots that show how the Bayesian estimates and their two-standard-errors (SE) bounds evolve in comparison with the true values for μ, σ, and ρ respectively. The figure clearly shows that the estimates of μ and σ converge to their true values, the two-SE bounds shrink, and the true values are within the two-SE bounds. Also, the true value of ρ is always within the three-SE bounds. Similarly, the Bayes estimates of $X(t)$ are close to their true values, which are always within the two-SE bounds.

The final posteriors of μ, σ, and ρ are presented in Table 5.1. Note that the lattices do not include the true values in order to ensure a fair comparison. The true values, final Bayes estimates, and their standard errors are listed below the table. The Bayes estimates are close to the true values, which are within three SE. The SE of μ is much larger than the SEs of σ and ρ (comparing to the Bayes estimates). This makes sense because μ is a trend parameter, and its estimation accuracy depends on the length of time covered by the data (23 days), but the accuracy of σ and ρ estimates mainly depends on the number of observations (32,500).

5.4. An Application to Real Data

In this section, the recursive algorithm is applied to the Microsoft data to obtain the Bayes estimates, to show the algorithm is fast enough to provide real-time estimates, and to present a straightforward financial application of the simplified model.

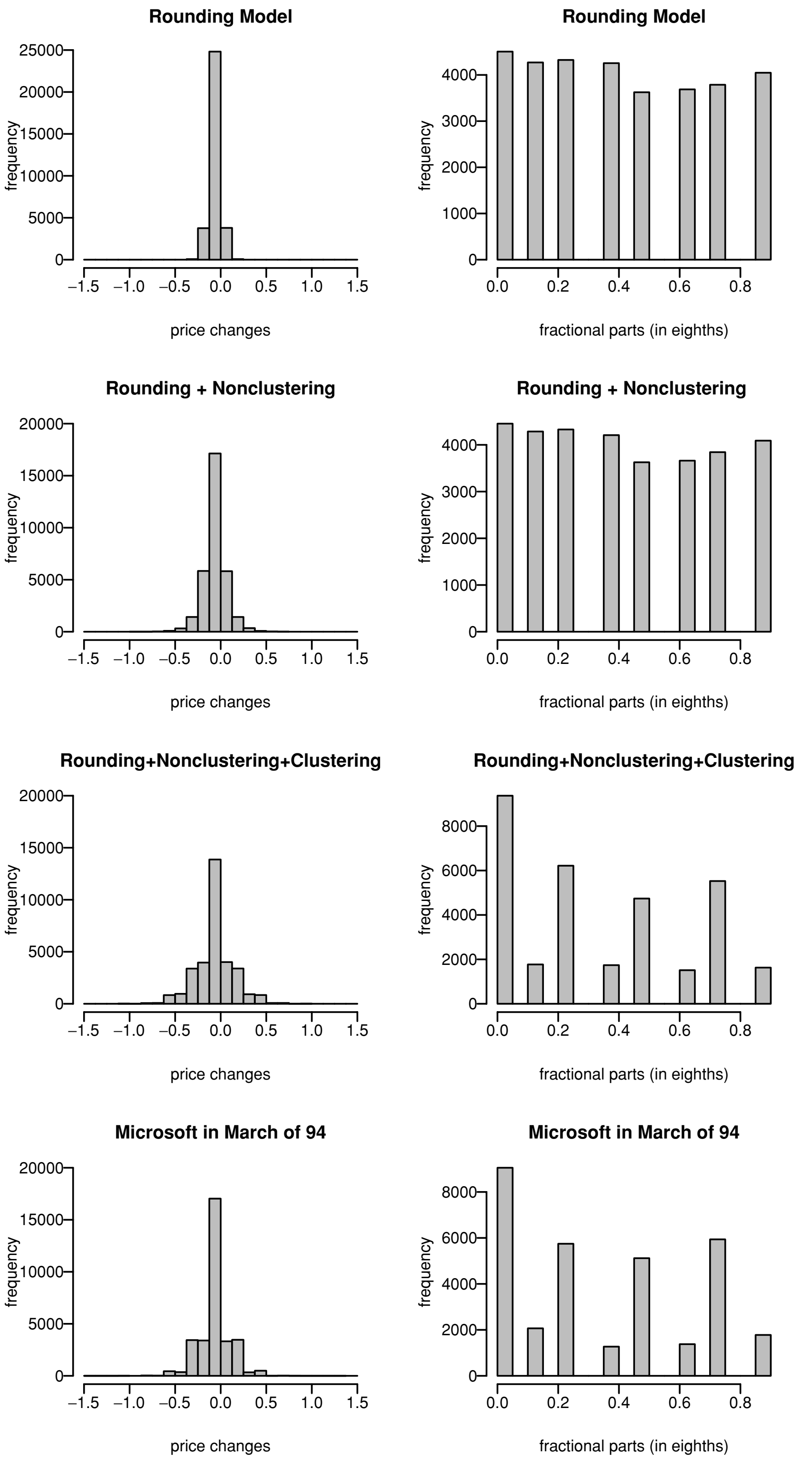

FIGURE 5.1. The impact of nonclustering and clustering noise on the distribution of price changes and on the fractional parts of prices.

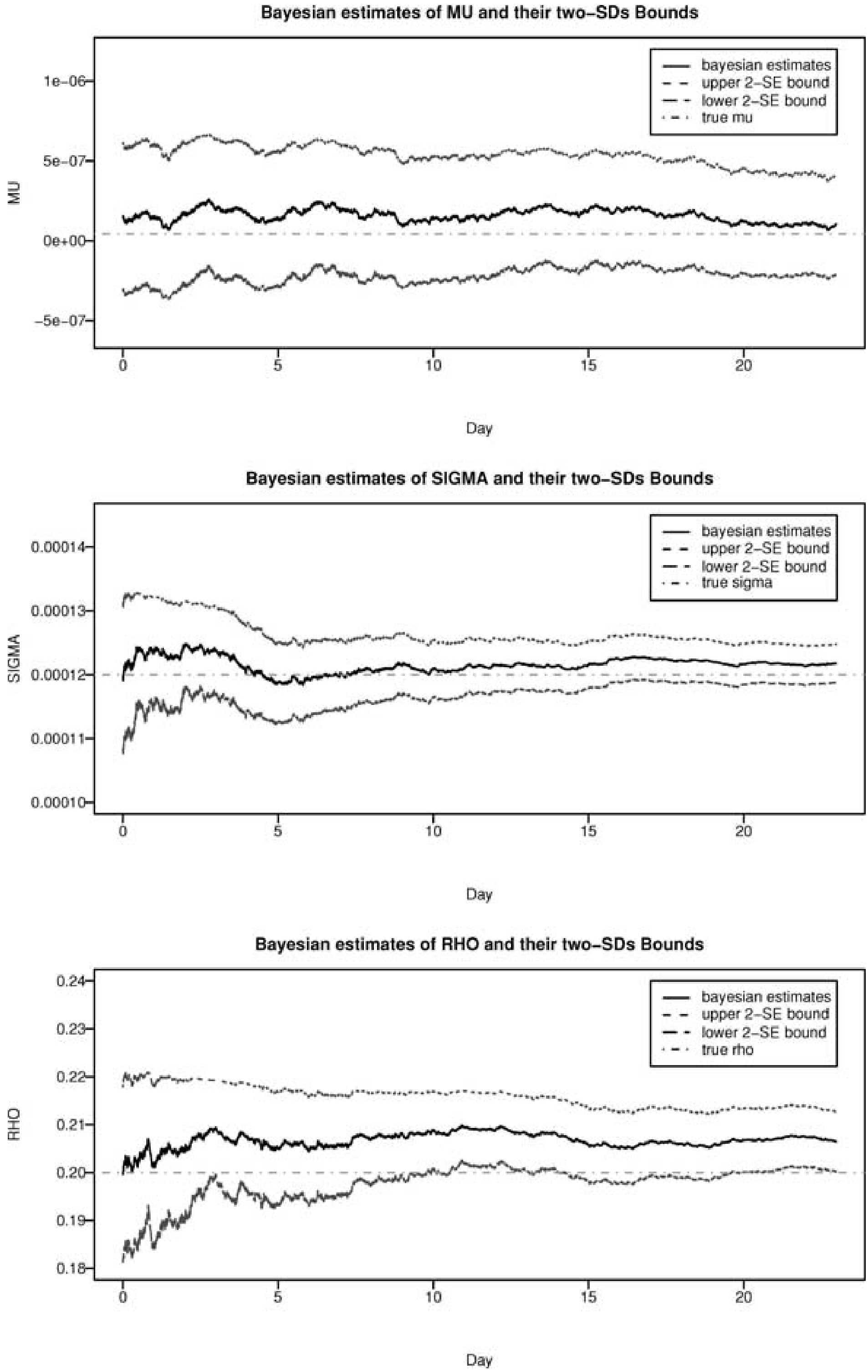

FIGURE 5.2. Bayesian estimates of μ, σ, and ρ and the two-standard-errors bounds.

TABLE 5.1
Final Posteriors of μ, σ, and ρ

μ	p(μ)	σ	p(σ)	ρ	p(ρ)
−2.0e−7	0.0481	1.110e−4	3.804e−13	0.186	2.105e−10
−1.0e−7	0.1187	1.135e−4	5.646e−8	0.190	4.638e−7
0.0	0.2041	1.160e−4	2.825e−4	0.194	1.844e−4
1.0e−7	0.2446	1.185e−4	0.0586	0.198	0.0135
2.0e−7	0.2041	1.210e−4	0.5861	0.202	0.1878
3.0e−7	0.1186	1.235e−4	0.3420	0.206	0.5043
4.0e−7	0.0481	1.260e−4	0.0134	0.210	0.2661
5.0e−7	0.0136	1.285e−4	4.108e−5	0.214	0.0280

True $\mu = 4.40e{-8}$, $E(\mu) = 1.054e{-7}$, and $SE(\mu) = 1.561e{-7}$.
True $\sigma = 1.20e{-4}$, $E(\sigma) = 1.218e{-4}$, and $SE(\sigma) = 1.498e{-6}$.
True $\rho = 0.20$, $E(\rho) = 0.2064$, and $SE(\rho) = 0.0031$.

Based on the relative frequencies of the fractional parts of the price for Microsoft, we estimate $\alpha = .224$, $\beta = .066$, and $\gamma = .31$. 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 simplified model. The mean duration is 16.64 seconds with a standard error of 28.08 seconds.

Now, the Fortran program is applied to the transaction data of Microsoft. It takes about three hours to obtain the Bayes estimates for each trade in the whole data set on a Compaq XP1000 Alpha computer. This running time is much less than the real-time estimates required.[9] Only the final Bayes estimates are presented.

5.4.1. Bayes Estimates for Microsoft Data. As a comparison, we first look at the MLE of GBM for daily closing data. There were 23 daily closing data for March 1994. Assuming 260 business days per year, the annualized results are summarized as item 1 in Table 5.2. The annualized Bayes estimates based on the transaction data are presented as item 2 in Table 5.2.

These two estimates are comparable. The Bayes estimates are more accurate because the model is closer to the real data and much more information is used. Note that the standard errors of the Bayesian estimates are smaller than those of MLEs, as expected.

5.4.2. Does Price Clustering Matter?. Price discreteness and clustering are two striking features of micromovement prices. Hausman, Lo, and Mackinlay (1992) asked whether price discreteness matters for the parameter estimation, and they provided a strong affirmative answer.

It is equally important to ask whether the related price clustering matters, and our simple analysis supplies strong evidence in the affirmative. When clustering noise is ignored, the variability due to clustering noise moves to σ and ρ. If the estimates of σ and ρ increase, supportive evidence is provided on the importance of price clustering.

[9] 23 days × 6.5 hours per day = 149.5 hours, the total time required for real-time estimates.

TABLE 5.2
Annualized MLE and Bayes Estimates for MSFT, 94/3

Model	μ	σ	ρ
1. MLE of daily closing prices	35.12%	26.49%	na
	(140.40%)	(6.24%)	na
2. Bayes estimates of transaction prices	26.53%	28.09%	0.211
with clustering	(46.25%)	(0.42%)	(.0024)
3. Bayes estimates of transaction prices	26.40%	31.42%	0.319
without clustering	(46.73%)	(0.46%)	(0.0032)

Standard errors are in parenthesess.

Ignoring price clustering is equivalent to setting $\alpha = \beta = \gamma = 0$. Then, μ, σ, and ρ are estimated. The Bayes estimates are presented as item 3 in Table 5.2. The Bayes estimate of μ differs little, but the estimates of σ and ρ have increased significantly. This increase in parameter values provides strong evidence that ignoring the existence of price clustering can significantly bias the statistical inference.

6. CONCLUSION

A partially observed micromovement model for asset prices is proposed. The model ties the sample characteristics of micromovement and macromovement in a consistent manner. It can be applied to a variety of asset prices, including stocks prices, exchange rates and commodity prices. The proposed model is purely statistical, though the price process it generates is fully consistent with many behavioral models (for an excellent survey, see O'Hara 1995). The model is a marked point process, which is so general that it includes many structurally different models.[10] Engle (2000) also proposed a marked point process to model micromovement of stock prices. Unlike Engle, the model proposed in this paper is partially observed and complete information of prices and trading times is utilized for parameter estimation.

The main appeal of the proposed model is that the transactional price process can be framed as a filtering problem with counting process observations. Under this framework, complete information can be used for parameter estimation. We introduced Bayes estimation via filtering for the parameters as well as the underlying value process. The core of Bayes estimation via filtering is to construct a consistent recursive algorithm. The algorithm can compute the Bayes estimates for time-invariant as well as time-dependent parameters. To illustrate the mechanics of implementing the proposed model, a simplified version of the model was studied in detail.

This paper opens up at least five directions for future research: model extension, model selection, computation, option pricing, and deeper studies in market microstructure.

[10] Such as the autoregressive conditional duration model by Engle and Russell (1998), the autoregressive conditional multinomial model by Russell and Engle (1998), the threshold autoregressive conditional duration model by Zhang, Russell, and R. S. Tsay (2000), the univariate activity, direction and size model by Rydberg and Shephard (1998), and the stochastic-intensity point process by Rogers and Zane (1998).

The simplified model can be extended in several ways to improve the empirical applications. First, the value process can be generalized to stochastic volatility and jump-diffusion models. Second, the clustering and nonclustering noises can assume other forms or distributions to accommodate different tick sizes and different sample characteristics. Third, the trading intensity $a(t)$ can be generalized to $a(\theta(t), X(t), t)$. Furthermore, the model itself can be extended to a multivariate setting, and to include other observable variables such as ask and bid quotes and trading volume. Some of these extensions are currently being investigated by the author.

Model selection is a significant and persistent area of research as many competing models can be fitted to intraday price data. Based on the Bayes factor, a Bayesian approach provides a methodology for hypothesis testing and model selection that accommodates irregular likelihood functions such as those under the general framework of this paper. The Bayesian model selection method provides a powerful tool for testing economic hypothesis related to market microstructure theory. Some of these issues are also being pursued by the author.

The recursive algorithm suffers from the "curse of dimensionality" as the simplified model is extended. It is worthwhile to develop new computing algorithms to improve efficiency.

The proposed model describes trade-by-trade price behavior and can provide trade-by-trade Bayes parameter estimates. Both of these supply the potential to develop a practically important trade-by-trade option pricing formula and trade-by-trade hedge portfolio under our framework, which incorporates transitory noise.[11]

There are other direct applications of the proposed model. For example, one can study the estimation bias induced by trading noises as in Gottlieb and Kalay (1985), Cho and Frees (1988), and Ball (1988), compare the Bayes estimates with other simpler approaches such as the discrete-time state-space model of Kitagawa (1987), assess the quality of security market as in Hasbrouck (1993), compare information flows in trading and nontrading periods and compare noise in call and auction markets as in Forster and George (1995) and George and Hwang (1998).

APPENDIX A

Proof of Theorem 3.1. There are two basic approaches to derive filtering equation: filtering via innovation and filtering via reference measure (see the classic book, Kallianpur 1980). We adopt the latter here. There are three steps in the proof.

Step 1. Determine the SDE for $\phi(f, t)$.

We start with integration by parts (see Protter 1992, p. 60):

$$U(t)V(t) = U(0)V(0) + \int_0^t U(s-)\,dV(s) + \int_0^t V(s-)\,dU(s) + [U, V]_t$$

and take $U(t) = f(\theta(t), X(t))$, $V(t) = L(t)$. Assumption 2.3 implies $f(\theta(t), X(t))$ and $L(t)$ have no simultaneous jumps w.p.1, which implies $[f(\theta, X), L]_t = 0$. Coupled with the fact that $[\int \mathbf{A} f(\theta(s), X(s))\,ds, L]_t = 0$, it implies $[M_f, L]_t = 0$. Then the Girsanov-Meyer

[11] Bertsimas, Kogan, and Lo (2000) studied the discrete-time approximate error in option pricing, but they did not include transitory noise.

Theorem (see Protter 1992, p. 109) implies that $M_f(t)$ is also a martingale under Q. Then

(A.1)

$$
\begin{aligned}
f(\theta(t), X(t))L(t) = {} & f(\theta(0), X(0))L(0) + \int_0^t L(s-)\,dM_f(s) \\
& + \sum_{k=1}^n \int_0^t f(\theta(s-), X(s-))[\lambda_k(\theta(s-), X(s-), s-) - 1]L(s-)\,dY_k(s) \\
& + \int_0^t L(s)\Bigg\{\mathbf{A} f(\theta(s), X(s)) - \sum_{k=1}^n \int_0^t f(\theta(s), \\
& \qquad X(s))[\lambda_k(\theta(s), X(s), s) - 1]\Bigg\}\,ds.
\end{aligned}
$$

We take conditional expectations with respect to the reference measure Q given the observed history of $\mathcal{F}_t^{\vec{Y}}$ on both sides of equation (A.1). We state four lemmas that deal with the four terms on the right-hand side of equation (A.1). The proofs of these lemmas are available upon request.

LEMMA A.1. *Suppose that X has finite expectation and is $\mathcal{H}$-measurable, and that $\mathcal{D}$ is independent of $\mathcal{H} \vee \mathcal{G}$. Then, $E[X \mid \mathcal{G} \vee \mathcal{D}] = E[X \mid \mathcal{G}]$.*

Let $\mathcal{D} = \mathcal{F}_{0<s\le t}^{\vec{Y}}$, which is independent of $\mathcal{G} = \mathcal{F}_0^{\vec{Y}}$ and $\mathcal{H} = \mathcal{F}_0^{\theta, X, \vec{Y}}$. Then Lemma A.1 implies $E^Q[f(\theta(0), X(0))L(0) \mid \mathcal{F}_t^{\vec{Y}}] = E^Q[f(\theta(0), X(0))L(0) \mid \mathcal{F}_0^{\vec{Y}}] = \phi(f, 0)$.

LEMMA A.2. *Suppose $\vec{X}$ and $\vec{Y}$ are independent. Let $M(t)$ be a martingale with respect to $\{\mathcal{F}_t^{\vec{X}}\}$ and let U be $(\mathcal{F}^{\vec{X}} \vee \mathcal{F}^{\vec{Y}})_t$-predictable. If $E[|\int_0^t U(s)\,dM(s)|] < \infty$. Then $E[\int_0^t U(s)\,dM(s) \mid \mathcal{F}_t^{\vec{Y}}] = 0$.*

Under the reference measure Q, (θ, X) and $\vec{Y}$ are independent and $M_f(t)$ is still a $\mathcal{F}_t^{\theta, X}$-martingale. $U(t) = L(t)$ is $(\mathcal{F}^{\theta, X} \vee \mathcal{F}^{Y})_t$-predictable and Assumption 2.5 ensures that $L(t)$ is uniformly bounded. Since $M_f(t)$ is bounded for every t, $E^Q[|\int_0^t L(s-)\,dM_f(s)|] < \infty$. Therefore, Lemma A.2 implies $E^Q[\int_0^t L(s-)\,dM_f(s) \mid \mathcal{F}_t^{\vec{Y}}] = 0$.

LEMMA A.3. *Suppose that $\vec{Y} = (Y_1, Y_2, \ldots, Y_n)$ and $\vec{X}, Y_1, Y_2, \ldots, Y_n$ are independent. U is $(\mathcal{F}^{\vec{X}} \vee \mathcal{F}^{\vec{Y}})_t$-predictable. Suppose that Y_k is a unit Poisson process and $E[\int_0^t |U(s)|\,ds] < \infty$. Then,*

$$
E\left[\int_0^t U(s-)\,dY_k(s) \,\Big|\, \mathcal{F}_t^{\vec{Y}}\right] = \int_0^t E\big[U(s-) \mid \mathcal{F}_{s-}^{\vec{Y}}\big]\,dY_k(s).
$$

Under Q, similarly, the conditions of Lemma A.3 are satisfied. Then

$$
\begin{aligned}
& E^Q\left[\int_0^t f(\theta(s-), X(s-))[\lambda_k(\theta(s-), X(s-), s) - 1]L(s-)\,dY_k(s) \,\Big|\, \mathcal{F}_t^{\vec{Y}}\right] \\
& \quad = \int_0^t \phi((\lambda_k - 1)f, s-)\,dY_k(s).
\end{aligned}
$$

LEMMA A.4. *Suppose that $\vec{X}$ and $\vec{Y}$ are independent. If U is $(\mathcal{F}^{\vec{X}} \vee \mathcal{F}^{\vec{Y}})_t$-adapted, satisfying $\int_0^t E[\,|\,U(s)\,|\,]\,ds < \infty$, then*

$$E^Q\left[\int_0^t U(s)\,ds \mid \mathcal{F}_t^{\vec{Y}}\right] = \int_0^t E^Q\big[U(s) \mid \mathcal{F}_s^{\vec{Y}}\big]\,ds.$$

Lemma A.4 implies

$$E^Q\left[\int_0^t L(s)\left\{\mathbf{A}f(\theta(s), X(s)) - \sum_{k=1}^n \int_0^t f(\theta(s), X(s))[\lambda_k(\theta(s), X(s), s) - 1]\right\} ds \mid \mathcal{F}_t^{\vec{Y}}\right]$$
$$= \int_0^t \phi\left(\mathbf{A}f - \sum_{k=1}^n (\lambda_k - 1)f, s\right) ds = \int_0^t \phi(\mathbf{A}f - (a-n)f, s)\,ds.$$

Summarizing the above, we have the SDE for $\phi(f, t)$, which is equation (3.1).

Step 2. Determine the SDE for $\pi(f, t)$.

Note $\pi(f, t) = \frac{\phi(f,t)}{\phi(1,t)}$. Apply Itô's formula (see Protter 1992, p. 74) to $f(X, Y) = \frac{X}{Y}$ with $X = \phi(f, t)$ and $Y = \phi(1, t)$. After some simplifications we have

(A.2)

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

A last step remains to make the integrand of the last integral predictable. Assume a trade at Y_k occurs. Then

$$\pi(f, s) = \frac{\phi(f, s)}{\phi(1, s)} = \frac{\phi(f, s-) + \phi(f(\lambda_k - 1), s-)}{\phi(1, s-) + \phi(\lambda_k - 1, s-)} = \frac{\phi(f\lambda_k, s-)}{\phi(\lambda_k, s-)} = \frac{\pi(fap_k, s-)}{\pi(ap_k, s-)}.$$

Hence, equation (A.2) implies equation (3.2).

When $a(\theta(t), X(t), t) = a(t)$, two additional observations further simplify equation (3.2) to equation (3.3). First, $\pi(fa, t) = a(t)\pi(f, t)$, and $\pi(a, t) = a(t)$. Then, the two terms in the integrand of ds in equation (3.2) cancel out. Second, $\frac{\pi(fap_k,s-)}{\pi(ap_k,s-)} = \frac{\pi(fp_k,s-)}{\pi(p_k,s-)}$.

Step 3. Prove the Uniqueness.

We follow Kliemann Koch, and Marchetti (1990). They formulate the problem as a *filtered martingale problem* (FMP) proposed by Kurtz and Ocone (1988) and then apply Kurtz and Ocone's result about the unique solution for the FMP. For our case, Assumption 2.5 ensures the existence of a measure, under which the $\vec{Y}$ are counting processes with intensities $\pi(\lambda_k, t)$ for $k = 1, 2, \ldots, n$ and the FMP can be defined. Then, the uniqueness of FMP gives the uniqueness of the filtering equations. We refer the interested reader to Kliemann et al. and Kurtz and Ocone for details of the proof. □

APPENDIX B

Proof of Theorem 5.1: Set $\vec{\xi} = (\mu, \sigma, \rho)'$. We know $E[f(\vec{\xi}) \mid \mathcal{F}_t^{\vec{Y}}]$ is a martingale converging to $E[f(\vec{\xi}) \mid \mathcal{F}_\infty^{\vec{Y}}]$ a.s. If we can show that $\vec{\xi}$ is $\mathcal{F}_\infty^{\vec{Y}}$-measurable, then $E[f(\vec{\xi}) \mid \mathcal{F}_\infty^{\vec{Y}}] = f(\vec{\xi})$ and the consistency follows.

To show that $\vec{\xi}$ is $\mathcal{F}_\infty^{\vec{Y}}$-measurable, it suffices to construct a.s. consistent estimates of μ, σ, and ρ based on $\mathcal{F}_\infty^{\vec{Y}}$.

First, we construct a consistent estimate for $\mu - \frac{1}{2}\sigma^2$. It is simple to check that $\frac{1}{t}(\log Y_t - \log Y_0) \to \mu - \frac{1}{2}\sigma^2$ a.s. as $t \to \infty$. Then, $\frac{1}{t}[\log Y(t) - \log Y(0)]$ is an a.s. consistent estimate for $\mu - \frac{1}{2}\sigma^2$.

Next, we construct an a.s. consistent estimate for σ^2. Required terms are defined prior to Lemma B.1. First, we define a sequence of stopping times $\{\tau_i\}$ for each $n > 0$. Define $\tau_1 = t_1$, which is the first trading time. Then recursively define $\tau_{i+1} = \inf_j\{t_j, t_j > \tau_i + n\}$, which is the first trading time after $\tau_i + n$.

Second, for any $K > 1.5$, we define the "simplified-K" model, which is the simplified model described in Section 6.1 except that the nonclustering noise V is replaced by a truncated version $\tilde{V}$, defined as

$$\tilde{V}_i = \begin{cases} V_i & \text{if } V_i \geq -K + 1 \\ -K + 1 & \text{if } V_i < -K + 1 \end{cases}.$$

We focus on constructing an a.s. consistent estimate of σ^2 for the simplified-K model, because the truncation makes the following inequality work for the case when $a = Y(\tau_{i,1})$ and $b = X(\tau_{i,1})$. The inequality is when $a > 1$ and $b > 1$, $|\log(a) - \log(b)| \leq |a - b|$, which is used in proving Lemma B.1.

Third, based on $\{\tau_i\}$ and for any $K > 1.5$, we define a sequence of paired stopping times $\{(\tau_{i,1}^K, \tau_{i,2}^K)\}$ as follows. The first pair $(\tau_{1,1}, \tau_{1,2})$ is the first pair in $\{\tau_i\}$ that satisfies three conditions: (i) $\tau_{1,1}$ and $\tau_{1,2}$ are consecutive in $\{\tau_i\}$, (ii) $\tau_{1,1} < \tau_{1,2}$, and (iii) $Y(\tau_{1,1}) \geq K$ and $Y(\tau_{1,2}) \geq K$. Then, recursively define the $(i+1)$th pair $(\tau_{i+1,1}, \tau_{i+1,2})$ in $\{\tau_i\}$ as the first pair after $\tau_{i,1}$ (note that $\tau_{i+1,1}$ can be $\tau_{i,2}$) that satisfies the similar three conditions: (i) $\tau_{i+1,1}$ and $\tau_{i+1,2}$ are consecutive in $\{\tau_i\}$, (ii) $\tau_{i+1,1} < \tau_{i+1,2}$, and (iii) $Y(\tau_{i+1,1}) \geq K$ and $Y(\tau_{i+1,2}) \geq K$.

Since $\mu - \frac{1}{2}\sigma^2 > 0$, we can always have such a sequence of paired stopping times for any $n > 0$ and $K > 1.5$.

Fourth, for each positive integer n and $K > 1.5$, we define

$$D_{i,n}^K = \frac{\log\left(Y(\tau_{i,2}^K)\right) - \log\left(Y(\tau_{i,1}^K)\right)}{\sqrt{n}}$$

for $i = 1, 2, \ldots, n$.

Finally, define

$$\hat{\sigma}_{n,K}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(D_{i,n}^K - \bar{D}_n^K\right)^2,$$

where $\bar{D}_n^K = \frac{1}{n}\sum_{i=1}^{n} D_{i,n}^K$.

LEMMA B.1. *For $K > 1.5$, $\hat{\sigma}_{n,K}^2 \to \sigma^2$ a.s. as $n \to \infty$.*

The proof is available upon request.

Then, letting $K \to \infty$, we have $\hat{\sigma}^2_{n,K}$ as the a.s. consistent estimate of σ^2.

To construct a consistent estimate for ρ, we choose two consecutive trades $Y(t_{i-1})$ and $Y(t_i)$ such that both fractional parts of prices are odd eighths and $t_i - t_{i-1} < \delta$. For any small $\delta > 0$, we can always obtain infinitely many such pairs of consecutive trades as $t \to \infty$, because the probability that two trades occur within the time span of δ is strictly positive given $C_1 < a(t)$.

Note that when $Y(t_i)$ is an odd eighth, then $Y(t_i) = R[X(t_i), \frac{1}{M}] + V_i$. When $t_{i+1} - t_i < \delta$ and as δ goes to zero, $R[X(t_{i+1}), \frac{1}{M}] = R[X(t_i), \frac{1}{M}]$. So, when δ is very small, $Y(t_{i+1}) - Y(t_i) = V_{i+1} - V_i$. And when both $Y(t_{i+1})$ and $Y(t_i)$ are odd eighths, $Y(t_{i+1}) - Y(t_i)$ must be an even eighth.

In such an i.i.d. sequence of $\{Y(t_i) - Y(t_{i-1})\}$, we can observe the empirical relative frequency that $Y(t_i) - Y(t_{i-1}) = 0$ for a fixed time span t, which is denoted by $f_{0,t}$.

We compute

$$P\{V_i - V_{i-1} = 0 \mid V_i - V_{i-1} = 2k, k = 0, \pm 1, \pm 2, \cdots\} = \frac{(2-\rho^2)(1-\rho^2)}{2(1+\rho^2)}.$$

Set $(2-\rho^2)(1-\rho^2)/[2(1+\rho^2)] = f_{0,t}$.

It is simple to check that the above equation has only one root between [0, 1]. This root is an estimate of method of relative frequency. The strong consistency of the estimates of method of relative frequency is well known (Serfling 1980). $\square$

REFERENCES

AÏT-SAHALIA, Y. (2002): Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach, *Econometrica* 70, 223–262.

AMIHUD, Y., and H. MENDELSON (1987): Trading Mechanisms and Stock Returns: An Empirical Investigation, *J. Finance* 42, 533–553.

BAGEHOT, W. (1971): The Only Game in Town, *Financial Anal. J.* 27, 12–14, 22.

BALL, C. (1988): Estimation Bias Induced by Discrete Security Prices, *J. Finance* 43, 841–865.

BARCLAY, M., W. CHRISTIE, J. HARRIS, E. KANDEL, and P. H. SCHULTZ (1999): The Effects of Market Reform on the Trading Costs and Depths of NASDAQ Stocks, *J. Finance* 54 (1), 1–34.

BERTSIMAS, D., L. KOGAN, and A. LO (2000): When Is Time Continuous?, *J. Financial Econ.* 55, 173–204.

BLACK, F. (1986): Noise, *J. Finance* 41, 529–543.

BREMAUD, P., (1981): *Point Processes and Queues: Martingale Dynamics*. New York: Springer-Verlag.

CHAN, L. K. C., and J. LAKONISHOK (1993): Institutional Trades and Intraday Stock Price Behavior, *J. Financial Econ.* 33, 173–199.

CHO, D., and E. W. FREES (1988): Estimating the Volatility of Discrete Stock Prices, *J. Finance* 43, 451–466.

CHRISTIE, W., and P. H. SCHULTZ (1994): Why Do NASDAQ Market Makers Avoid Odd-Eighth Quotes?, *J. Finance* 49, 1813–1840.

COHEN, K., G. HAWAWINI, S. MAIER, R. SCHWARTZ, and D. WHITCOMB (1980): Implications of Microstructure Theory for Empirical Research on Stock Price Behavior, *J. Finance* 35, 249–257.

DUFFIE, D., J. PAN, and K. SINGLETON (2000): Transform Analysis and Asset Pricing for Affine Jump-Diffusions, *Econometrica* 68, 1343–1376.

EASLEY, D., and M. O'HARA (1992): Time and the Process of Security Price Adjustment, *J. Finance* 47, 577–605.

ELLIOTT, R. J., A. LAKHDAR, and J. B. MOORE (1995): *Hidden Markov Models: Estimation and Control*. New York: Springer-Verlag.

ENGLE, R. (2000): The Econometrics of Ultra-High-Frequency Data, *Econometrica* 68, 1–22.

ENGLE, R., and J. RUSSELL (1998): Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data, *Econometrica* 66, 1127–1162.

ETHIER, S., and T. KURTZ (1986): *Markov Processes: Characterization and Convergence*. New York: Wiley.

FORSTER, M., and T. GEORGE (1995): Trading Hours, Information Flow and International Cross-Listing, *Inter. Rev. Financial Anal*. 4, 19–34.

GEORGE, T., and C. HWANG (1998): Information Flow and Pricing Errors: A Unified Approach to Estimation and Testing. Working paper, University of Iowa.

GHOSAL, S., J. K. GHOSH, and T. SAMANTA (1995): On Convergence of Posterior Distributions, *Ann. Statistics* 23, 2145–2152.

GODEK, P. (1996): Why NASDAQ Market Makers Avoid Odd-Eighth Quotes, *J. Financial Econ*. 41, 465–474.

GOGGIN, E. (1994): Convergence in Distribution of Conditional Expectations, *Ann. Probab*. 22, 1097–1114.

GOTTLIEB, G., and A. KALAY (1985): Implications of the Discreteness of Observed Stock Prices, *J. Finance* 40, 135–153.

GROSSMAN, S., M. MILLER, K. CONE, D. FISCHEL, and D. ROSS (1997): Clustering and Competition in Asset Markets, *J. Law & Econ*. 40, 23–60.

HARRIS, L. (1990): Estimation of Stock Price Variances and Serial Covariances from Discrete Observations, *J. Financial Quant. Anal*. 25, 291–306.

HARRIS, L. (1991): Stock Price Clustering and Discreteness, *Rev. Financial Stud*. 4, 389–415.

HASBROUCK, J. (1988): Trades, Quotes, Inventories and Information, *J. Financial Econ*. 42, 229–252.

HASBROUCK, J. (1993): Assessing the Quality of a Security Market: A New Approach to Transaction-Cost Measurement, *Rev. Financial Stud*. 6, 191–212.

HASBROUCK, J. (1996): Modeling Market Microstructure Time Series; in *Handbook of Statistics*, Vol. 14, eds. G. Maddala and C. Rao. Amsterdam: North-Holland, pp. 647–692.

HASBROUCK, J. (1999): Security Bid/Ask Dynamics with Discreteness and Clustering: Simple Strategies for Modeling and Estimation, *J. Financial Markets* 2, 1–28.

HAUSMAN, J., A. LO, and C. MACKINLAY (1992): An Ordered Probit Analysis of Stock Transaction Prices, *J. Financial Econ*. 31, 319–379.

KALLIANPUR, G. (1980): *Stochastic Filtering Theory*. New York: Springer-Verlag.

KASS, R. E., and A. E. RAFTERY (1995): Bayes Factors and Model Uncertainty, *J. Am. Stat. Assoc*. 90, 773–795.

KITAGAWA, G. (1987): Non-Gaussian State-Space Modeling of Nonstationary Time Series, *J. Am. Stat. Assoc*. 82, 1032–1041.

KLIEMANN, W., G. KOCH, and F. MARCHETTI (1990): On the Unnormalized Solution of the Filtering Problem with Counting Process Observations, *IEEE Trans. Info. Theory* 36, 1415–1425.

KURTZ, T., and D. OCONE (1988): Unique Characterization of Conditional Distributions in Nonlinear Filtering, *Ann. Probab*. 16, 80–107.

KURTZ, T., and P. PROTTER (1991): Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations, *Ann. Probab*. 19, 1035–1070.

KUSHNER, H., and P. DUPUIS (1994): *Numerical Methods for Stochastic Control Problems in Continuous Time*. New York: Springer-Verlag.

MENDELSON, H. (1982): Market Behavior in a Clearing House, *Econometrica* 50, 1505–1524.

MERTON, R. (1976): Option Pricing When Underlying Stock Returns Are Discontinuous, *J. Financial Econ*. 3, 125–144.

NELSON, D. B. (1990): ARCH Models as Diffusion Approximations, *J. Econometrics* 45, 7–39.

NIEDERHOFFER, V. (1965): Clustering of Stock Prices, *Oper. Res*. 13, 258–265.

NIEDERHOFFER, V. (1966): A New Look at Clustering of Stock Prices, *J. Business* 39, 309–313.

O'HARA, M. (1995): *Market Microstructure Theory*. Oxford: Blackwell.

OSBORNE, M. (1962): Periodic Structure in the Brownian Motion of Stock Prices, *Oper. Res*. 10, 345–379.

PROTTER, P. (1992): *Stochastic Integration and Differential Equations*. New York: Springer-Verlag.

ROGERS, L. and O. ZANE (1998): Designing and Estimating Models of High-Frequency Data, Working paper, University of Bath.

RUSSELL, J., and R. ENGLE (1998): Econometric Analysis of Discrete-Valued Irregularly-Spaced Financial Transactions Data Using a New Autoregressive Conditional Multinomial Model. Discussion paper 98-10, University of California, San Diego, Department of Economics.

RYDBERG, T. H., and N. SHEPHARD (1988): Dynamics of Trade-by-Trade Price Movements: Decomposition and Models. Working paper, Nuffield College, Oxford.

SERFLING, R. J. (1980): *Approximation Theorems of Mathesmatical Statistics*. New York: Wiley.

ZENG, Y. (1999): A Class of Partially-Observed Models with Discrete, Clustering and Non-Clustering Noises: Application to Micro-movement of Stock Prices. Unpublished doctoral dissertation, The University of Wisconsin–Madison, Statistics Department.

ZHANG, M. Y., J. R. RUSSELL, and R. S. TSAY (2000): A Nonlinear Autoregressive Conditional Duration Model with Application to Financial Transaction Data. Working paper, Graduate School of Business, University of Chicago.