

Filtering with Counting Process Observations: Application to the Statistical Analysis of the Micromovement of Asset Price *

Laurie C. Scott [†] Yong Zeng [‡]

December 4, 2008

Abstract

This chapter surveys the recent developments in a general filtering model with counting process observations for the micromovement of asset price and its related statistical analysis. Mainly, the Bayesian inference (estimation and model selection) via filtering for the filtering model. The normalized and un-normalized filtering equations as well as the system of evolution equations for Bayes factors are reviewed. Markov chain approximation method is used to construct recursive algorithms and their consistency is proven. We employ a specific micromovement model built upon the model of LSDE (linear stochastic differential equation) to show the steps to develop a micromovement model with specific types of trading noises. The model is further utilized to show the steps to construct consistent recursive algorithms for computing the trade-by-trade Bayes estimates and the Bayes factors for model selection. Monte Carlo and real-data examples are provided.

KEY WORDS: filtering, counting process, Markov chain approximation, Bayesian statistical inference, price discreteness, price clustering and ultra high frequency data.

*The authors are indebted to an anonymous referee for detailed constructive written comments. The research is supported in part by the National Science Foundation under grant DMS-0604722.

[†]Department of Mathematics and Statistics, University of Missouri at Kansas City, Kansas City, MO 64110, USA; and H&R Block Company at Kansas City. Email: laurie.scott@hrblock.com

[‡]Department of Mathematics and Statistics, University of Missouri at Kansas City, Kansas City, MO 64110, USA. Tel: (816) 235 5850. Fax: (816) 235 5517. Email: zengy@umkc.edu. Website: <http://mendota.umkc.edu/>.

1 Introduction

1.1 Intuition Behind the Modeling

Existing asset price models can be broadly divided into two groups: macro- and micro-movement models. Daily, weekly, and monthly closing price behaviors are regarded as macromovement. Figure 1, the time series plot of the daily closing prices of MicroSoft during 1/1/1993 - 3/31/1994, is an example of the macromovement. There were 316 business dates and there are 316 data. Clearly, the price fluctuates and is usually modeled by either discrete-time models such as ARIMA or ARCH-type models or continuous-time models such as GBM (geometric Brownian motion), SV (stochastic volatility), or jump diffusion models in econometric or mathematical finance literature.

Micromovement refers to trade-by-trade, transactional price behavior. Such data, containing the trading times and prices for all trades, are regarded as UHF (*ultra high frequency*) data by Engle [10]. The first stylized fact of UHF data is the trading times or *durations* (that is, the inter-trading times) are random¹. Figure 2 plots all the transaction prices of MicroSoft during the same period. There are about 480,000 transaction prices, much more than the 316 data in the macromovement. First we observe that the overall shapes of these two plots are the same and this is not surprising because Figure 1 is a daily sub-sample of Figure 2. This implies that the micromovement model should be closely related to the macromovement model. Second, we observe that there are much more fluctuations in the micromovement and by just looking at these two pictures, it is very tempting to conclude that the micromovement looks even more like a GBM. However, if we cut a small piece of Figure 2 and look through it under a microscope, namely, we plot about one-half day of the prices in Figure 3, then we clearly see that the price does not move continuously in the state space as GBM suggests, but moves tick-by-tick or level-by-level. Here, a tick is the minimum price variation set by trading regulations and was \$1/8 dollar at that time. Moreover, Figure 3 clearly shows the second stylized fact of UHF data. Namely, there are trading noises in UHF prices. Trading noises include discrete noise due to the price discreteness, clustering noise due to price clustering (more prices were traded at even eighths than at odd eighths), and non-clustering noise which includes all other noise. The down spike in Figure 3 is an evidence of the existence of the non-clustering noise. Finally, the topics of the two recent presidential addresses to the American Finance Association were “noise” (Black [4]) and “friction” (Stoll [26]). Both are about market microstructure noise. This indicates that noise is an essential matter in finance and in the modeling of asset price.

By looking at these three pictures, we gain the simple intuition that a micromovement model should be built upon a macromovement model by incorporating trading noises. In the other words of economics, the trade-by-trade price is formed from the intrinsic value process of an asset by combining the market microstructure noises. Also, we can see that when we deal with macromovement or daily prices, the noises are negligible. However, when we deal with micromovement or UHF prices, the noises are not negligible anymore. One representation of the model reviewed in this paper is built upon such intuition. Another representation comes from Figure 3, where each price level is modeled by a counting process.

¹From the viewpoint of time series, econometricians naturally view such data as an *irregularly-spaced* time series and Engle [10] develops a general framework under this view with many developments.

Figure 1:

Daily Closing Prices of Microsoft, 93.01.01--94.03.31

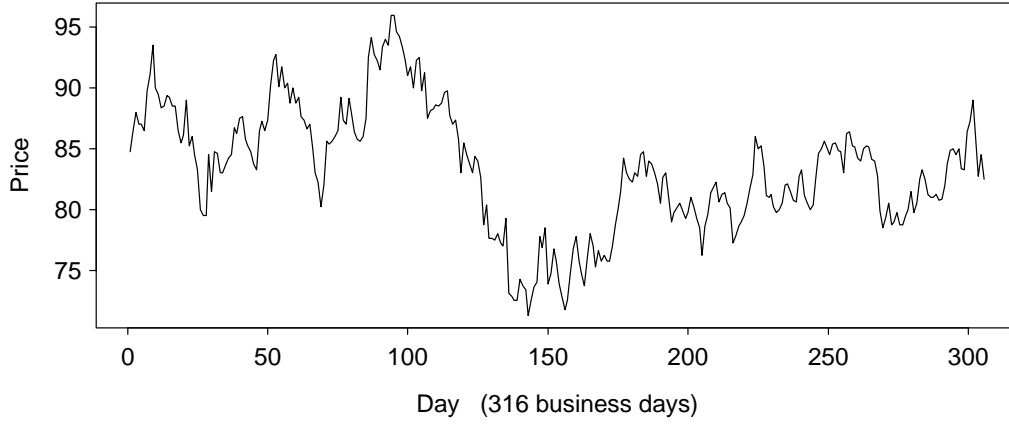


Figure 2:

Transaction Data of Microsoft, 93.01.01--94.03.31

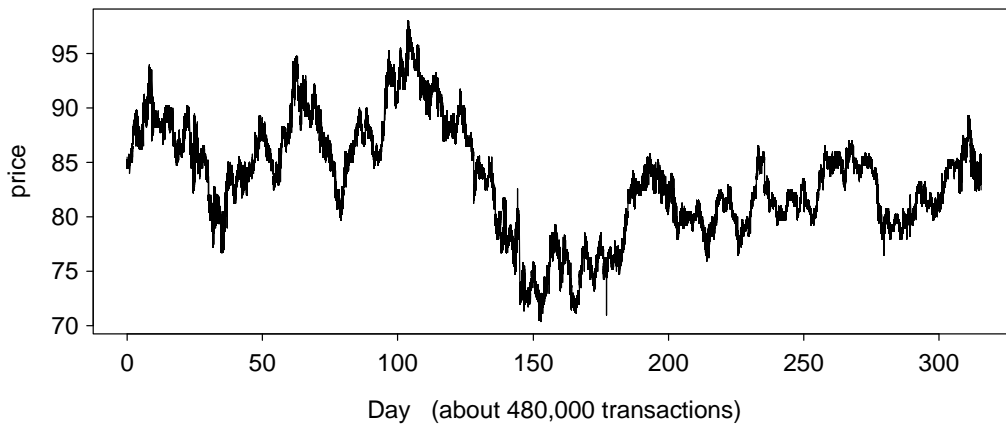
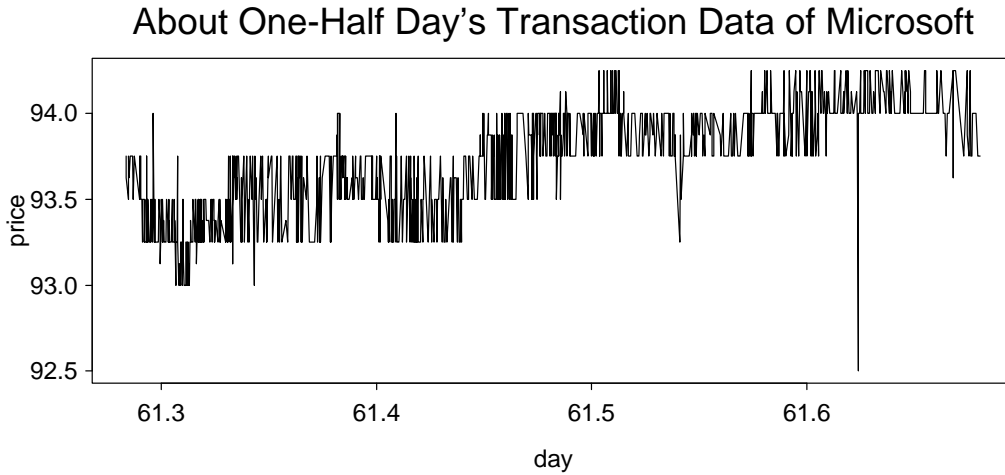


Figure 3:



1.2 Related Literature on Non-linear Filtering for UHF data

To the best of our knowledge, Frey [14] and Frey and Runggaldier [15] are the first papers that employ the non-linear filtering technique to model UHF data. Their viewpoint is to model the unobserved volatility process, which is crucial for option pricing. Their model is able to capture the Poisson random arrival times in UHF data. Cvitanic, Lipster and Rozovskii [6] extends the previous model to a more general setting and further allowing general random times of observation, not just doubly stochastic Poisson processes. Cvitanic, Rozovskii and Zaliapin [7] numerically implements the filtering procedure and estimates the unobserved volatility. However, market microstructure noise is missing in these models.

1.3 An Overview of this Chapter

Zeng [30] develops a general filtering micromovement model (FM model, as we simply call it) for asset price, where both stylized facts of micromovement are taken care of. In the FM model, there is an unobservable intrinsic value process for an asset, which corresponds to the macromovement. The intrinsic value process is the permanent component and has a long-term impact on price. Prices are observed only at random trading times which are driven by a conditional Poisson process, whose intensity may depend on the intrinsic value. Prices are distorted observations of the intrinsic value process at the trading times. Market microstructure noise is explicitly and flexibly modeled by a random transformation with a transition probability from the intrinsic value to the price at trading time. Noise is the transient component and only has a short-term impact (when a trade happens) on price. One important feature of the FM model is that the model can be framed as a filtering problem with counting process observations. This connects the model to the filtering literature, which has found great success in engineering and networking². Then, the continuous-time likelihoods and posterior not only exist, but also are uniquely characterized, respectively, by the unnormalized

²Early and recent literature on related filtering problems with counting process observations includes, but not limited to, [27], [28] [8], [25], [21], and [11].

Duncan-Mortensen-Zakai-like filtering equation and the normalized, Kushner-Stratonovich (KS) (or Fujisaki-Kallianpur-Kunita)-like filtering equations. The related numerical solution based on Markov chain approximation method and the Bayes estimation via filtering for the intrinsic value process and the related parameters in the model are developed. Furthermore, Kouritzin and Zeng [22] characterizes the continuous-time likelihood ratio and Bayes factors by a system of evolution equations and develops the Bayesian hypothesis testing or model selection via filtering for the FM model.

The first aim of this chapter is to survey the general filtering model with counting process observations for the micromovement of asset price and its related statistical analysis recently developed in [30] and [22]. The statistical analysis contains Bayesian inference (estimation and model selection) via filtering for the FM model. We adopt the Bayesian paradigm, because it offers extra model flexibility as well as the ability to incorporate real prior information. The second aim is to give a worked example to help those who wish to carry out a similar analysis in practice. A specific FM model with LSDE as the intrinsic value process is built up to show the modeling development technique. This FM model is further employed to show the steps to carry out Bayes estimation via filtering as well as to compute Bayes factors for model selection. Namely, we construct recursive algorithms for computing the trade-by-trade Bayes estimates and Bayes factors. The consistency of the recursive algorithms are established. We provide simulation results to show the consistency of Bayes estimates and the effectiveness of Bayes factors for model selection. The recursive algorithms are applied to an actual Microsoft data set to obtain trade-by-trade Bayes parameter estimates as well as to implement a simple model selection.

In Section 2, we present the general FM model in three equivalent fashions. In Section 3, we present the continuous-time Bayesian inference via filtering for the FM model including the filtering equations and the consistency theorem. In Section 4, we focus on developing a specific FM model built on LSDE and developing the related recursive algorithms. Simulation and real-data examples are provided for Bayes estimation and Bayesian model selection. Section 5 concludes.

2 The General FM Model

The general FM model has three equivalent representations.

2.1 Representation I: Constructing Price from Intrinsic Value

Based on the simple intuition obtained in Section 1.1 that the price is formed from an intrinsic value by incorporating the noises that arise from the trading activity, we build up the FM model.

In general, there are three steps in constructing the price process Z from the intrinsic value process X . First, we specify X . In order to permit time-dependent parameters such as stochastic volatility and to prepare for parameter estimation, we enlarge the partially-observed model (X, Y) to (θ, X, Y) . Assume (θ, X, Y) is defined in a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $\{\mathcal{F}_t\}_{0 \leq t \leq \infty}$. Since the macro-movement models are appropriate for the value processes, we invoke below a mild assumption on (θ, X) so that all relevant stochastic processes are included.

Assumption 1 (θ, X) is the unique solution of a martingale problem for a generator \mathbf{A} such that for a function f in the domain of \mathbf{A} , $M_f(t) = f(\theta(t), X(t)) - \int_0^t \mathbf{A}f(\theta(s), X(s))ds$, is a $\mathcal{F}_t^{\theta, X}$ -martingale, where $\mathcal{F}_t^{\theta, X}$ is the σ -algebra generated by $(\theta(s), X(s))_{0 \leq s \leq t}$.

The generator and martingale problem approach (see for example, Ethier and Kurtz [12]) furnishes a powerful tool for the characterization of Markov processes. Assumption 1 includes all relevant stochastic processes such as diffusion and jump-diffusion processes for modeling asset price. One examples is given in Section 4.

However, in UHF data, the price can not be observed continuously in time, neither can it move continuously as GBM suggests. Therefore, two more steps are necessary. Step 2 takes care of the trading times and Step 3 the trading noise.

In Step 2, we assume trading times $t_1, t_2, \dots, t_i, \dots$, are driven by a conditional Poisson process with an intensity $a(X(t), \theta(t), t)$. In Step 3, $Y(t_i)$, the price at time t_i , is corrupted from $X(t_i)$, the intrinsic value, with trading noise. Namely, $Y(t_i) = F(X(t_i))$, where $y = F(x)$ is a random transformation with the transition probability $p(y|x)$, modeling the trading noise. The random transformation, $F(x)$, is flexible and Section 4.1 constructs one to accommodate the three observed types of noise in the MicroSoft data: discrete, clustering and non-clustering.

Under this construction, information affects $X(t)$, the value of an asset, and has a permanent influence on the price while noise modeled by $F(x)$ (or $p(y|x)$) only has a transitory impact on price. The formulation is similar to the time series VAR structural models used in many market microstructure papers (see a survey paper [17] by Hasbrouck and a recent paper [18]). Furthermore, the formulation is closely related the recent two-scale frameworks incorporating market microstructure noises in literature of realized volatility estimators. See [34], [1], [2], and [13]. Especially, Li and Mykland in [24] shows that rounding noise in UHF data may severely distort even the two-scale estimators of realized volatility, and the error could be infinite.

2.2 Representation II: Filtering with Counting Process Observations

From Figure 3 and because of price discreteness, we can formulate the prices of an asset as a collection of counting processes in the following form:

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

where $Y_j(t) = N_j(\int_0^t \lambda_j(\theta(s), X(s), s)ds)$ is the counting process recording the cumulative number of trades that have occurred at the j th price level (denoted by y_j) up to time t . We make four more mild assumptions on the model.

Assumption 2 $\{N_j\}_{j=1}^n$ are unit Poisson processes under measure \mathbb{P} .

Then, $Y_j(t) = N_j(\int_0^t \lambda_j(\theta(s), X(s), s)ds)$ is conditional Poisson process with the stochastic intensity, $\lambda_j(\theta(t), X(t), t)$. Given $\mathcal{F}_t^{\theta, X}$, the filtration of θ and X , $Y_j(t)$ has a Poisson dis-

tribution with parameter $\int_0^t \lambda_j(\theta(s), X(s), s) ds$. Moreover, $Y_j(t) - \int_0^t \lambda_j(\theta(s), X(s), s) ds$ is a $\mathcal{F}_t^{\theta, X, Y}$ -martingale.

Assumption 3 $(\theta, X), N_1, N_2, \dots, N_n$ are independent under measure \mathbb{P} .

Let $a(\theta, X(t), t)$ be the total trading intensity at time t .

Assumption 4 There exists a positive constant, C , such that $0 \leq a(\theta, x, t) \leq C$ for all $t > 0$ and (θ, x) .

The total trading intensity $a(\theta, x, t)$, which is bounded by the above assumption, determines the expected rate of trading at time t . These three assumptions imply that there exists a reference measure Q and that after a suitable change of measure to Q , $(\theta, X), Y_1, \dots, Y_n$ become independent, and Y_1, Y_2, \dots, Y_n become unit Poisson processes (Bremaud [5]).

Assumption 5 The intensities are of the form: $\lambda_j(\theta, x, t) = a(\theta, x, t)p(y_j|x)$, where $p(y_j|x)$ is the transition probability from x to y_j , the j th price level.

This assumption imposes a desirable structure for the intensities of the model. It means that the total trading intensity $a(\theta(t), X(t), t)$ determines when the next trade will occur and $p(y_j|X(t))$ determines at which price level the next trade will occur given the value is $X(t)$. Note that $p(y_j|X(t))$ models how the trading noise enters the price process.

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

2.3 An Integral Form of the Price

To solve problems in mathematical finance such as the option pricing and hedging and the portfolio selection, the stochastic differential or integral equation form of the most recent price is needed. However, the previous two representations do not provide such a form of price. That is why a third representation is given below. There are also three steps in constructing such representation.

Step 1 : We define a random counting measure. Let $U = \{0, \frac{1}{M}, \frac{2}{M}, \dots\}$ be a mark space containing all the possible price levels, and $u = \frac{j}{M}$ be a generic point in U . Note that M can be 8, 16, 64, 100, 128 or 256 or others according to the asset. For stock price, the tick size was 1/8, 1/16 and it is 1/100 of one dollar. For treasure note, the tick sizes are 1/64, 1/128 or 1/256 of one percentage.

For $A \in U$, we define $m(A, t)$ as a random counting measure that counts the cumulative number of trades whose price levels are in A up to time t . When $A = \{\frac{j}{M}\}$, $m(\{\frac{j}{M}\}, t) = Y_j(t)$, which counts the number of trades occurring at the j price level. More generally, we can express $m(A, t) = \sum_{u \in A} Y_{uM}(t)$.

A random measure is characterized by its compensator. To express the compensator of $m(A, t)$, we define a counting measure $\eta(A) = \text{number of element in } A$ for $A \in U$. Note that η has the following two properties: For $A \in U$, $\eta(A) = \int I_A \eta(du)$ and $\int_A f(u) \eta(du) = \sum_{u \in A} f(u)$.

Step 2 : We write $\gamma_m(A, t)$, the compensator of $m(A, t)$ with respect to \mathcal{F}_t^X , as

$$\begin{aligned}\gamma_m(A, t) &= \int_0^t \int_A p(u|X(s))a(\theta, X(s), s)\eta(du)ds \\ &= \sum_{u \in A} \int_0^t p(u|X(s))a(\theta, X(s), s)ds.\end{aligned}$$

Again, when $A = \{\frac{j}{M}\}$,

$$\gamma_m(\{\frac{j}{M}\}, t) = \int_0^t \lambda_j(\theta, X(s), s)ds = \int_0^t p(\frac{j}{M}|X(s))a(\theta, X(s), s)ds.$$

Moreover, we can write

$$\gamma_m(du, dt) = p(u|X_t)a(\theta, X_t, t)\eta(du)dt.$$

Step 3 : We are in the position to define the integral form needed. Let $Y(t)$ be the price of the most recent transaction at or before time t . Then,

$$Y(t) = Y(0) + \int_{[0, t] \times U} (u - Y(s-))m(du, ds). \quad (2)$$

Note that $m(du, ds)$ is zero most of time, and becomes one only at trading time t_i with $u = Y(t_i)$, the trading price. The above expression is but a telescoping sum: $Y(t) = Y(0) + \sum_{t_i < t} (Y(t_i) - Y(t_{i-1}))$. Alternatively, in differential form,

$$dY(t) = \int_U (u - Y(t-))m(du, dt). \quad (3)$$

To understand the above differential equation, assuming there is a price change from $Y(t-)$ to u occurs at time t , then $Y(t) - Y(t-) = (u - Y(t-))$ implying $Y(t) = u$.

With the above representation of price, Lee and Zeng [23] studies the option pricing and hedging through local risk minimizing criterion, and Xiong and Zeng [29] studies the portfolio selection problem.

2.4 The Equivalence of the Three Representation

In Representation I, the price is constructed from the intrinsic value. In Representation II, the price process is a collection of counting processes. In Representation III, the SDE form of the price is given. The following proposition states their equivalence in distribution. This guarantees the statistical inference based on the second representation is also valid to the other two.

Proposition 1 *The three representations of the model in Sections 2.1, 2.2 and 2.3, respectively, have the same probability law.*

The proof of the equivalence of Representations of I and II can be found in [32]. The intensity structure of Assumption 5 plays an essential role in the equivalence of the two approaches of modeling. The main idea is to show both marked point processes have the same stochastic intensity kernel. Similarly, we can show the third one is equivalent with the first two.

3 Bayesian Inference via Filtering

This section summarizes the theoretical results of the Bayesian inference via filtering. We present the statistical foundations for the general FM model, the related filtering equations, and a convergence theorem. The convergence theorem not only provides a blueprint through the Markov chain approximation method to construct recursive algorithms, but also ensures the consistency of such algorithms, which compute the approximate continuous-time likelihoods, the posterior, and the Bayes factors.

3.1 The Statistical Foundations

In this section, we study the continuous-time joint likelihood, the likelihood function (from frequentists' viewpoint), the integrated likelihood (from Bayesians' viewpoint), the posterior of the proposed model as well as the continuous-time likelihood ratio and Bayes factors for model selections. These terminologies are used in statistics. To connect to the terminologies used in the filtering community, the continuous-time joint likelihood corresponds to the Girsanov-type exponential martingale, the likelihood to the total un-normalized conditional measure, the posterior to the conditional distribution (or the normalized conditional measure), and the Bayes factors to the total conditional ratio measures. These conditional measures are characterized by the unnormalized and normalized filtering equations as well as the system of evolution equations, respectively.

3.1.1 The Continuous-time Joint Likelihood

The probability measure \mathbb{P} of (θ, X, \vec{Y}) can be written as $\mathbb{P} = \mathbb{P}_{\theta, x} \times \mathbb{P}_{y|\theta, x}$, where $\mathbb{P}_{\theta, x}$ is the probability measure for (θ, X) such that $M_f(t)$ in Assumption 1 is a $\mathcal{F}_t^{\theta, X}$ -martingale, and $\mathbb{P}_{y|\theta, x}$ is the conditional probability measure on $D_{R^n}[0, \infty)$ for \vec{Y} given (θ, X) (where $D_{R^n}[0, \infty)$ is the space of right continuous with left limit functions). Under \mathbb{P} , \vec{Y} relies on (θ, X) . Recall that there exists a reference measure Q such that under Q , $(\theta, X), \vec{Y}$ become independent, (θ, X) remains the same probability law and Y_1, Y_2, \dots, Y_n become unit Poisson processes. Therefore, Q can be decomposed as $Q = \mathbb{P}_{\theta, x} \times Q_y$, where Q_y is the probability measure for n independent unit Poisson processes. One can obtain the Radon-Nikodym derivative of the model, that is the joint likelihood of (θ, X, \vec{Y}) , $L(t)$, as (see [5] pg 166),

$$\begin{aligned} L(t) &= \frac{d\mathbb{P}}{dQ}(t) = \frac{d(\mathbb{P}_{\theta, x} \times \mathbb{P}_{y|\theta, x})}{d(\mathbb{P}_{\theta, x} \times dQ_y)}(t) = \frac{d\mathbb{P}_{y|\theta, x}}{dQ_y}(t) \\ &= \prod_{j=1}^n \exp \left\{ \int_0^t \log \lambda_j(\theta(s-), X(s-), s-) dY_j(s) - \int_0^t [\lambda_j(\theta(s), X(s), s) - 1] ds \right\}. \end{aligned} \quad (4)$$

or in SDE form:

$$L(t) = 1 + \sum_{j=1}^n \int_0^t [\lambda_j(\theta(s-), X(s-), s-) - 1] L(s-) d(Y_j(s) - s).$$

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

It is clear that we can not observe X (and the stochastic components of θ such as stochastic volatility if exists in the intrinsic value process) and the joint likelihood is not computable. To do statistical analysis, what we need is the likelihood of \vec{Y} alone. Therefore, we would like to integrate out X . In the expression of probability, this can be done by using conditional expectation. Let $\mathcal{Y}_t = \sigma\{\vec{Y}(s) | 0 \leq s \leq t\}$ be all the available information up to time t . We use $E^Q[X]$ and $\mathbb{E}[X]$ to indicate that the expectation is taken with respect to the measures Q and \mathbb{P} , respectively.

Definition 1 Let ρ_t be the conditional measure of $(\theta(t), X(t))$ given \mathcal{Y}_t defined as

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

Definition 2 Let

$$\rho(f, t) = E^Q[f(\theta(t), X(t))L(t) | \mathcal{Y}_t] = \int f(\theta, x)\rho_t(d\theta, dx).$$

If $(\theta(0), X(0))$ is given, then the likelihood of Y is $E^Q[L(t) | \mathcal{Y}_t] = \rho(1, t)$, the total conditional measure. In a Bayesian framework, a prior is placed on $(\theta(0), X(0))$, and the integrated (or marginal) likelihood of Y is also $\rho(1, t)$.

3.1.3 The Continuous-time Posterior

Definition 3 Let π_t be the conditional distribution of $(\theta(t), X(t))$ given \mathcal{Y}_t and let

$$\pi(f, t) = \mathbb{E}[f(\theta(t), X(t)) | \mathcal{Y}_t] = \int f(\theta, x)\pi_t(d\theta, dx).$$

Again, in a Bayesian framework, a prior is placed on $(\theta(0), X(0))$, and π_t becomes the continuous-time posterior, which is determined by $\pi(f, t)$ for all continuous and bounded f . Bayes Theorem (see [5], page 171) provides the relationship between $\rho(f, t)$ and $\pi(f, t)$: $\pi(f, t) = \rho(f, t) / \rho(1, t)$. That is, $\pi(f, t)$ is the normalized conditional measure. Hence, the equation governing the evolution of $\rho(f, t)$ is called the *un-normalized filtering equation*, and that of $\pi(f, t)$ is called the *normalized filtering equation*.

3.1.4 Continuous-time Bayes Factors

Given one data set, there are many different models³ to fit the data set. Bayes factor is a model selection criterion, first developed by Jeffreys [19] in a Bayesian framework. Suppose there are two models. Bayes factor quantifies the evidence provided by the data in favor of Model 1 over Model 2. For $c = 1, 2$, denote Model c by $(\theta^{(c)}, X^{(c)}, \vec{Y}^{(c)})$. Denote the joint likelihood of Model c by $L^{(c)}(t)$, as in Equation (4). Denote the un-normalized conditional measure of Model c as $\rho_c(f_c, t) = E^{Q^{(c)}}[f_c(\theta^{(c)}(t), X^{(c)}(t))L^{(c)}(t) | \mathcal{Y}_t^{(c)}]$. Then, the integrated likelihood of \vec{Y} is $\rho_c(1, t)$, for Model c .

³Here means different models, not the same model with different representations as described in Section 2.

Table 1: Interpretation of Bayes Factor

B_{21}	Evidence against Model 1
1 to 3	Not worth more than a bare mention
3 to 20	Positive
20 to 150	Strong
> 150	Decisive

Jeffreys [19] defined the Bayes factor of Model 2 over Model 1, B_{21} , as the ratio of integrated likelihoods of Model 2 over Model 1. In our setting, that is, $B_{21}(t) = \rho_2(1, t)/\rho_1(1, t)$. Bayes Factors is designed to measure the relative fit of one model vs. another one given the observed, or in our case partially-observed, data. To select a ‘best’ model among a set of models, Bayes factor can achieve this goal via pair-wise comparison. When B_{21} has been calculated, it can be interpreted using Table 1 furnished by Kass and Raftery [20] as guideline. Similarly, we can define B_{12} , the Bayes factor of Model 1 over Model 2.

In order to characterize the evolution of the Bayes factors, we would like to define two more conditional measures. Instead of normalizing by its integrated likelihood to obtain an normalized conditional measure, we let the conditional measure ρ_c be divided by the integrated likelihood of the other model. In this way, we define two conditional ratio measures as below.

Definition 4 For $c = 1, 2$, let $q_t^{(c)}$ be the conditional ratio measure of $(\theta^{(c)}(t), X^{(c)}(t))$ given $\mathcal{Y}_t^{(c)}$:

$$\begin{aligned} & q_t^{(c)} \left\{ (\theta^{(c)}(t), X^{(c)}(t)) \in A \right\} \\ &= \frac{E^Q \left[\mathbf{I}_{\{(\theta^{(c)}(t), X^{(c)}(t)) \in A\}} (\theta^{(c)}(t), X^{(c)}(t)) L^{(c)}(t) | \mathcal{Y}_t^{(c)} \right]}{\rho_{3-c}(1, t)} \\ &= E^Q \left[\mathbf{I}_{\{(\theta^{(c)}(t), X^{(c)}(t)) \in A\}} (\theta^{(c)}(t), X^{(c)}(t)) \tilde{L}^{(c)}(t) | \mathcal{Y}_t^{(c)} \right] \end{aligned}$$

where $\tilde{L}^{(c)}(t) = L^{(c)}(t)/\rho_{3-c}(1, t)$.

The second equality is because $\rho_c(1, t)$, $c = 1, 2$, depend on the same filtration $\mathcal{F}^{Y^{(1)}} = \mathcal{F}^{Y^{(2)}}$ and $\rho_{3-c}(1, t)$ can be moved inside the conditioning.

Definition 5 Let the conditional ratio processes for f_1 and f_2 be:

$$q_1(f_1, t) = \frac{\rho_1(f_1, t)}{\rho_2(1, t)}, \quad \text{and} \quad q_2(f_2, t) = \frac{\rho_2(f_2, t)}{\rho_1(1, t)}.$$

The reason that we define the two conditional ratio measure is given by the observation that the Bayes factors can be expressed by $B_{12}(t) = q_1(1, t)$ and $B_{21}(t) = q_2(1, t)$. Moreover, observe that $q_c(f_c, t)$ can be written as $q_c(f_c, t) = \int f_c(\theta^{(c)}, x^{(c)}) q_t^{(c)}(d\theta^{(c)}, dx^{(c)})$. The integral forms of $\rho(f, t)$, $\pi(f, t)$, and $q_c(f_c, t)$ are important in deriving the recursive algorithms where f (or f_c) is taken to be a lattice-point indicator function.

Using Bayes factors for model selection has at least two advantages over the likelihood-based approaches. First, unlike likelihood-based model selection approaches, a Bayes factor neither requires the models to be nested, nor does it require the probability measures of the models to be absolutely continuous with respect to those of the total models in hypotheses. This feature is important for the model selection of stochastic process, since absolute continuity is not as common in the probability measures of continuous-time stochastic processes as those of discrete-time ones. Second, Kass and Raftery [20] show that under some conditions, Bayes factor \approx BIC (Bayesian Information Criterion), which penalizes according to both the number of parameters and the number of data. This suggests that Bayes factor might have this desirable property also.

3.2 Filtering and Evolution Equations

The stochastic partial differential equations (SPDEs) provide a powerful machinery to characterize the infinite dimensional conditional measures. Similar filtering or related equations for the related filtering problems with counting process observations can be found, for example, in [27], [28], [8], [25] [21], and [11]. The conditional measures, in turn, determine the continuous-time likelihoods, posteriors, likelihood ratios and Bayes factors. The following two theorems summarize all the useful SPDEs.

Theorem 1 *Suppose that (θ, X, \vec{Y}) satisfies Assumptions 1 - 5. Then, ρ_t is the unique measure-valued solution of the SPDE, the unnormalized filtering equation,*

$$\rho(f, t) = \rho(f, 0) + \int_0^t \rho(\mathbf{A}f, s) ds + \sum_{j=1}^n \int_0^t \rho((ap_j - 1)f, s-) d(Y_j(s) - s), \quad (5)$$

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

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

$$\begin{aligned} \pi(f, t) &= \pi(f, 0) + \int_0^t \pi(\mathbf{A}f, s) ds \\ &+ \sum_{j=1}^n \int_0^t \left[\frac{\pi(f p_j, s-)}{\pi(p_j, s-)} - \pi(f, s-) \right] d(Y_j(s) - \pi(ap_j, s)s). \end{aligned} \quad (6)$$

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

$$\pi(f, t) = \pi(f, 0) + \int_0^t \pi(\mathbf{A}f, s) ds + \sum_{j=1}^n \int_0^t \left[\frac{\pi(f p_j, s-)}{\pi(p_j, s-)} - \pi(f, s-) \right] dY_j(s). \quad (7)$$

In Theorem 1, the unnormalized filtering equation characterizes the evolution of the conditional measure, and its total conditional measure, which is the likelihoods. The normalized filtering equation characterizes the evolution of the normalized conditional measure, which is the posteriors in a Bayesian framework.

Theorem 2 Suppose Model c ($c = 1, 2$) has generator $\mathbf{A}^{(c)}$ for $(\theta^{(c)}, X^{(c)})$, the trading intensity $a_c = a_c(\theta^{(c)}(t), X^{(c)}(t), \vec{Y}^{(c)}(t))$, and the transition probability $p_j^{(c)} = p^{(c)}(y_j|x)$ from x to y_j for the random transformation $F^{(c)}$. Suppose that $(\theta^{(c)}, X^{(c)}, \vec{Y}^{(c)})$ satisfies Assumptions 1 - 5. Then, $(q_t^{(1)}, q_t^{(2)})$ are the unique measure-valued pair solution of the following system of SPDEs,

$$\begin{aligned} q_1(f_1, t) &= q_1(f_1, 0) + \int_0^t q_1(\mathbf{A}^{(1)} f_1, s) ds \\ &+ \sum_{j=1}^n \int_0^t \left[\frac{q_1(f_1 a_1 p_j^{(1)}, s-)}{q_2(a_2 p_j^{(2)}, s-)} q_2(1, s-) - q_1(f_1, s-) \right] d(Y_j(s) - \frac{q_2(a_2 p_j^{(2)}, s)}{q_2(1, s)} ds) \end{aligned} \quad (8)$$

$$\begin{aligned} q_2(f_2, t) &= q_2(f_2, 0) + \int_0^t q_2(\mathbf{A}^{(2)} f_2, s) ds \\ &+ \sum_{j=1}^n \int_0^t \left[\frac{q_2(f_2 a_2 p_j^{(2)}, s-)}{q_1(a_1 p_j^{(1)}, s-)} q_1(1, s-) - q_2(f_2, s-) \right] d(Y_j(s) - \frac{q_1(a_1 p_j^{(1)}, s)}{q_1(1, s)} ds) \end{aligned} \quad (9)$$

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

$$\begin{aligned} q_1(f_1, t) &= q_1(f_1, 0) + \int_0^t q_1(\mathbf{A}^{(1)} f_1, s) ds \\ &+ \sum_{j=1}^n \int_0^t \left[\frac{q_1(f_1 p_j^{(1)}, s-)}{q_2(p_j^{(2)}, s-)} q_2(1, s-) - q_1(f_1, s-) \right] dY_j(s), \end{aligned} \quad (10)$$

$$\begin{aligned} q_2(f_2, t) &= q_2(f_2, 0) + \int_0^t q_2(\mathbf{A}^{(2)} f_2, s) ds \\ &+ \sum_{j=1}^n \int_0^t \left[\frac{q_2(f_2 p_j^{(2)}, s-)}{q_1(p_j^{(1)}, s-)} q_1(1, s-) - q_2(f_2, s-) \right] dY_j(s). \end{aligned} \quad (11)$$

In Theorem 2, the system of evolution equations for $q_c(f_c, t)$, $c = 1, 2$, characterizes the evolution of the conditional ratio measures and their total measures, namely, the likelihood ratios or the Bayes factors.

The proof of Theorem 1 is in [30] and that of Theorem 2 is [22]. Note that all the unsimplified filtering equations are rewritten in the semimartingale form.

Note that $a(t)$ disappears in Equations (7), (10) and (11). This reduces the computation greatly in computing the Bayes estimates and Bayes factors. The tradeoff of taking a_i independent of $(\theta^{(c)}, X^{(c)})$ is that the relationship between trading intensity and other parameters (such as stochastic volatility) is excluded.

Let the trading times be t_1, t_2, \dots , then, for example, Equation (7) can be written in two parts. The first is called the *propagation equation*, describing the evolution without trades and the second is called the *updating equation*, describing the update when a trade occurs. The propagation equation has no random component and is written as

$$\pi(f, t_{i+1}-) = \pi(f, t_i) + \int_{t_i}^{t_{i+1}-} \pi(\mathbf{A}f, s) ds. \quad (12)$$

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

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

$$\pi(f, t_{i+1}) = \frac{\pi(fp_j, t_{i+1}-)}{\pi(p_j, t_{i+1}-)}. \quad (13)$$

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

3.3 A Convergence Theorem and Recursive Algorithms

Theorems 1 and 2 provide the evolutions of the continuous-time versions, which are all infinite dimensional. To compute them, one needs to reduce the infinite dimensional problem to a finite dimensional problem and constructs algorithms based on it. The algorithms, based on the evolutions of SPDEs, are naturally recursive, handling a datum at a time. Moreover, the algorithms are easily parallelizable. Thus, the algorithm can make real-time updates and handle large data sets. One basic requirement for the recursive algorithms is consistency: The approximate versions, computed by the recursive algorithms, converges to the true ones. The following theorem proves the consistency of the approximate versions and provides a blueprint for constructing consistent algorithms through Kushner's Markov chain approximation methods.

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

$$\vec{Y}_\epsilon^{(c)}(t) = \begin{pmatrix} N_1^{(c)}(\int_0^t \lambda_1(\theta_\epsilon^{(c)}(s), X_\epsilon^{(c)}(s), s) ds) \\ N_2^{(c)}(\int_0^t \lambda_2(\theta_\epsilon^{(c)}(s), X_\epsilon^{(c)}(s), s) ds) \\ \vdots \\ N_{n_c}^{(c)}(\int_0^t \lambda_{n_c}(\theta_\epsilon^{(c)}(s), X_\epsilon^{(c)}(s), s) ds) \end{pmatrix}, \quad (14)$$

set $\mathcal{F}_t^{\vec{Y}_\epsilon^{(c)}} = \sigma(\vec{Y}_\epsilon^{(c)}(s), 0 \leq s \leq t)$, take $L_\epsilon^{(c)}(t) = L((\theta_\epsilon^{(c)}(s), X_\epsilon^{(c)}(s), Y_\epsilon^{(c)}(s))_{0 \leq s \leq t})$ as in Equation (4). We use the notation, $X_\epsilon \Rightarrow X$, to mean X_ϵ converges weakly to X in the Skorohod topology as $\epsilon \rightarrow 0$. Suppose that $(\theta_\epsilon^{(c)}, X_\epsilon^{(c)}, \vec{Y}_\epsilon^{(c)})$ lives on $(\Omega_\epsilon^{(c)}, \mathcal{F}_\epsilon^{(c)}, P_\epsilon^{(c)})$, and Assumptions 1 - 5 also hold for $(\theta_\epsilon^{(c)}, X_\epsilon^{(c)}, \vec{Y}_\epsilon^{(c)})$. Then, there also exists a reference measure $Q_\epsilon^{(c)}$ with similar properties. Next, we define the approximations of $\rho_c(f_c, t)$, $\pi_c(f_c, t)$, and $q_c(f_c, t)$.

Definition 6 For $c = 1, 2$, let

$$\begin{aligned} \rho_{\epsilon,c}(f_c, t) &= E^{Q_\epsilon^{(c)}} \left[f_c(\theta_\epsilon^{(c)}(t), X_{\epsilon_x}^{(c)}(t)) L_\epsilon^{(c)}(t) | \mathcal{F}_t^{\vec{Y}_\epsilon^{(c)}} \right], \\ \pi_{\epsilon,c}(f_c, t) &= E^{P_\epsilon^{(c)}} \left[f_c(\theta_\epsilon^{(c)}(t), X_{\epsilon_x}^{(c)}(t)) | \mathcal{F}_t^{\vec{Y}_\epsilon^{(c)}} \right], \\ q_{\epsilon,1}(f_1, t) &= \rho_{\epsilon,1}(f_1, t) / \rho_{\epsilon,2}(1, t) \text{ and } q_{\epsilon,2}(f_2, t) = \rho_{\epsilon,2}(f_2, t) / \rho_{\epsilon,1}(1, t). \end{aligned}$$

Theorem 3 Suppose that Assumptions 1 - 5 hold for the models $(\theta^{(c)}, X^{(c)}, \vec{Y}^{(c)})_{c=1,2}$ and that Assumptions 1 - 5 hold for the approximate models $(\theta_\epsilon^{(c)}, X_\epsilon^{(c)}, \vec{Y}_\epsilon^{(c)})$. Suppose $(\theta_\epsilon^{(c)}, X_\epsilon^{(c)}) \Rightarrow (\theta^{(c)}, X^{(c)})$ as $\epsilon \rightarrow 0$.

Then, as $\epsilon \rightarrow 0$, for all bounded continuous functions, f_1 and f_2 , and $c = 1, 2$,

(i) $\vec{Y}_\epsilon^{(c)} \Rightarrow \vec{Y}^{(c)}$; (ii) $\rho_{\epsilon,c}(f_c, t) \Rightarrow \rho_c(f_c, t)$; (iii) $\pi_{\epsilon,c}(f_c, t) \Rightarrow \pi_c(f_c, t)$; (iv) $q_{\epsilon,1}(f_1, t) \Rightarrow q_1(f_1, t)$ and $q_{\epsilon,2}(f_2, t) \Rightarrow q_2(f_2, t)$ simultaneously.

The proofs for (i) and (iii) are in [30] and those for (ii) and (iv) are in [22].

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

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

$$\pi_\epsilon(f, t_{i+1}-) = \pi_\epsilon(f, t_i) + \int_{t_i}^{t_{i+1}-} \pi_\epsilon(\mathbf{A}_\epsilon f, s) ds, \quad (15)$$

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

$$\pi_\epsilon(f, t_{i+1}) = \frac{\pi_\epsilon(f p_{\epsilon,j}, t_{i+1}-)}{\pi_\epsilon(p_{\epsilon,j}, t_{i+1}-)}. \quad (16)$$

Step 3 converts Equations (15) and (16) to the recursive algorithm in discrete state space and in discrete times by two substeps: (a) represents $\pi_\epsilon(\cdot, t)$ as a finite array with the components being $\pi_\epsilon(f, t)$ for lattice-point indicator f and (b) approximates the time integral in (15) with an Euler scheme.

4 A LSDE FM Model with Simulation and Empirical Results

We exemplify first how to build a specific FM model and then how to construct the recursive algorithms for computing the joint posteriors and the Bayes estimates as well as for computing the Bayes factors. Simulation and real-data examples are provided in the end of this section.

4.1 The LSDE FM Model

Since it is intuitive to construct the model, we do so. Step 1: we specify the intrinsic value process and its infinitesimal generator. Suppose $X(t)$ follows a LSDE with the SDE given by

$$dX(t) = (a + bX(t))dt + (c + dX(t))dB(t). \quad (17)$$

where $B(t)$ is a standard Brownian Motion, and $a, b, c,$ and d are constants. Conceptually, this model considers both the stock's instantaneous rate of return and volatility as linear

functions of the price instead of constants. LSDE contains GBM, LBM and O-U processes. Its generator is:

$$Af(x) = (a + bx)\frac{\partial f}{\partial x} + \frac{1}{2}(c + dx)^2\frac{\partial^2 f}{\partial x^2} \quad (18)$$

Step 2: We simply assume the trading times follow a Poisson process with a deterministic trading intensity, $a(t)$. Then, the simplified normalized filtering equation for posterior and the evolution equations for Bayes factors can be employed. A time-dependent deterministic intensity $a(t)$ fits the trade duration data better than the time-invariant one since trading activities are higher in the opening and the closing periods.

Step 3: We incorporate the trading noises on the intrinsic values at trading times to produce the price process. There are three important types of noise that have been identified as we have shown in Figure 3 and have been extensively studied in the finance literature (for example, see Harris [16]): discrete, clustering, and nonclustering. First, intraday prices move discretely, resulting in “discrete noise”. Second, because prices do not happen evenly on all ticks, but more concentrate on integer and half ticks, “price clustering” is obtained. Third, the “non-clustering noise” contains all other unspecified noise. Let $R[\cdot, \frac{1}{M}]$ be the rounding function to the closest $\frac{1}{M}$. For simple notation, at a trading time t_i , let $x = X(t_i)$, $y = Y(t_i)$, and $y' = Y'(t_i) = R[X(t_i) + V_i, \frac{1}{M}]$, where V_i is to be defined as the non-clustering noise and Instead of directly formulating $p(y|x)$, we construct $y = F(x)$ in three steps:

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

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

Step (ii): Incorporate discrete noise by rounding off x' to its closest tick, $y' = R[x', \frac{1}{M}]$.

Step (iii): Incorporate clustering noise by biasing y' through a random biasing function $b_i(\cdot)$ at trade i . $\{b_i(\cdot)\}$ is assumed independent of $\{y'_i\}$. To be consistent with the Microsoft data analyzed, we construct a simple random biasing function only for the tick of 1/8 dollar (i.e. $M = 8$). For other tick size, it can be done similarly. The data to be fitted has this clustering occurrence: integers and halves are most likely and have about the same frequencies; odd quarters are the second most likely and have about the same frequencies; and odd eighths are least likely and have about the same frequencies. To generate such clustering, a random biasing function is constructed based on the following rules: if the fractional part of y' is an even eighth, then y stays on y' with probability one; if the fractional part of y' is an odd eighth, then y stays on y' with probability $1 - \alpha - \beta$, y moves to the closest odd quarter with probability α , and moves to the closest half or integer with probability β . In brief,

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

The detail of $b_i(\cdot)$, and the explicit $p(y|x)$ for F can be found in Appendix A. Simulations can demonstrate that the constructed $F(x)$ are able to capture the tick-level sample characteristics of transaction data. For more details see [31].

The parameters of clustering noise, α and β , can be estimated through the method of relative frequency. The other parameters, a, b, c, d and ρ , are estimated by Bayes estimation via filtering through the recursive algorithm to be constructed.

4.2 The Recursive Algorithms for Bayes Estimates and Bayes Factors

We exemplify the three-step blueprint of Markov chain approximation method summarized in the end of Section 3 to construct the recursive algorithms. Finally, we show the consistency of the algorithms. For notational simplicity, we use the superscript (c) , for $c = 1, 2$, to distinguish the two models and let $\bar{\theta}_\epsilon^{(c)} = (a_{\epsilon_a}^{(c)}, b_{\epsilon_b}^{(c)}, c_{\epsilon_c}^{(c)}, d_{\epsilon_d}^{(c)}, \rho_{\epsilon_\rho}^{(c)})$ to denote an approximate discretized parameter signal, which is random in the Bayesian framework.

4.2.1 Step 1: Construct $(\bar{\theta}_\epsilon^{(c)}, X_{\epsilon_x}^{(c)})$

First, we latticize the parameter spaces of $a^{(c)}, b^{(c)}, c^{(c)}, d^{(c)}, \rho^{(c)}$ and the state space of $X^{(c)}$. Suppose there are $n_a^{(c)} + 1, n_b^{(c)} + 1, n_c^{(c)} + 1, n_d^{(c)} + 1, n_\rho^{(c)} + 1$ and $n_x^{(c)} + 1$ lattices in the latticized spaces of $a^{(c)}, b^{(c)}, c^{(c)}, d^{(c)}, \rho^{(c)}$ and $X^{(c)}$ respectively, e.g.

$$a^{(c)} : [\alpha_a^{(c)}, \beta_a^{(c)}] \rightarrow \{\alpha_a^{(c)}, \alpha_a^{(c)} + \epsilon_a^{(c)}, \dots, \alpha_a^{(c)} + (n_a^{(c)} - 1)\epsilon_a^{(c)}, \beta_a^{(c)}\}$$

where the number of lattices is $n_a^{(c)} + 1$. Define $a_v^{(c)} = \alpha_a^{(c)} + v\epsilon_a^{(c)}$, the v th element in the latticized parameter space of $a^{(c)}$, and define $b_h^{(c)}, c_l^{(c)}, d_m^{(c)}, \rho_r^{(c)}$ and $X_w^{(c)}$ similarly. Let

$$\bar{\theta}_{\vec{v}}^{(c)} = (a_v^{(c)}, b_h^{(c)}, c_l^{(c)}, d_m^{(c)}, \rho_r^{(c)})$$

as where \vec{v} is (v, h, l, m, r) .

We construct a birth and death generator $\mathbf{A}_\epsilon^{(c)}$, such that $\mathbf{A}_\epsilon^{(c)} \rightarrow \mathbf{A}^{(c)}$. Namely, we construct a birth and death process $(\bar{\theta}_\epsilon^{(c)}, X_{\epsilon_x}^{(c)})$, a simple example of Markov chain, to approximate $(\theta, X^{(c)}(t))$ using the generator for the LSDE process.

$$\begin{aligned} & \mathbf{A}_\epsilon^{(c)} f_c(\bar{\theta}_{\vec{v}}^{(c)}, x_w^{(c)}) \\ &= (a_v^{(c)} + b_h^{(c)} x_w^{(c)}) \left(\frac{f_c(\bar{\theta}_{\vec{v}}^{(c)}, x_w^{(c)} + \epsilon_x^{(c)}) - f_c(\bar{\theta}_{\vec{v}}^{(c)}, x_w^{(c)} - \epsilon_x^{(c)})}{2\epsilon_x^{(c)}} \right) \\ &+ \frac{1}{2} (c_l^{(c)} + d_m^{(c)} x_w^{(c)})^2 \left(\frac{f_c(\bar{\theta}_{\vec{v}}^{(c)}, x_w^{(c)} + \epsilon_x^{(c)}) + c(\bar{\theta}_{\vec{v}}^{(c)}, x_w^{(c)} - \epsilon_x^{(c)}) - 2f_c(\bar{\theta}_{\vec{v}}^{(c)}, x_w^{(c)})}{(\epsilon_x^{(c)})^2} \right) \\ &= \beta^{(c)}(\bar{\theta}_{\vec{v}}^{(c)}, x_w^{(c)}) (f_c(\bar{\theta}_{\vec{v}}^{(c)}, x_w^{(c)} + \epsilon_x^{(c)}) - f_c(\bar{\theta}_{\vec{v}}^{(c)}, x_w^{(c)})) \\ &+ \delta^{(c)}(\bar{\theta}_{\vec{v}}^{(c)}, x_w^{(c)}) (f_c(\bar{\theta}_{\vec{v}}^{(c)}, x_w^{(c)} - \epsilon_x^{(c)}) - f_c(\bar{\theta}_{\vec{v}}^{(c)}, x_w^{(c)})), \end{aligned} \tag{19}$$

where

$$\beta^{(c)}(\bar{\theta}_{\vec{v}}^{(c)}, x_w^{(c)}) = \frac{1}{2} \left(\frac{(c_l^{(c)} + d_m^{(c)} x_w^{(c)})^2}{\epsilon_x^2} + \frac{a_v^{(c)} + b_h^{(c)} x_w^{(c)}}{\epsilon_x} \right),$$

and

$$\delta^{(c)}(\bar{\theta}_{\vec{v}}^{(c)}, x_w^{(c)}) = \frac{1}{2} \left(\frac{(c_l^{(c)} + d_m^{(c)} x_w^{(c)})^2}{\epsilon_x^2} - \frac{a_v^{(c)} + b_h^{(c)} x_w^{(c)}}{\epsilon_x} \right).$$

Note that $\beta^{(c)}(\bar{\theta}_{\vec{v}}^{(c)}, x_w)$ and $\delta^{(c)}(\bar{\theta}_{\vec{v}}^{(c)}, x_w)$ are the birth and death rates, respectively, and should be nonnegative. If necessary ϵ_x can be made smaller to ensure the nonnegativity.

Clearly, $\mathbf{A}_\varepsilon^{(c)} \rightarrow \mathbf{A}^{(c)}$ and we have $(\vec{\theta}_\varepsilon^{(c)}, X_{\varepsilon_x}^{(c)}) \Rightarrow (\vec{\theta}^{(c)}, X^{(c)})$ as $\varepsilon \rightarrow 0$ where $\varepsilon = \max(\varepsilon_a, \varepsilon_b, \varepsilon_c, \varepsilon_d, \varepsilon_\rho, \varepsilon_x)$.

Now, we have the approximate model $(\vec{\theta}_\varepsilon^{(c)}, X_{\varepsilon_x}(t))^{(c)}$ of $(\vec{\theta}^{(c)}, X^{(c)}(t))$. Then, we have the approximate $Y_\varepsilon^{(c)}$ which is defined by Equation (14). Now the counting process observations can be viewed as $Y^{(c)}(t)$ defined by Equation (1) or $Y_\varepsilon^{(c)}(t)$ defined by Equation (14) depending on whether the driving process is $(\vec{\theta}^{(c)}, X^{(c)}(t))$ or $(\vec{\theta}_\varepsilon^{(c)}, X_{\varepsilon_x}^{(c)}(t))$. When we model the parameters and the stock value as $(\vec{\theta}^{(c)}, X^{(c)}(t))$, the counting process observations of stock price are regarded as $Y^{(c)}(t)$. When we intend to compute the Bayes factors for the comparison of two models, we use $(\vec{\theta}_\varepsilon^{(c)}, X_{\varepsilon_x}^{(c)}(t))$ to approach $(\vec{\theta}^{(c)}, X^{(c)}(t))$ and the counting process observations of stock price are regarded as $Y_\varepsilon^{(c)}(t)$.

The recursive algorithm is to compute the joint posterior and Bayes estimates and Bayes factors for the approximate model $(\vec{\theta}_\varepsilon^{(c)}, X_{\varepsilon_x}^{(c)}, Y_\varepsilon^{(c)})$, which is close to the joint posterior and Bayes estimates and Bayes factors of the model $(\vec{\theta}^{(c)}, X^{(c)}, Y^{(c)})$, by Theorem 3, when ε is small.

4.2.2 Step 2: Obtain the SPDEs of the Approximate Model

When $(\vec{\theta}^{(c)}, X^{(c)})$ is approximated by $(\vec{\theta}_\varepsilon^{(c)}, X_{\varepsilon_x}^{(c)})$, $\mathbf{A}^{(c)}$ by $\mathbf{A}_\varepsilon^{(c)}$, and Y by Y_ε , there accordingly exist probability measures $P_\varepsilon^{(c)}$ and $Q_\varepsilon^{(c)}$, which approximate $P^{(c)}$ and $Q^{(c)}$. It can be checked that Assumptions 1 - 5 hold for $(\vec{\theta}_\varepsilon^{(c)}, X_{\varepsilon_x}^{(c)}, Y_\varepsilon^{(c)})$ for $c = 1, 2$ satisfying the conditions of Theorems 1 and 2.

For the posterior, there is one model only and the superscript "(c)" is omitted. Let $(\vec{\theta}_\varepsilon, X_{\varepsilon_x})$ denote the discretized signal.

Definition 7 Let $\pi_{\varepsilon,t}$ be the conditional probability mass function of $(\vec{\theta}_\varepsilon, X_{\varepsilon_x}(t))$ on the discrete state space given $\mathcal{F}_t^{\vec{Y}_\varepsilon}$. Let

$$\pi_\varepsilon(f, t) = E^{P_\varepsilon} [f(\vec{\theta}_\varepsilon, X_{\varepsilon_x}(t)) | \mathcal{F}_t^{\vec{Y}_\varepsilon}] = \sum_{\vec{\theta}_\varepsilon, x_w} f(\vec{\theta}_\varepsilon, x_w) \pi_{\varepsilon,t}(\vec{\theta}_\varepsilon, x_w),$$

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

Then, the normalized filtering equation for the approximate model is given by Equations (15) and (16).

Next, we approximate the Bayes factors.

Definition 8 Let $q_{\varepsilon,t}^{(c)}$ as a conditional mass finite measure of $(\vec{\theta}_\varepsilon, X_{\varepsilon_x}(t))$ on the discrete state space given $\mathcal{F}_t^{\vec{Y}_\varepsilon}$. $q_{\varepsilon,t}^{(c)}$ approximates $q_t^{(c)}$. Let

$$q_\varepsilon^{(c)}(f_c, t) = \sum_{\vec{\theta}_\varepsilon, x_w^{(c)}} f_c(\vec{\theta}_\varepsilon, x_w^{(c)}) q_{\varepsilon,t}(\vec{\theta}_\varepsilon, x_w^{(c)}), \quad (20)$$

where $(\vec{\theta}_\varepsilon, x_w^{(c)})$ goes over all the lattices in the approximate state spaces.

Similarly to the normalized filtering equation, the systems of SPDEs for $q_\varepsilon^{(c)}(f_c, t)$ can be separated into the propagation equation:

$$q_\varepsilon^{(c)}(f_c, t_{i+1}-) = q_\varepsilon^{(c)}(f_c, 0) + \int_{t_i}^{t_{i+1}-} q_\varepsilon^{(c)}(\mathbf{A}_\varepsilon^{(c)} f_c, s) ds, \quad (21)$$

and the updating equation:

$$q_\varepsilon^{(c)}(f_c, t_{i+1}) = \frac{q_\varepsilon^{(c)}(f_c p_j^{(c)}, t_{i+1}-)}{q_\varepsilon^{(3-c)}(f_{3-c} p_j^{(3-c)}, t_{i+1}-)} q_\varepsilon^{(3-c)}(1, t_{i+1}-). \quad (22)$$

Together, these two components form the key to deriving the recursive algorithms.

4.2.3 Step 3: Convert to the Recursive Algorithm

First, we convert Equations (15) and (16) to the recursive algorithm for computing the approximate joint posterior. We show details for how to obtain the algorithm for this one. For converting Equations (21) and (22) to the recursive algorithm, we only provide the algorithm for computing the Bayes factor without giving details of derivation. But the procedure is similar and interested readers are referred to [22].

We define the posterior that the recursive algorithm computes.

Definition 9 *The posterior of the approximate model at time t is denoted by*

$$p_\varepsilon(\vec{\theta}_{\vec{v}}, x_w; t) = \pi_{\varepsilon, t} \{ \vec{\theta}_{\vec{e}} = \vec{\theta}_{\vec{v}}, X_\varepsilon(t) = x_w \}.$$

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

$$\mathbf{I}_{\{ \vec{\theta}_{\vec{e}} = \vec{\theta}_{\vec{v}}, X_\varepsilon(t) = x_w \}}(\vec{\theta}_{\vec{e}}, X_\varepsilon(t)) \quad (23)$$

Then, the following fact emerges:

$$\pi_\varepsilon \left(\beta(\vec{\theta}_{\vec{e}}, X_\varepsilon(t)) \mathbf{I}_{\{ \vec{\theta}_{\vec{e}} = \vec{\theta}_{\vec{v}}, X_\varepsilon(t) + \epsilon_x = x_w \}}(\vec{\theta}_{\vec{e}}, X_\varepsilon(t) + \epsilon_x), t \right) = \beta(\vec{\theta}_{\vec{v}}, x_{w-1}) p_\varepsilon(\vec{\theta}_{\vec{v}}, x_{w-1}; t).$$

Along with similar results, Equation (15) becomes

$$\begin{aligned} p_\varepsilon(\vec{\theta}_{\vec{v}}, x_w; t_{i+1}-) &= p_\varepsilon(\vec{\theta}_{\vec{v}}, x_w; t_i) + \int_{t_i}^{t_{i+1}-} \left(\beta(\vec{\theta}_{\vec{v}}, x_{w-1}) p_\varepsilon(\vec{\theta}_{\vec{v}}, x_{w-1}; t) \right. \\ &\quad \left. - (\beta(\vec{\theta}_{\vec{v}}, x_w) + \delta(\vec{\theta}_{\vec{v}}, x_w)) p_\varepsilon(\vec{\theta}_{\vec{v}}, x_w; t) + \delta(\vec{\theta}_{\vec{v}}, x_{w+1}) p_\varepsilon(\vec{\theta}_{\vec{v}}, x_{w+1}; t) \right) dt, \end{aligned} \quad (24)$$

If a trade at j th price level occurs at time t_{i+1} , the updating Equation (16) can be written as,

$$p_\varepsilon(\vec{\theta}_{\vec{v}}, x_w; t_{i+1}) = \frac{p_\varepsilon(\vec{\theta}_{\vec{v}}, x_w; t_{i+1}-) p(y_j | x_w, \rho_r)}{\sum_{\vec{v}', w'} p_\varepsilon(\vec{\theta}_{\vec{v}'}, x_{w'}; t_{i+1}-) p(y_j | x_{w'}, \rho_{r'})}, \quad (25)$$

where the summation goes over the total discretized space, and $p(y_j | x_w, \rho_r)$, the transition probability from x_w to y_j , is specified by Equation (31) in Appendix A.

In the second substep, we approximate the time integral in Equation (24) with an Euler scheme to obtain a recursive algorithm further discrete in time. After excluding the probability-zero event that two or more jumps occur at the same time, there are two possible cases for the inter-trading time. Case 1, if $t_{i+1} - t_i \leq LL$, the length controller in the Euler scheme, then we approximate $p(\vec{\theta}_{\vec{v}}, x_w; t_{i+1}-)$ as

$$p(\vec{\theta}_{\vec{v}}, x_w; t_{i+1}-) \approx p(\vec{\theta}_{\vec{v}}, x_w; t_i) + \left[\beta(\vec{\theta}_{\vec{v}}, x_{w-1})p(\vec{\theta}_{\vec{v}}, x_{w-1}; t_i) - (\beta(\vec{\theta}_{\vec{v}}, x_w) + \delta(\vec{\theta}_{\vec{v}}, x_w))p(\vec{\theta}_{\vec{v}}, x_w; t_i) + \delta(\vec{\theta}_{\vec{v}}, x_{w+1})p(\vec{\theta}_{\vec{v}}, x_{w+1}; t_i) \right] (t_{i+1} - t_i). \quad (26)$$

Case 2, if $t_{i+1} - t_i > LL$, then we can choose a fine partition $\{t_{i,0} = t_i, t_{i,1}, \dots, t_{i,n} = t_{i+1}\}$ of $[t_i, t_{i+1}]$ such that $\max_j |t_{i,j+1} - t_{i,j}| < LL$ and then approximate $p(\vec{\theta}_{\vec{v}}, x_l; t_{i+1}-)$ by applying repeatedly Equation (26) from $t_{i,0}$ to $t_{i,1}$, then $t_{i,2}, \dots$, until $t_{i,n} = t_{i+1}$.

Equations (25) and (26) consist of the recursive algorithm we employ to calculate the approximate posterior at time t_{i+1} for $(\vec{\theta}, X(t_{i+1}))$ based on the posterior at time t_i . At time t_{i+1} , the Bayes estimates of $\vec{\theta}$ and $X(t_{i+1})$ are the expected values of the corresponding marginal posteriors.

To complete the algorithm for posterior, we choose a reasonable prior. Assume independence between $X(0)$ and $\vec{\theta}$. The prior for $X(0)$ can be set by $P\{X(0) = Y(t_1)\} = 1$ where $Y(t_1)$ is the first trade price of a data set because they are very close. For other parameters, we can simply take a uniform prior to the discretized state space of $\vec{\theta}$, which is used also for the algorithm for Bayes factors and in the simulation and the real data example. At $t = 0$, we select the prior as below:

$$p(\vec{\theta}_{\vec{v}}, x_w; 0) = \begin{cases} \frac{1}{(1+n_a)(1+n_b)(1+n_c)(1+n_d)(1+n_\rho)} & \text{if } x_w = Y(t_1) \\ 0 & \text{otherwise} \end{cases}.$$

In the rest of this subsection, we briefly present the algorithm for computing the Bayes factors.

We define the conditional ratio measure that the recursive algorithm computes.

Definition 10 *The conditional ratio measure of the approximate model at time t is denoted by*

$$q_\varepsilon^{(c)}(\vec{\theta}_{\vec{v}}^{(c)}, x_w^{(c)}; t) = q_{\varepsilon, t}^{(c)}\{\vec{\theta}_{\vec{v}}^{(c)} = \vec{\theta}_{\vec{v}}^{(c)}, X_\varepsilon^{(c)}(t) = x_w^{(c)}\}.$$

Take $f^{(c)}$ as the indicator function

$$\mathbf{I}^{(c)}(\vec{\theta}_{\vec{v}}^{(c)}, x_w^{(c)}) = \mathbf{I}_{\{\vec{\theta}_{\vec{v}}^{(c)} = \vec{\theta}_{\vec{v}}^{(c)}, X_\varepsilon^{(c)}(t) = x_w^{(c)}\}}^{(c)}(\vec{\theta}_{\vec{v}}^{(c)}, X_\varepsilon^{(c)}(t)). \quad (27)$$

Then,

$$q_\varepsilon^{(c)}(\mathbf{I}^{(c)}(\vec{\theta}_{\vec{v}}^{(c)}, x_w^{(c)}), t) = q_\varepsilon^{(c)}(\vec{\theta}_{\vec{v}}^{(c)}, x_w^{(c)}; t)$$

Equations (21) and (22) become

$$q_\varepsilon^{(c)}(\vec{\theta}_{\vec{v}}^{(c)}, x_w^{(c)}; t_{i+1}-) \approx q_\varepsilon^{(c)}(\vec{\theta}_{\vec{v}}^{(c)}, x_w^{(c)}; t_i) + \left[\beta^{(c)}(\vec{\theta}_{\vec{v}}^{(c)}, x_{w-1}^{(c)})q_\varepsilon^{(c)}(\vec{\theta}_{\vec{v}}^{(c)}, x_{w-1}^{(c)}; t_i) + \delta^{(c)}(\vec{\theta}_{\vec{v}}^{(c)}, x_{w+1}^{(c)})q_\varepsilon^{(c)}(\vec{\theta}_{\vec{v}}^{(c)}, x_{w+1}^{(c)}; t_i) \right]$$

$$-(\beta^{(c)}(\vec{\theta}_{\vec{v}}^{(c)}, x_w^{(c)}) + \delta^{(c)}(\vec{\theta}_{\vec{v}}^{(c)}, x_w^{(c)}))q_\varepsilon^{(c)}(\vec{\theta}_{\vec{v}}^{(c)}, x_w^{(c)}; t_i)](t_{i+1} - t_i) \quad (28)$$

and

$$\begin{aligned} q_\varepsilon^{(c)}(\vec{\theta}_{\vec{v}}^{(c)}, x_w^{(c)}; t_{i+1}) = & \\ & \frac{q_\varepsilon^{(c)}(\vec{\theta}_{\vec{v}}^{(c)}, x_w^{(c)}; t_{i+1-})p_j^{(c)}(y_j | (x_w^{(c)}, \rho_r^{(c)}))}{\sum_{\vec{v}', w'} q_\varepsilon^{(3-c)}(\vec{\theta}_{\vec{v}'}^{(3-c)}, x_{w'}^{(3-c)}; t_{i+1-})p_j^{(3-c)}(y_j | (x_{w'}^{(3-c)}, \rho_{r'}^{(3-c)}))} \\ & \times \left(\sum_{\vec{v}', w'} q_\varepsilon^{(3-c)}(\vec{\theta}_{\vec{v}'}^{(3-c)}, x_{w'}^{(3-c)}, t_{i+1-}) \right) \end{aligned} \quad (29)$$

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

Equations (28) and (29) compose the recursive algorithm we employ to calculate the approximate conditional ratio measures. At time t_{i+1} , the Bayes factor

$$B_{21}(t_{i+1}) \approx q_\varepsilon^{(2)}(1, t_{i+1}) = \sum_{\vec{v}', w'} q_\varepsilon^{(2)}(\vec{\theta}_{\vec{v}'}^{(2)}, x_{w'}^{(2)}; t_{i+1}),$$

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

Finally, we note that the statistical and computational concerns for a prior on a parameter have two aspects: suitable range and mesh size. Usually, the marginal posterior of a parameter obtained from a large data set is concentrated on a small area around the true value. This implies that one needs to run the program several times in order to identify suitable range and mesh size of parameters for a real world data set. If the true parameter is out of the range, say, smaller (larger) than the lower (upper) boundary, then the marginal posterior of this parameter would be one or very close to one on the lower (upper) boundary. Such indication gives the direction to adjust the range of parameter space of the prior until no spike on the boundary. After having a suitable range, we may choose a suitable mesh size, which ideally produces a posterior with a unique modal and bell-shaped distribution as shown in Table 5.1 of [30]. After obtaining suitable ranges and mesh sizes for both models in Bayes estimation via filtering, we then employ those suitable ranges and mesh sizes for the parameters in the computer program for computing Bayes factors in order to obtain the Bayes factors.

4.2.4 Consistency of the Recursive Algorithms

There are two approximations in our recursive algorithms to compute the posteriors and Bayes factors. One is to approach the integral in the propagation equations by Euler scheme, whose convergence is well-known. The other one, which is more important, is the approximation of Equations (7) by Equations (15) and (16), and the approximation of Equations (10) and (11) by Equations (21) and (22). Since, for $c = 1, 2$, $(\vec{\theta}_{\vec{c}^{(c)}}, X_{\vec{c}^{(c)}}^{(c)}) \Rightarrow (\vec{\theta}^{(c)}, X^{(c)})$ by construction, Theorem 3 warrants these convergence in the sense of the weak convergence in the Skorohod topology, that is, the consistency of the Bayes factors.

4.3 A Monte Carlo Example

Having constructed the recursive algorithm to compute the Bayes factors, we move on to develop the software to implement the algorithms. For computing the posterior, a Fortran

program for the recursive algorithm is developed to compute, at each trading time t_i , the joint posterior of $(\vec{\theta}, X(t))$, their marginal posteriors, their Bayes estimates and their standard errors (SE), respectively. The recursive algorithm is fast enough to generate real-time Bayes estimates. We test the algorithm extensively and verify it on Monte Carlo data, where we know the true parameters. Care is required in selecting the lattice of parameter values for the prior. Ideally, the approximate posterior should resemble a bell-shaped curve. That is, the posterior should be weighted towards the interior of the lattice, but it should not be weighted so strongly to any specific point that the relative probability of that lattice point being the true value is close to 1 (i.e. a bar chart of the relative distribution should be shaped like a normal or bell-shaped curve, not a single spike at one point). Monte Carlo studies indicate that the posterior is robust to the prior. Namely, for reasonable priors, as long as the range covers the true values and the lattices are reasonably fine, the Bayes estimates will converge to the true values. Below we give one Monte Carlo example with 50,000 data to show the effectiveness of Bayes estimates. For parameters $\vec{\theta}$, their Bayes estimates converge to their true values and the two-SE bounds become smaller and smaller, and goes to zero as in the case of GBM in Zeng [30]. Hence, only the final Bayes estimates, their SE, and true values are presented in Table 2. The true values are close to the Bayes estimates and all within two SE bounds.

Table 2: Bayes Estimates for the LSDE FM Model (Simulation)

Parameter	True Value	Bayes Estimate	St. Error
a	3.000E-7	2.951E-7	1.013E-6
b	5.000E-7	4.542E-8	1.001E-6
c	2.000E-3	2.4990E-3	1.001E-3
d	3.000E-6	1.1750E-6	1.074E-6
ρ	2.260E-1	2.247E-1	3.367E-3

The value of parameters are for per second.

Using the same Monte Carlo data, we calculate the Bayes factor of the full LSDE FM model versus the restricted LSDE FM model with $c = 0$ (a wrong model) for checking the effectiveness for model selection. The final Bayes factor is 13,152, which is larger than 150, the decisive benchmark for rejecting the restricted LSDE FM model.

4.4 Real Data Example

For this example the data used was the Microsoft trade by trade stock prices taken from January and February 1994 as provided by the TAQ (Trade and Quote) from the NYSE. After suitable filtering the resulting data used for the estimation was comprised of 49,937 Microsoft trades. The basic summary statistics can be found in Table 3 and the breakdown of the trades by the fractional portions of the price appears in Table 4.

Again, we note that clustering away from the odd eighths is present in this data. Method of relative frequency produces the estimations for the clustering parameters: $\alpha = .2414$ and $\beta = .3502$.

Table 3: Summary Statistics for MSFT Jan. & Feb. 1994

	Size	Mean	Median	St. Dev.	Skewness	Kurtosis
MSFT	49937	82.465	83.25	1.854	-0.201	-1.446

Table 4: Freq. of the Fractional Parts of the Prices for MSFT Jan. & Feb. 1994

	0	1/8	1/4	3/8	1/2	5/8	3/4	7/8
MSFT Freq.	11218	2791	9136	2199	10009	2826	9376	2382
Rel. Freq.	.2246	.0559	.183	.044	.2004	.0566	.1878	.0477

We then estimate two models for this data each having a value process with the form of an LSDE. For the full LSDE model the Bayes estimate for each parameter along with the associated standard error appear in Table 5. For a restricted model with $c = 0$, the estimates are given in Table 6.

Table 5: Bayes Est. for the Full LSDE FM Model (MSFT Jan. & Feb. 1994)

Parameter	Bayes Estimate	St. Error
a	2.196E-6	1.777E-6
b	9.600E-8	1.001E-6
c	-1.6760E-5	5.115E-5
d	6.999E-5	1.002E-6
ρ	4.133E-1	3.150E-3

The value of parameters are for per second.

We then determined the Bayes factors to compare the full model vs. the restricted model of $c = 0$. The Bayes factor is $2.90E28$, which is much larger than 150 again. So, we reject the restricted model of $c = 0$ and conclude the full model fits better than the restricted model for this Microsoft UHF data set.

5 Conclusion

This chapter reviews recent development of a rich class of filtering models with counting process observations for the micromovement of asset price and the related Bayesian inference via filtering. A specific FM model built upon LSDE is used to exemplify how to develop a FM model and further exemplify how to develop the recursive algorithms for computing the posterior and the Bayes factors. The general model and its developed statistical analysis offer strong potential to relate or illuminate aspects of the rich theoretical literature on market microstructure and trading mechanism. Furthermore, Bayesian model selection via filtering provides a general, powerful tool to test related market microstructure theories, represented by the FM models. We may test whether NASDAQ has less trading noise after a market reform as in Barclay et al [3], test whether information affects trading intensity as in Easley

Table 6: Bayes Est. for the Restricted-'c=0' Case (MSFT Jan. & Feb. 1994)

Parameter	Bayes Estimate	St. Error
a	-5.000E-6	1.008E-6
b	1.999E-7	1.000E-6
d	7.000E-4	1.000E-6
ρ	4.149E-1	8.312E-4

The value of parameters are for per second.

and O'Hara [9] and Engle [10]. Finally, a more general FM model with statistical analysis can be found in [33].

A Appendix: More on Clustering Noise

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

$$r(y) = \begin{cases} 3 & \text{if the fractional part of } y \text{ is odd eighth} \\ 2 & \text{if the fractional part of } y \text{ is odd quarter} \\ 1 & \text{if the fractional part of } y \text{ is a half or zero} \end{cases} \quad (30)$$

The biasing rules specify the transition probabilities from y' to y , $p(y|y')$. Then, $p(y|x)$, the transition probability can be computed through $p(y|x) = \sum_{y'} p(y|y')p(y'|x)$ where $p(y'|x) = P\{V = y' - R[x, \frac{1}{8}]\}$. Suppose $D = 8|y - R[x, \frac{1}{8}]|$. Then, $p(y|x)$ can be calculated as, for example, when $r(y) = 2$,

$$p(y|x) = \begin{cases} (1 - \rho)(1 + \alpha\rho) & \text{if } r(y) = 2 \text{ and } D = 0 \\ \frac{1}{2}(1 - \rho)[\rho + \alpha(2 + \rho^2)] & \text{if } r(y) = 2 \text{ and } D = 1 \\ \frac{1}{2}(1 - \rho)\rho^{D-1}[\rho + \alpha(1 + \rho^2)] & \text{if } r(y) = 2 \text{ and } D \geq 2 \end{cases} \quad (31)$$

References

- [1] Ait-Sahalia, Y., P. A. Mykland and L. Zhang, 2005. *How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise*. Review of Financial Studies, **18**, 351-416.
- [2] Bandi, F. M. and J. R. Russell, 2006. *Separating microstructure noise from volatility*. Journal of Financial Economics, **79**, 655-692.
- [3] 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*, Journal of Finance, **54**, 1-34.
- [4] Black, F., 1986. *Noise*, Journal of Finance, **41**, 529-543.
- [5] Bremaud, P., 1981. *Point Processes and Queues: Martingale Dynamics*. Springer-Verlag, New York.

- [6] Cvitanic, J., R. Liptser, and B. Rozovskii, 2006. *A filtering approach to tracking volatility from prices observed at random times*. Annals of Applied Probability, **16**, 1633-1652.
- [7] Cvitanic, J., B. Rozovskii, and I. Zaliapin, 2006. *Numerical estimation of volatility values from discretely observed diffusion data*. Journal of Computational Finance, **4**, 1-36.
- [8] Davis, M. H. A., A. Segall, and T. Kailath, 1975. *Nonlinear filtering with counting observations*, IEEE Trans. Inf. Theory, **21**, 1431-49.
- [9] Easley, D. and M. O'Hara., 1992 *Time and the process of security price adjustment*, Journal of Finance, **47**, 577-605.
- [10] Engle, R., 2000. *The econometrics of ultra-high-frequency data*, Econometrica, **68**,1-22.
- [11] Elliott, R. J. and W. P. Malcolm, 2005. *General smoothing formulas for Markov-modulated Poisson observations*, IEEE Trans. Aut. Control, **50**, 1123-1134.
- [12] Ethier, S. N. and T.G. Kurtz, 1986. *Markov Processes : Characterization and Convergence*. Wiley, New York.
- [13] Fan, J. and Y. Wang, 2007. *Multi-scale jump and volatility analysis for high-Frequency financial data*, Journal of American Statistical Association, **102**, 1349-1362.
- [14] Frey, R., 2000. *Risk-minimization with incomplete information in a model for high-frequency data*, Mathematical Finance **10**, 215-225.
- [15] Frey, R. and W.J. Runggaldier, 2001, *A nonlinear filtering approach to volatility estimation with a view towards high frequency data*. International Journal of Theoretical and Applied Finance **4**, 199-210.
- [16] Harris, L., 1991. *Stock price clustering and discreteness*, Review of Financial Studies, **4**, 389-415.
- [17] Hasbrouck, J., 1996. *Modeling market microstructure time series*, in G. Maddala and C. Rao, eds, 'Handbook of Statistics', Vol. 14, North-Holland, Amsterdam, pp. 647-692.
- [18] Hasbrouck, J., 2002. *Stalking the "efficient price" in market microstructure specifications: an overview*, Journal of Financial Markets **5**, 329-339.
- [19] Jeffreys, H., 1961. *Theory of Probability*, Oxford University, London.
- [20] Kass, R.E. and A. E. Raftery, 1995. *Bayes factors and model uncertainty*, Journal of the American Statistical Association, **90**, 773-795.
- [21] Kliemann, W., W. Koch, and F. Marchetti, *On the unnormalized solution of the filtering problem with counting process observations*, IEEE Trans. Inf. Theory, **36**, 1415-1425.
- [22] Kouritzin, M. and Y. Zeng, 2005. *Bayesian model selection via filtering for a class of micro-movement models of asset price*, International Journal of Theoretical and Applied Finance, **8**, 97-121.

- [23] Lee K. and Y. Zeng, 2007. *Risk minimization for a filtering micromovement model of asset price*. Working paper. University of Missouri at Kansas City.
- [24] Li, Y. and P. A. Mykland, 2007. *Are volatility estimators robust with respect to modeling assumptions?* Bernoulli, **13**, 601-622.
- [25] Liptser, R.S. and A.N. Shiryaev, 2002. *Statistics of Random Processes*, 2nd Edition, Volume 2, Springer-Verlag, New York.
- [26] Stoll, H., 2000. *Friction*, Journal of Finance, **55**, 1479-1514.
- [27] Snyder, D. L. 1972a. *Filtering and detection for doubly stochastic poisson processes*. IEEE Trans. Inf. Theory, **18**, 91-102.
- [28] Snyder, D. L. 1972b. *Smoothing for doubly stochastic poisson processes*. IEEE Trans. Inf. Theory, **18**, 558-562.
- [29] Xiong J. and Y. Zeng, 2008 *Mean-variance portfolio selection for A filtering micromovement model of asset price*. Working paper. University of Tennessee.
- [30] Zeng, Y., 2003. *A partially observed model for micromovement of asset prices with Bayes estimation via filtering*, Mathematical Finance, **13**, 411-444.
- [31] Zeng, Y., 2004. *Estimating stochastic volatility via filtering for the micro-movement of asset prices*, IEEE Transactions on Automatic Control, **49**, 338-348.
- [32] Zeng, Y., 2005. *Bayesian inference via filtering for a class of counting processes: Application to the micromovement of asset price*, Statistical Inference for Stochastic Processes, **8**, 331 - 354.
- [33] Zeng, Y., 2008. *Econometric analysis via filtering for financial ultra-high frequency data*, Working paper. University of Missouri at Kansas City.
- [34] Zhang, L., Mykland, P. A. and Ait-Sahalia, Y., 2005. *A tale of two time scales: Determining integrated volatility with noisy high frequency data*, Journal of the American Statistical Association, **100**, 1394-1411.