

Bayes estimation via filtering equation through implicit recursive algorithms for financial ultra-high frequency data

BRENT BUNDICK, NOAH RHEE AND YONG ZENG^{*,†}

We review a recently proposed general partially-observed framework of Markov processes with marked point process observations for financial ultra-high frequency (UHF) data, and the related Bayes estimation via filtering equation (BEFE), a stochastic PDE approach. In this paper, we show how the BEFE through explicit recursive algorithms becomes bottlenecked when the tick size is reduced from \$1/8 to \$1/100, and we develop the BEFE through implicit recursive algorithms, greatly improving the computational efficiency. We demonstrate the substantial computation gained in implementing real-time BEFE for an illustrating but practical model using simulated data. The new implicit recursive algorithm is applied to a real stock price UHF data set, and is capable of producing real time Bayes parameter estimates of the model.

AMS 2000 SUBJECT CLASSIFICATIONS: Primary 62M05, 62F15, 62P05; secondary 60H35, 60G55, 65C60, 93E11.

KEYWORDS AND PHRASES: Bayes estimation, Implicit methods, Marked point process, Market microstructure noise, Markov chain approximation method, Nonlinear filtering, Partially observed model, Ultra-high frequency data.

1. INTRODUCTION

Transactions data in financial markets have become widely available in recent years. Such time-stamped (“tick”) data, containing the most detailed information for price evolution and quote revisions, are referred to as *ultra-high frequency* (UHF) data in [11]. The direct modeling and analysis of these data, which make full use of the information (including timing) at one’s disposal, is essential for insight concerning market microstructure theory.

UHF data contains all transaction prices arrived at irregular and random times and can be represented by two kinds of random variables. One is transaction time, T_i , and the other is a random element, Y_i , describing such as the observed price and/or the trading volume. Hence, UHF data

are naturally modeled as a *Marked Point Process* (MPP), $\Phi = \{(T_i, Y_i)\}_{i \geq 1}$, which consists of a *point process*, $\{T_i\}$, describing an increasing sequence of transaction times, and *marks*, $\{Y_i\}$, representing, as mentioned, prices and volumes associated with the trades.

There are two different views on how to treat the observations, leading to the different formulations of the statistical foundation for analyzing UHF data. Focused on modeling random arrival times and from the viewpoint of time series, econometricians naturally treat UHF data as an irregularly-spaced time series and this view is well expressed in [11]. UHF data has a second stylized fact, namely, UHF data contains microstructure (or trading) noise due to price discreteness, price clustering, bid-ask bounce and other market microstructure issues. To accommodate the second stylized fact and from the standpoint of stochastic process, the second view normally perceived by probabilists is to treat the transaction observations as *an observed sample path of MPP from a partially observed Markov process*. This view was advocated in [33], where the data is treated as a collection of counting processes, a special case of MPP observations.

In this paper, we first review a general partially-observed model for financial UHF data recently proposed in [20] to unify the two different views and merge their strengths. The model connects to five literatures: (1) state-space models; (2) hidden Markov models; (3) Vector AutoRegressive (VAR) structural models in market microstructure theory surveyed in [18]; (4) estimating Markov processes sampled at discrete or random times via the operator approach in [17], [1], [9], [3] and [7] among others; and (5) realized volatility with microstructure noise in [38], [4], [37], [15], [25], and [5] among many others. The model is presented in two equivalent representations in the sense that they have the same probability law. Then, we review the corresponding Bayes estimation via filtering equation (BEFE). Namely, we review the normalized filtering equation to characterize the evolution of the posterior measure-valued process, review a weak convergence theorem and the blueprint for constructing consistent, easily-parallelizable, recursive algorithms to calculate the joint posteriors and the related Bayes estimates for streaming UHF data.

One key step in constructing recursive algorithms is to approximate the time integral by explicit methods such as an

*Corresponding author.

†Supported by NSF grants DMS-0604722, TG-DMS090039, TG-DMS100019, and DMS-1228244 as well as a University Missouri Research Board grant.

Euler scheme. BEFE through explicit recursive algorithms has found successes of real-time Bayes estimation in models such as geometric Brownian motion (GBM) and jumping stochastic volatility among others for UHF stock prices when the tick size was 1/8 or 1/16 (that is, before the year of 2000). See, for example, [36], [34], [22], [35], and [30]. However, explicit recursive algorithms from BEFE suffer two curses when the tick size was reduced to 1/100 after the year of 2,000. One is the curse of grid size, which further increases exponentially as the number of parameter increases. The other is the curse of time-step size, which shrinks dramatically (even over one hundred folds) as the tick size is reduced to \$1/100 and the volatility keeps high, such as 30% annually for higher.

In this paper, we propose two strategies, each as a cure to one of the two curses. The first curse of grid size can be solved by a well-designed parallel computing according to each parameter set. And the second curse of time-step size can be cured by developing BEFE through implicit methods instead of explicit methods. Based on these two strategies, we develop BEFE through implicit recursive algorithms, a general approach to constructing consistent, easily-parallelized, efficient recursive algorithms for real time Bayes estimation of the model.

Using a partially-observed geometric Brownian motion (GBM) sampled at random times with trading noises, we show the computation challenge of real time Bayes estimation through explicit methods when the tick size is in cent. A key for valid posterior approximation computed by the explicit recursive algorithm is to keep the approximate probability masses positive over time. To achieve this, however, the time-step size has to shrink greatly when the tick size becomes \$1/100. We explain how implicit methods can improve the computation efficiently and show the nonnegativity of probability masses computed in the implicit recursive algorithm. We demonstrate the substantial speed gained in implementing real-time BEFE using simulated data when the volatility is high. The new implicit recursive algorithm is applied to a real stock price UHF data set, and is capable of producing reasonable real time Bayes parameter estimates of the model.

The rest of this paper goes as follows: Section 2 presents the general model in two equivalent forms and reviews the BEFE. Section 3 develops BEFE through implicit recursive algorithms with parallel computing for the general model. Section 4 presents a specific GBM model sampled at random times with trading noises at tick size of \$1/100, presents the old recursive algorithm constructed via the explicit Euler scheme, and develops the new implicit recursive algorithm. Moreover, we prove the nonnegativity for the implicit recursive algorithm, demonstrate the tremendous computational gain using simulation data and apply the implicit recursive algorithm to a real stock transactions price data. Section 5 concludes.

2. THE MODEL AND BAYES ESTIMATION VIA FILTERING EQUATION

We review the partially-observed model and the corresponding BEFE in [20].

2.1 The model

We present the model in two equivalent forms and briefly discuss its generality.

2.1.1 Representation I: Random arrival time state-space model

This representation has the similar components: state process, observation times, observations and noise. However, there are two significant differences from the usual state-space models: random arrival times and continuous-time state process.

State process

The extended state process (θ, X) , allowing θ to potentially change in continuous time, satisfies a mild Markov assumption.

Assumption 2.1. (θ, X) is a $p + m$ -dimension vector Markov process that is the solution of a martingale problem for a generator \mathbf{A}_v such that

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

is a $\mathcal{F}_t^{\theta, X, V}$ -martingale, where $\mathcal{F}_t^{\theta, X, V}$ is the σ -algebra generated by $(\theta(s), X(s), V(s))_{0 \leq s \leq t}$, and f is in the domain of \mathbf{A}_v .

Here, V , which can be a vector process, represents other observable economic or market factors. V can influence X , which can be interpreted as an intrinsic value vector process for m assets, through the generator, \mathbf{A}_v . The generator and martingale problem approach (see, for example, [14]) furnishes a powerful tool for the characterization of Markov processes.

Observation times

$T_1, T_2, \dots, T_i, \dots$, are allowed to be a general point process, specified by a nonnegative \mathcal{F}_t -predictable stochastic (or conditional) intensity in the following form: $\bar{\lambda}(t) = \bar{\lambda}(\theta(t), X(t), V^{t-}, \Phi^{t-}, t)$ where $V^t = V(\cdot \wedge t)$ denotes the sample path of V up to time t and similarly $\Phi^{t-} = \{(T_i, Y_i) : T_i < t\}$.

Observations

Taking a value in mark space \mathbb{Y} , the noisy observation at event time T_i , $Y(T_i)$, is modeled by $Y(T_i) = F(X(T_i))$. Note that $F(\cdot)$ in $y = F(x)$ is a random transformation from x to y , specified by a transition probability $p(y|x)$, which is \mathcal{F}_t -predictable.

2.1.2 Representation II: Filtering with MPP observations

Under this representation, (θ, X) becomes the Markov signal, which is partially observed through a MPP, $\Phi =$

$\{T_i, Y_i\}_{i \geq 1}$. An auxiliary predictable process V is incorporated in both the signal and MPP, and (θ, X, Φ, V) is framed as a filtering model with MPP observations.

Let $(\mathbb{Y}, \mathcal{Y}, \mu)$ be a measure space for \mathbb{Y} , the mark space of observations with a finite measure μ ($\mu(\mathbb{Y}) < \infty$). An MPP can alternatively be described by a random counting measure, $\Phi(t, B)$, recording the cumulative number of marks in a set $B \in \mathcal{Y}$ up to time t . Let $\Phi(t, B) = \int_0^t \int_B \Phi(dt, dy)$ where $\Phi(dt, dy)$ is given by

$$(1) \quad \Phi(dt, dy) = \sum_{i \geq 1} \delta_{\{T_i, Y_i\}}(t, y) dt dy$$

with $\Phi(\{0\} \times \mathbb{Y}) = 0$ and $\delta_{\{T_i, Y_i\}}(t, y)$ is the Dirac delta-function on $\mathbb{R}^+ \times \mathbb{Y}$.

The assumption below ensures that this representation is equivalent with the previous one:

Assumption 2.2. Under \mathbb{P} , the stochastic intensity kernel of $\Phi = \{(T_n, Y_n)\}_{n \geq 1}$ is given by

$$(2) \quad \begin{aligned} \lambda(t, dy) &= \lambda(t, dy; \theta(t), X(t), V^{t-}, \Phi^{t-}, t-) \\ &= \bar{\lambda}(\theta(t), X(t), V^{t-}, \Phi^{t-}, t-) \\ &\quad \times p(dy|X(t); \theta(t), V^{t-}, \Phi^{t-}, t-). \end{aligned}$$

In the next assumption, we assume the existence of a simple unit Poisson random measure as a reference measure with a key independence property.

Assumption 2.3. There exists a *reference measure* Q , $\mathbb{P} \ll Q$ so that under Q , (θ, X) and V are independent of $\Phi = \{(T_n, Y_n)\}_{n \geq 1}$ and the compensator of MPP Φ is $\gamma_Q(d(t, y)) = \mu(dy)dt$.

We use $E^Q[X]$ or $E^{\mathbb{P}}[X]$ to indicate that the expectation is taken with respect to a specific probability measure. The next mild assumption ensures that such a reference measure Q exists. Let the ratio of the mark distributions under \mathbb{P} and Q be

$$(3) \quad \begin{aligned} r(y) &= r(y; \theta(t), X(t), V^{t-}, \Phi^{t-}, t-) \\ &= \frac{p(dy|X(t); \theta(t), V^{t-}, \Phi^{t-}, t-)}{\mu(dy)}. \end{aligned}$$

and the ratio of the compensators under \mathbb{P} and Q be

$$(4) \quad \begin{aligned} \zeta(t, y) &= \frac{\gamma_P(d(t, y))}{\gamma_Q(d(t, y))} = \frac{\lambda(t, dy)}{\mu(dy)} \\ &= \bar{\lambda}(\theta(t), X(t), V^{t-}, \Phi^{t-}, t-) \\ &\quad \times r(y; \theta(t), X(t), V^{t-}, \Phi^{t-}, t-). \end{aligned}$$

Let $L(t) = \frac{dP}{dQ}(t)$ be the Radon-Nikodym derivative given below with more explanation in [21].

$$L(t) = L(0) \exp \left\{ \int_0^t \int_{\mathbb{Y}} \log \zeta(s-, y) \Phi(d(s, y)) - \int_0^t \int_{\mathbb{Y}} [\zeta(s-, y) - 1] \mu(dy) ds \right\}.$$

Clearly, $\zeta(t, y)$ determines $L(t)$.

Assumption 2.4. The $\zeta(t, y)$ defined in (4) satisfies the condition that $E^Q[L(T)] = 1$ for all $T > 0$.

The final assumption is a technical condition that ensures that $\{T_i\}$ is non-explosive and that the filtering equation is well-defined.

Assumption 2.5. $\int_0^t E^{\mathbb{P}}[\bar{\lambda}(s)] ds < \infty$, for $t > 0$.

In summary, this rich framework subsumes two classes of existing models: *Class I* contains models that lack a continuous-time latent X , while *Class II* contains models that incorporate continuous-time latent X , but do not include confounding, observable factors V . For surveys on Class I, see [27] and [13]. As an alternative, Class II contains many classical filtering problems with MPP observations (see for example, [10] and references therein) and the model of [33] and its multivariate generalization ([28] and [29]), and filtering models that estimate market volatility from UHF data such as [16] and [8].

Moreover, the proposed model provides a new class of interesting models whose new key element is V , containing observable factors. One example is provided in Section 4 of [20], where V includes two variables incorporated into the volatility of a GBM intrinsic value process X . Moreover, the flexibility of the model allows for the marks $\{Y_j\}$ to represent trade price, volume, bid-ask quotes, or some combination and to be influenced by V . Brought together, these characteristics provide new possibilities in empirical studies of market microstructure theory.

2.2 Bayes estimation via filtering equation

We review the normalized filtering equation for the posterior process, a weak convergence theorem, and the blueprint for constructing recursive algorithms.

2.2.1 The Continuous-time posterior of $(\theta(t), X(t))$ and its filtering equation

We first define the conditional distribution.

Definition 2.1. Let π_t be the conditional distribution of $(\theta(t), X(t))$ given $\mathcal{F}_t^{\Phi, V}$ and $(\theta(0), X(0))$, and let

$$\begin{aligned} \pi(f, t) &= E^P[f(\theta(t), X(t)) | \mathcal{F}_t^{\Phi, V}] \\ &= \int f(\theta, x) \pi_t(d\theta, dx). \end{aligned}$$

When $(\theta(0), X(0))$ has a prior distribution in a Bayesian paradigm, π_t is the continuous-time posterior distribution.

The collection of $\pi(f, t)$ for all continuous and bounded f specifies π_t . The posterior is the key for Bayesian point and interval estimation.

For each $t > 0$, π_t is a probability distribution, which is a measure. Hence, $\{\pi_t\}_{t \geq 0}$ is a measure-valued process, which is of infinite dimension and is uniquely characterized by the following normalized filtering equation, a stochastic partial differential equation (SPDE).

Theorem 2.1. *Suppose that (θ, X, Φ, V) satisfies Assumptions 2.1 – 2.5. Then, π_t is the unique measure-valued solution of the SPDE under P , the normalized filtering equation,*

$$(5) \quad \begin{aligned} \pi(f, t) &= \pi(f, 0) + \int_0^t \pi(\mathbf{A}_v f, s) ds \\ &+ \int_0^t \int_{\mathbb{Y}} \left[\frac{\pi(\zeta(y)f, s-)}{\pi(\zeta(y), s-)} - \pi(f, s-) \right] \\ &(\Phi(d(s, y)) - \pi(\zeta(y), s) \mu(dy) ds). \end{aligned}$$

Moreover, when the stochastic intensity $\bar{\lambda}(t)$ is $\mathcal{F}_t^{\Phi, V}$ -predictable, the normalized filtering equation is simplified as

$$(6) \quad \begin{aligned} \pi(f, t) &= \pi(f, 0) + \int_0^t \pi(\mathbf{A}_v f, s) ds \\ &+ \int_0^t \int_{\mathbb{Y}} \left[\frac{\pi(r(y)f, s-)}{\pi(r(y), s-)} - \pi(f, s-) \right] \Phi(d(s, y)), \end{aligned}$$

where $r(y) = r(y; \theta(s), X(s), \Phi^{s-}, V^{s-}, s-)$ is defined in (3).

The case that $\{T_i\}$'s stochastic intensity $\bar{\lambda}(t)$ is $\mathcal{F}_t^{\Phi, V}$ -predictable is general and includes Cox model and the more general proportional hazard formulations, Autoregressive Conditional Duration (ACD) model proposed in [12] and its variants, and the Hawkes process (see, for example [24] and [2] for recent applications in finance). In that case, Equation (5) is reduced to (6), significantly reducing the computation. Further note that the stochastic intensity disappears in the simplified filter. Hence, the Bayes estimates of $(\theta(t), X(t))$ are *model-free* of the assumptions on event times. The example in Section 4 makes use of such advantages. The tradeoff is that the relationship between event times and the state (θ, X) is ignored.

We focus on the recursive filter (6), which can be separated into the *propagation* equation between trading times t_i and t_{i+1} –:

$$(7) \quad \pi(f, t_{i+1}-) = \pi(f, t_i) + \int_{t_i}^{t_{i+1}-} \pi(\mathbf{A}_v f, s) ds,$$

and the *updating* equation at time t_{i+1} :

$$(8) \quad \pi(f, t_{i+1}) = \frac{\pi(r(y)f, t_{i+1}-)}{\pi(r(y), t_{i+1}-)},$$

where the mark at time t_{i+1} is assumed to be y .

2.2.2 A weak convergence theorem

To compute the infinite dimensional posterior measure-valued process characterized in Theorem 2.1, we need to approximate the infinite dimensional problem by a finite dimensional problem and to develop a recursive algorithm. One fundamental requirement for any algorithm is consistency, namely, the approximate posterior measure-valued process converges to the true one. The following weak convergence theorem provides a blueprint for constructing consistent recursive algorithms through Kushner's Markov chain approximation methods ([23]).

Following the literature of weak convergence, we use the notation, $X_\epsilon \Rightarrow X$, to mean X_ϵ converges weakly to X in the Skorohod topology as $\epsilon \rightarrow 0$. Let $(\theta_\epsilon, X_\epsilon)$ be an approximation of the signal, (θ, X) and let the observations for an approximate signal $(\theta_\epsilon, X_\epsilon)$ be $\Phi_\epsilon = \{(T_{i,\epsilon}, Y_{i,\epsilon})\}_{i \geq 1}$.

Theorem 2.2. *Suppose that (θ, X, Φ, V) on (Ω, \mathcal{F}, P) satisfies Assumptions 2.1 – 2.5. Suppose that for any $\epsilon > 0$, $(\theta_\epsilon, X_\epsilon, \Phi_\epsilon, V)$ on $(\Omega_\epsilon, \mathcal{F}_\epsilon, P_\epsilon)$ also satisfies Assumptions 2.1 – 2.5. If $(\theta_\epsilon, X_\epsilon) \Rightarrow (\theta, X)$ as $\epsilon \rightarrow 0$, then, for any bounded continuous functions, f ,*

- (1) $\Phi_\epsilon \Rightarrow \Phi$ under the physical measure P ; and
- (2) $\pi_\epsilon(f, \cdot) \Rightarrow \pi(f, \cdot)$, as $\epsilon \rightarrow 0$.

This theorem states that if the approximate signal converges to the true signal in distribution, then, (1) the observation of an approximate model converges to that of the true model in distribution, and more importantly, (2) the posterior measure-valued process of the approximate model converges to that of the true one in distribution.

2.2.3 A blueprint for developing recursive algorithms

Based on Theorem 2.2, we present a three-step blueprint for developing consistent recursive algorithms using Markov chain approximation method to compute the current joint posterior distribution and the Bayes estimates. After respectively discretizing the state and the mark spaces into two grids, the blueprint has three steps in developing a recursive algorithm. In Step 1, we construct $(\theta_\epsilon, X_\epsilon)$, a Markov chain approximation to (θ, X) , and approximate $r(y) = r(y|\theta, x)$ by $r_\epsilon(y) = r(y_\epsilon|\theta_\epsilon, x_\epsilon)$, where $(\theta_\epsilon, x_\epsilon)$ and y_ϵ are respectively confined to the grids of $(\theta_\epsilon, X_\epsilon)$ and \mathbb{Y}_ϵ . In Step 2, we apply Theorem 2.1 to get the approximate filter for $\pi_\epsilon(f, t)$ corresponding to $(\theta_\epsilon, X_\epsilon, Y_\epsilon, r_\epsilon)$, which can also be separated into:

$$(9) \quad \pi_\epsilon(f, t_{i+1}-) = \pi_\epsilon(f, t_i) + \int_{t_i}^{t_{i+1}-} \pi_\epsilon(\mathbf{A}_{v,\epsilon} f, s) ds,$$

and with the assumption that a mark at y_{i+1} happens at time t_{i+1} :

$$(10) \quad \pi_\epsilon(f, t_{i+1}) = \frac{\pi_\epsilon(f r_\epsilon(y_{i+1}), t_{i+1}-)}{\pi_\epsilon(r_\epsilon(y_{i+1}), t_{i+1}-)}.$$

In Step 3, we turn Equations (9) and (10) to a recursive algorithm for the state grid and in discrete times via two substeps: (a) turn $\pi_\epsilon(\cdot, t)$ into a finite array with the components being $\pi_\epsilon(f, t)$ for lattice-point indicators f and (b) approaches the time integral in (9) with a numerical scheme.

3. CONSISTENT IMPLICIT PARALLEL RECURSIVE ALGORITHMS

We construct well-designed recursive algorithms to produce real-time joint posterior and Bayes estimates. For notation simplicity, we exclude V , the other observable factors, in this section.

We separate $\theta(t)$ into time-invariant and time-variant parameters, namely, set $\theta(t) = (\theta, v(t))$ where θ and $v(t)$ can be vectors. This natural way to separate $\theta(t)$ has the advantage in computing parallelism and memory management as shown later. The goal is to compute π_t , the joint posterior distribution of $(\theta, v(t), X(t))$ given $\mathcal{F}_t^\Phi = \sigma((T_i, Y_i) : T_i \leq t)$, the observations up to time t . Suitable priors such as uninformative priors can be chosen as in [33]. Then, $\pi(f, t) = E^P[f(\theta, v(t), X(t)) | \mathcal{F}_t^\Phi] = \int f(\theta, v, x) \pi_t(d(\theta, v, x))$. Just as we approximate a continuous distribution by a histogram, we approximate π_t by $\pi_{\epsilon, t}$, a discretized approximation. The corresponding $\pi_\epsilon(f, t)$ is defined by $\pi_\epsilon(f, t) = \sum_{\theta, v, x} f(\theta, v, x) \pi_{\epsilon, t}(\theta, v, x)$ where the sum is over the state grid described below.

In preparation for constructing the recursive algorithm, we discretize the state space of $(\theta, v(t), X(t))$ as a state grid, \mathcal{G} . The total number of grid point in \mathcal{G} is $N_\theta \times N_v \times N_x$ where N_θ is the number of grid point for the time-invariant parameter space of θ and similarly for N_v and N_x . For a time-invariant parameter space such as θ , the discretized parameter space is a natural approximation. However, Markov chain is a natural approximation for stochastic processes $(v(t), X(t))$.

We can order the parameter grid points in the grid of θ in some way and let θ_j be a parameter lattice point. For each fixed θ_j , let \mathcal{G}_j be the corresponding grid of $(v(t), X(t))$. Then, we can partition \mathcal{G} by \mathcal{G}_j , namely, $\mathcal{G} = \cup_{j=1}^{N_\theta} \mathcal{G}_j$. Of course, if there is no time-variant parameter, $(v(t), X(t))$ is reduced to $X(t)$ only.

3.1 Recursive algorithms

Below, we follow the three steps in Section 2.2.3.

Step 1: Construct $(\theta_\epsilon, v_\epsilon(t), X_\epsilon(t))$, a Markov chain approximation to $(\theta, v(t), X(t))$. This can be transformed to construct a Markov chain generator, \mathbf{A}_ϵ , such that $\mathbf{A}_\epsilon \rightarrow \mathbf{A}$ pointwise as $\epsilon \rightarrow 0$. Then, Corollary 4.8.5 on page 230 of [14] implies $(\theta_\epsilon, v_\epsilon(t), X_\epsilon(t)) \Rightarrow (\theta, v(t), X(t))$. We give three typical examples, diffusion, jump-diffusion and stochastic volatility, on how to construct approximate Markov chain generators.

Example 3.1. GBM Model This is the example used in Section 4. X_t follows a SDE:

$$\frac{dX_t}{X_t} = \mu dt + \sigma dB_t,$$

where B_t is a standard Brownian motion. Let $\theta = (\mu, \sigma, \rho)$. The generator associated with the martingale problem in Assumption 2.1 for X is

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

The generator has the first and second order derivatives. Hence, we employ finite difference method, namely, the central difference approximation to the first and second order differentials so as to construct the approximate generator as below:

$$\begin{aligned} & \mathbf{A}_{1, \epsilon} f(\theta, x) \\ &= \mu x \left(\frac{f(\theta, x + \epsilon_x) - f(\theta, x - \epsilon_x)}{2\epsilon_x} \right) \\ (11) \quad &+ \frac{1}{2} \sigma^2 x^2 \left(\frac{f(\theta, x + \epsilon_x) + f(\theta, x - \epsilon_x) - 2f(\theta, x)}{\epsilon_x^2} \right) \\ &= a_x(\theta, x) (f(\theta, x + \epsilon_x) - f(\theta, x)) \\ &+ b_x(\theta, x) (f(\theta, x - \epsilon_x) - f(\theta, x)) \end{aligned}$$

where

$$a_x(\theta, x) = \frac{1}{2} \left(\frac{\mu x}{\epsilon_x} + \frac{\sigma^2 x^2}{\epsilon_x^2} \right),$$

and

$$b_x(\theta, x) = \frac{1}{2} \left(-\frac{\mu x}{\epsilon_x} + \frac{\sigma^2 x^2}{\epsilon_x^2} \right).$$

Example 3.2. Jump-GBM Model popularized by [26]. $X(t)$ follows

$$\frac{dX(t)}{X(t)} = \mu dt + \sigma dB(t) + (J_{N(t)} - 1) dN(t),$$

where $N(t)$ is a Poisson process independent of $B(t)$ and has intensity λ . Representing the jump magnitudes, $\{J_i\}$ are i.i.d. lognormal, namely, $\log J_i$ are i.i.d. $N(\mu_J, \sigma_J)$. The generator is

$$\mathbf{A}_2 f(\theta, x) = \mathbf{A}_1 f(\theta, x) + \lambda \int [f(xz) - f(x)] q_J(z) dz,$$

where \mathbf{A}_1 is given in Example 3.1, $q_J(z)$ is the lognormal density and θ includes $(\mu, \sigma, \mu_J, \sigma_J)$. The generator further contains integration, which can be approximated by proper summation. The approximate Markov chain generator is given by

$$\begin{aligned} & \mathbf{A}_{2, \epsilon} f(\theta, x) = \mathbf{A}_{1, \epsilon} f(\theta, x) \\ &+ \lambda \sum_z [f(\theta, z) - f(\theta, x)] \hat{q}_J(z, x), \end{aligned}$$

where $\mathbf{A}_{1,\varepsilon}f(\theta, x)$ is given in (11), the summation is over all grid points of X , and $\hat{q}_J(z, x) = P\{J \in [(z - 0.5\epsilon_x)/x, (z + 0.5\epsilon_x)/x]\}$.

Example 3.3. Heston Stochastic Volatility Model proposed in [19] contains time variant stochastic volatility $v(t) = \sigma^2(t)$. $(v(t), X(t))$ follows

$$(12) \quad \begin{aligned} \frac{dX_t}{X_t} &= \mu dt + \sigma_t dB_t \\ dv_t &= \kappa(\alpha - v_t)dt + \gamma\sqrt{v_t}dW_t \end{aligned}$$

where (B_t, W_t) are independent standard Brownian motions. Let θ including $(\mu, \kappa, \alpha, \gamma)$ be a vector of the parameters in the model. The corresponding generator \mathbf{A} in Assumption (2.1) is given by

$$(13) \quad \begin{aligned} \mathbf{A}_3 f(\theta, v, x) &= \mu x \frac{\partial f}{\partial x}(\theta, v, x) + \frac{1}{2} v x^2 \frac{\partial^2 f}{\partial x^2}(\theta, v, x) \\ &+ \kappa(\alpha - v) \frac{\partial f}{\partial v}(\theta, v, x) + \frac{1}{2} \gamma^2 v \frac{\partial^2 f}{\partial v^2}(\theta, v, x) \end{aligned}$$

The generator can similarly be approximated by finite central difference.

$$(14) \quad \begin{aligned} \mathbf{A}_{3,\varepsilon} f(\theta, v, x) &= a_x(\theta, v, x) (f(\theta, v, x + \epsilon_x) - f(\theta, v, x)) \\ &+ b_x(\theta, v, x) (f(\theta, v, x - \epsilon_x) - f(\theta, v, x)) \\ &+ a_v(\theta, v, x) (f(\theta, v + \epsilon_v, x) - f(\theta, v, x)) \\ &+ b_v(\theta, v, x) (f(\theta, v - \epsilon_v, x) - f(\theta, v, x)) \end{aligned}$$

where

$$\begin{aligned} a_x(\theta, v, x) &= \frac{1}{2} \left(\frac{\mu x}{\epsilon_x} + \frac{v x^2}{\epsilon_x^2} \right), \\ b_x(\theta, v, x) &= \frac{1}{2} \left(-\frac{\mu x}{\epsilon_x} + \frac{v x^2}{\epsilon_x^2} \right), \end{aligned}$$

and

$$\begin{aligned} a_v(\theta, v, x) &= \frac{1}{2} \left(\frac{\kappa(\alpha - v)}{\epsilon_v} + \frac{\gamma^2 v}{\epsilon_v^2} \right), \\ b_v(\theta, v, x) &= \frac{1}{2} \left(-\frac{\kappa(\alpha - v)}{\epsilon_v} + \frac{\gamma^2 v}{\epsilon_v^2} \right). \end{aligned}$$

Step 2: Obtain the filtering equation for $\pi_\varepsilon(f, t)$, which are (9) and (10). Note that $r_\varepsilon(y_{i+1}) = p_\varepsilon(y_{i+1}|X(t_{i+1}); \theta)$, the approximate transition probability in Representation I. Note that $p(y|x)$ specifies the noise, which may include additive, rounding and biasing noise.

Step 3: Convert Equations (9) and (10) to recursive algorithms with two more substeps. A recursive algorithm computes the approximate posterior masses at all grid points. Let the approximate posterior mass at grid point (θ_j, v_m, x_l) be

$$\begin{aligned} p_\varepsilon(\theta_j, v_m, x_l; t) \\ = P\{\theta_\varepsilon = \theta_j, v_\varepsilon(t) = v_m, X_\varepsilon(t) = x_l | \mathcal{F}_t^\Phi\}. \end{aligned}$$

In the first substep, we take f as the following indicator function.

$$(15) \quad \mathbf{I}_{\{\theta_j, v_m, x_l\}}(\theta_\varepsilon, v_\varepsilon(t), X_\varepsilon(t))$$

which takes value one only when $(\theta_\varepsilon = \theta_j, v_\varepsilon(t) = v_m, X_\varepsilon(t) = x_l)$ and zero otherwise. Then, the following facts emerge: $\pi_\varepsilon(\mathbf{I}(\theta_j, v_m, x_l), t) = p_\varepsilon(\theta_j, v_m, x_l; t)$, $\pi_\varepsilon(\mathbf{I}(\theta_j, v_m, x_l)r_\varepsilon, t) = p_\varepsilon(\theta_j, v_m, x_l; t)r_\varepsilon(y_i|x_l; \theta_j)$, and $\pi_\varepsilon(r_\varepsilon, t) = \sum_{j', m', l'} p_\varepsilon(\theta_{j'}, v_{m'}, x_{l'}; t)r_\varepsilon(y_i|x_{l'}; \theta_{j'})$, where the summation is over the whole state grid.

Hence, the propagation Equation (9) becomes,

$$(16) \quad \begin{aligned} p_\varepsilon(\theta_j, v_m, x_l; t_{i+1}-) &= p_\varepsilon(\theta_j, v_m, x_l; t_i) \\ &+ \int_{t_i}^{t_{i+1}-} \pi_\varepsilon(\mathbf{A}_\varepsilon \mathbf{I}, s) ds, \end{aligned}$$

where $\pi_\varepsilon(\mathbf{A}_\varepsilon \mathbf{I}, s)$ depends on $\mathbf{A}_\varepsilon \mathbf{I}$, which is a linear combination of some indicator functions according to the three given examples. Therefore, $\pi_\varepsilon(\mathbf{A}_\varepsilon \mathbf{I}, s)$ is a linear combination of $p_\varepsilon(\theta_j, \cdot, \cdot; s)$. We assume the posteriors are zero outside the state grid.

The updating Equation (10) can be written as,

$$(17) \quad \begin{aligned} p_\varepsilon(\theta_j, v_m, x_l; t_{i+1}) \\ = \frac{p_\varepsilon(\theta_j, v_m, x_l; t_{i+1}-) r_\varepsilon(y_{i+1}|x_l, \theta_j)}{\sum_{j', m', l'} p_\varepsilon(\theta_{j'}, v_{m'}, x_{l'}; t) r_\varepsilon(y_{i+1}|x_{l'}; \theta_{j'})}. \end{aligned}$$

The second substep is to approximate the time integral in (16) via a numerical method. This is where the explicit and implicit recursive algorithms differ.

3.1.1 Explicit recursive algorithms

Explicit Euler scheme is the most straightforward one. We can choose LL , a controller of time-step size. If $t_{i+1} - t_i \leq LL$, then an explicit algorithm approximate (16) by

$$(18) \quad \begin{aligned} p_\varepsilon(\theta_j, v_m, x_l; t_{i+1}-) &\approx p_\varepsilon(\theta_j, v_m, x_l; t_i) \\ &+ \pi_\varepsilon(\mathbf{A}_\varepsilon \mathbf{I}, t_i)(t_{i+1} - t_i). \end{aligned}$$

It is clear that $p_\varepsilon(\theta_j, v_m, x_l; t_{i+1}-)$ at time $t_{i+1}-$ is directly computed by a linear combination of $p_\varepsilon(\theta_j, v_m, x_l; t_i)$ at time t_i . Then, it can roll over to the next time step.

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 approximate $p_\varepsilon(\theta_j, v_m, x_l; t_{i+1}-)$ by repeatedly applying the recursive algorithm given by Equation (18) from $t_{i,0}$ to $t_{i,1}$, then $t_{i,2}, \dots$, until $t_{i,n} = t_{i+1}$.

Equations (17) and (18) constitute the explicit recursive algorithm for calculating the posterior over time. The explicit recursive works well in previous studies ([33] and [34]) which the tick size of stock price was \$1/8 or \$ 1/16. However, when the tick size of a stock price becomes 0.01 dollar and the volatility is high (say, 30% annually), the explicit

recursive algorithms can become very inefficient as demonstrated in the next section. The reason is that the positivity constraint on $p_\epsilon(\theta_j, v_m, x_l)$ causes the shrinkage of the time step to zero. This is the curse of time-step size.

3.1.2 Implicit recursive algorithms

In order to cure this curse, we adopt the implicit method to approximate (16) and develop *implicit recursive algorithms* in this paper.

For $t_{i+1} - t_i \leq LL$, we instead approximate (16) by

$$(19) \quad p_\epsilon(\theta_j, v_m, x_l; t_{i+1}-) \approx p_\epsilon(\theta_j, v_m, x_l; t_i) + \pi_\epsilon(\mathbf{A}_\epsilon \mathbf{I}, t_{i+1}-)(t_{i+1} - t_i).$$

Now, for each fixed θ_j , we treat (19) as a $N_v \times N_x$ linear system with $N_v \times N_x$ unknown $p_\epsilon(\theta_j, v_m, x_l; t_{i+1}-)$, given $p_\epsilon(\theta_j, v_m, x_l; t_i)$. Through various matrix iterative techniques depending on the generator, we solve $p_\epsilon(\theta_j, v_m, x_l; t_{i+1}-)$ from the linear system. Moreover, the coefficient structure usually ensures the positivity of the solutions as demonstrated in the example of the next section.

Equations (17) and (19) constitute a class of implicit recursive algorithms to calculate the posterior.

3.2 Memory management and computing parallelism

We observe from (16) that the propagation can be carried out for each fixed θ_j , in either explicit or implicit recursive algorithms. Namely, we can *parallelize the propagation* in a way that each CPU can propagate $p_\epsilon(\theta_j, \cdot, \cdot; t)$ on $(v(t), X(t))$ for each θ_j . Only in computing the denominator in the updating equation (17), we need to *aggregate all* $p_\epsilon(\theta_{j'}, v_{m'}, x_{l'}; t)p_\epsilon(y_{i+1}|x_{l'}; \theta_{j'})$ from all CPUs. After obtaining the grand total, we can *parallelize the updating* in a way that each CPU updates its share of the joint posterior distribution of $(\theta, v(t), X(t))$. Then repeat the procedure as new observations flow in.

To achieve such parallel computing, we, at the beginning, partition the joint prior distribution of $(\theta_\epsilon, v_\epsilon(0), X_\epsilon(0))$ on the whole state grid according to each θ_j . Namely, we choose N_θ CPUs, each CPU's memory stores $N_v \times N_x$ number of $p_\epsilon(\theta_j, \cdot, \cdot; 0)$ of \mathcal{G}_j corresponding to θ_j . Hence, with proper memory management, the recursive algorithms' parallelism can be carried out nicely.

3.3 Consistency

Two approximations are employed in the construction of the above recursive algorithms. One is the approximation of the time integral in (16) by an explicit or implicit method, whose convergence is well-established. The other more important one is the approximation of the filtering equation in (7) and (8) of the true model by that in (9) and (10) for the approximated model. The weak convergence result (2) of Theorem 2.2 warrants the consistency of the second approximation.

Corresponding to a recursive algorithm, let

$$p(\theta_j, v_m, x_l; t) = \pi_t\{\theta \in N_{\theta_j}, v(t) \in N_{v_m}, X(t) \in N_{x_l}\},$$

where N_{θ_j} , N_{v_m} and N_{x_l} are appropriate neighborhoods. For example, $N_{x_w} = (x_w - \frac{1}{2}\epsilon_x, x_w + \frac{1}{2}\epsilon_x)$. Then, Theorem 2.2 implies that $p_\epsilon(\theta_j, v_m, x_l; \cdot) \Rightarrow p(\theta_j, v_m, x_l; \cdot)$ for all (θ_j, v_m, x_l) in the state grid as $\epsilon \rightarrow 0$.

Both explicit and implicit recursive algorithms are consistent and parallelizable with proper memory management. The computational advantage of the implicit recursive algorithms is demonstrated by an illustrating example in the next section.

4. AN UHF STOCK PRICE EXAMPLE

We construct a partially-observed GBM model with tick size \$1/100, and develop explicit and implicit recursive algorithms. We identify the cause for the deficiency of the explicit recursive algorithm and prove the nonnegativity of the implicit algorithm. We use simulated data to demonstrate the efficiency of the implicit algorithm and apply it to a real stock transaction price data set.

4.1 The GBM model sampled at random times with trading noise

The model for UHF stock price at tick size \$1/100 is presented in Representation I as it is intuitive. The state process becomes the intrinsic value process of a stock.

We assume the intrinsic value process follows a GBM given in Example 3.1. We assume the trade durations follow an exponential ACD or Weibull ACD model. Because such an ACD model has an $\mathcal{F}_t^{\Phi, V}$ -predictable stochastic intensity, we can use the simplified version of the normalized filtering equation, (7) and (8).

Trading noise enters into price at trading times. Following [33], we consider three sources of noise: non-clustering noise, price discreteness and clustering noise. To accommodate price discreteness, the mark space becomes a discrete space $\mathbf{Y} = \{\frac{a+1}{M}, \frac{a+2}{M}, \dots, \frac{a+m}{M}\}$, where $M = 100$ is set by market convention for stock. In addition, market trades tend to occur on coarse ticks such as 5 or 10 cents leading to price clustering. Finally, we assume that nonclustering noise includes all other unspecified noise in the market.

At trade time t_i , let $x = X(t_i)$, $y = Y(t_i)$, and $y' = Y'(t_i) = R[X(t_i) + U_i, \frac{1}{M}]$, where U_i is the non-clustering noise. We develop a random transformation $y = F(x)$ in three steps and point out a way to calculate the corresponding transition probability $p(y|x)$.

Step (a): Add non-clustering noise U ; $x' = x + U$, where U is the non-clustering noise at trade i . We assume $\{U_i\}$ are

independent of $X(t)$, *i.i.d.*, with a doubly geometric distribution:

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

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

Step (c): Accommodate clustering noise by biasing y' through a random biasing function $b_i(\cdot)$ at trade i . $\{b_i(\cdot)\}$ is assumed conditionally independent given $\{y'_i\}$. To be coherent with the stock price data analyzed, we develop a simple random biasing function for the tick of cent or \$1/100. The data to be studied possesses the clustering occurrence: multiples of ten cents are most likely; odd multiples of five cents are the second most likely; the others are least likely and each one has about the same frequencies.

To accommodate such clustering, a biasing function $b(\cdot)$ is developed according to these biasing rules: if the fractional part of y' is a multiple of five cents, then y stays on y' with probability one; if the fractional part of y' is not a multiple of five cents, then y stays on y' with probability $1 - \alpha - \beta$; y moves to the closest odd multiple of five cents with probability α , and moves to the closest multiple of ten cents with probability β .

In short, for coherence with market convention in a stock market, we develop a random transformation F for ticks with $M = 100$:

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

where F is specified by $p(y|x)$, which can be calculated using the above construction and the formula, $p(y|x) = \sum_{y'} p(y|y')p(y'|x)$. And the parameters related to clustering noise (α, β) can be estimated through the method of relative frequency.

4.2 An explicit recursive algorithm, its deficiency and the cause

The model above has time-varying $v(t)$ and θ becomes (μ, σ, ρ) . Let N_μ be the number of the grid for the parameter space of μ and similarly for N_σ and N_ρ . Then, $N_\theta = N_\mu \times N_\sigma \times N_\rho$. We assign uniform priors for (μ, σ, ρ) and set $P\{X(0) = Y(t_1)\} = 1$ where $Y(t_1)$ is the first observed price. Then, a reasonable prior is given below:

$$p(\theta_j, x_l; 0) = \begin{cases} \frac{1}{N_\mu N_\sigma N_\rho} & \text{if } x_l = Y(t_1) \text{ for all } \theta_j \\ 0 & \text{otherwise} \end{cases}.$$

Following the general approach in Section 3, for θ_j (a parameter grid point), the approximate generator is given by (11). Taking $f(\theta_\epsilon, X_\epsilon(t)) = \mathbf{I}_{\{\theta_j, x_l\}}(\theta_\epsilon, X_\epsilon(t))$, $\pi_\epsilon(\mathbf{A}_\epsilon \mathbf{I}, t)$ becomes

$$(20) \quad \begin{aligned} & \pi_\epsilon(\mathbf{A}_\epsilon \mathbf{I}, t) \\ &= a_x(\theta_j, x_l)(p_\epsilon(\theta_j, x_{l-1}; t) - p_\epsilon(\theta_j, x_l; t)) \\ & \quad + b_x(\theta_j, x_l)(p_\epsilon(\theta_j, x_{l+1}; t) - p_\epsilon(\theta_j, x_l; t)). \end{aligned}$$

Then, when $\Delta t = t_{i+1} - t_i \leq LL$, the explicit recursive algorithm for the model becomes

$$(21) \quad \begin{aligned} & p_\epsilon(\theta_j, x_l; t_{i+1}-) \\ & \approx p_\epsilon(\theta_j, x_l; t_i) + \left[a_x(\theta_j, x_l)p_\epsilon(\theta_j, x_{l-1}; t_i) \right. \\ & \quad \left. - (a_x(\theta_j, x_l) + b_x(\theta_j, x_l))p_\epsilon(\theta_j, x_l; t_i) \right. \\ & \quad \left. + b_x(\theta_j, x_l)p_\epsilon(\theta_j, x_{l+1}; t_i) \right] \Delta t, \end{aligned}$$

and

$$(22) \quad \begin{aligned} & p_\epsilon(\theta_j, x_l; t_{i+1}) \\ &= \frac{p_\epsilon(\theta_j, x_l; t_{i+1}-)p(y_{i+1}|x_l, \theta_j)}{\sum_{j', l'} p_\epsilon(\theta_{j'}, x_{l'}; t_{i+1}-)p_\epsilon(y_{i+1}|x_{l'}; \theta_{j'})}. \end{aligned}$$

4.2.1 The deficiency and the cause

Probabilities must be positive. From (21), one necessary condition to keep positive probabilities is that for all θ_j and x_l ,

$$(a_x(\theta_j, x_l) + b_x(\theta_j, x_l))\Delta t \leq 1,$$

or equivalently,

$$\Delta t \leq \max_{\theta_j, x_l} \frac{1}{a_x(\theta_j, x_l) + b_x(\theta_j, x_l)},$$

where the maximum is taken over the whole state grid.

When the tick size is 1/8, $a_x(\theta_j, x_l)$ and $b_x(\theta_j, x_l)$ are small and Δt can be reasonably large. For example, using commonly seen values from the stock market, we take $x = 105$, expected growth 10% annually, or $\mu = 0.1/(252 \times 6.5 \times 60 \times 60) = 1.696e - 8$ in second assuming 252 business days per year with 6.5 trading hours (9:30am - 4pm) per day, and volatility 60% annually, or $\sigma = 0.6/\sqrt{252 \times 6.5 \times 60 \times 60} = 2.471e - 4$. Then, we can compute $a_x(\theta, x) = 0.345$ and $\Delta t \leq 1/(2a_x(\theta, x)) = 1.45$ seconds. Setting $LL = 1$, the explicit method has no nonnegativity issue and the computer program works fast enough to produce real time Bayes estimates. However, when tick size becomes 1/100, $a_x(\theta_j, x_l) = 53.85$ and $\Delta t \leq 1/(2a_x(\theta, x)) = 0.00929$. In the simulation study presented later, we assume σ to be 30% annually. However, we don't know the true value and we can guess the maximum of σ to be as large as 60%, double the true value. When we experiment the size of LL with our computer program, we find that LL needs to be set at 0.005 to avoid nonnegativity issue. This imposes a too severe shrinkage to the size of Δt and the explicit recursive algorithm requires too many steps (beyond real time) to compute the joint posterior before next trade. This is the cause of deficiency of the explicit recursive algorithm. The implicit recursive algorithm developed below can resolve such a computation deficiency.

Table 1. Time Comparison of the Explicit and Implicit Recursive Algorithms

Volatility	NDATA = 100				NDATA = 1,000			
	Explicit		Implicit		Explicit		Implicit	
	LL	Time	LL	Time	LL	Time	LL	Time
10%	.05s	353s	1s	199s	.05s	4,503s	1s	2,310s
20%	.01s	1,979s	1s	187s	.01s	19,787s	1s	2,228s
30%	.005s	3,858s	1s	191s	.005s	41,440s	1s	1,939s

Table 2. Summary Statistics for AA Stock Prices, 11/07/2005 – 11/10/2005

AA stock	# trades	Min	Max	Median	Mean	Std	Skewness	Kurtosis
	28900	25.33	26.75	26.12	26.086	0.3393	-0.1387	2.1652

Table 3. Final Marginal Posterior of μ , σ , and ρ for AA Stock

μ	$p(\mu)$	σ	$p(\sigma)$	ρ	$p(\rho)$
-4.0e-7	0.0070	7.8e-5	1.1e-6	0.1100	1.2e-8
-2.0e-7	0.0405	7.9e-5	0.0002	0.1125	7.2e-6
0.0	0.1349	8.0e-5	0.0095	0.1150	0.0009
2.0e-7	0.2589	8.1e-5	0.1158	0.1175	0.0289
4.0e-7	0.2877	8.2e-5	0.3881	0.1200	0.2252
6.0e-7	0.1851	8.3e-5	0.3716	0.1225	0.4555
8.0e-7	0.0690	8.40e-5	0.1053	0.1250	0.2493
10.0e-7	0.0149	8.5e-5	0.0091	0.1275	0.0382
12.0e-7	0.0019	8.6e-5	0.0002	0.1300	0.0017
14.0e-7	0.0001	8.7e-5	2.2e-6	0.1325	2.2e-5

$$E(\mu) = 3.395e - 7 \text{ and } SE(\mu) = 2.693e - 7, \text{ or annually, } 200.20\% \text{ with SE } 158.80\%;$$

$$E(\sigma) = 8.247e - 5 \text{ and } SE(\sigma) = 8.996e - 7, \text{ or annually, } 20.03\% \text{ with SE } 0.22\%;$$

$$E(\rho) = 0.1226 \text{ and } SE(\rho) = 0.0022$$

is clear, and can be over 20 times when the volatility is as high as 30%, which is not uncommon in stock market.

4.5 A real data application

We use the implicit recursive algorithm to examine tick transaction data for the AA stock for the period of November 7–10, 2005. The data was from the TAQ database issued by NYSE. Table 2 reports summary statistics for the complete set of 28,900 trades in our data. Based on the relative frequencies for observed tick values and recorded trades, the estimated clustering parameters are $\alpha = .0113$ and $\beta = 0.0353$.

Even though the tick size is \$1/100, the algorithm is fast enough to provide real-time estimates. The final posteriors of μ , σ , and ρ are presented in Table 3. All the final marginal posteriors are unimodal and bell-shaped. The Bayes estimates and their standard errors are given below the table. Because the AA stock price began at \$ 25.49 on November 7, 2005 and ended at \$26.31 at the end of November 10, 2005, the Bayes estimates of μ is large (200.20%). But we also note that the standard error of μ is also large (158.80%), which implies that the expected return is not statistically different from zero. The standard error of μ is much larger than those of σ and ρ (comparing to the Bayes estimates) and this makes sense because μ is a trend parameter, and its

estimation accuracy depends on the length of time covered by the data (4 days); but the accuracy of σ and ρ estimates may increase as the frequency of sampling increases.

5. CONCLUSIONS

For a general nonlinear filtering framework for UHF data that uses MPP observations with confounding observable factors, we develop implicit recursive algorithms for the Bayes estimation via filtering equation. The implicit recursive algorithms mitigates the curse of time-step shrinkage in explicit recursive algorithms. This is demonstrated in a partially-observed GBM model, where the nonnegativity of the implicit recursive algorithm is proven. The algorithm is applied to a real stock transactions price data set to produce real-time Bayes estimates.

The consistent implicit parallel recursive algorithms proposed here can possibly harness the newly available green low-cost GPU supercomputing power, and be used for more complicated and important benchmark models of Stochastic Volatility (SV) models so as to produce real-time stochastic volatility estimates. This is under investigation and [6] is one work in this direction. The real time Bayes estimation of another important class of models, exponential Lévy models, can be studied also.

APPENDIX A. RELATED RESULTS AND THE PROOF OF NONNEGATIVITY

Before we prove the nonnegativity result in Proposition 4.1, we present two related concepts and three propositions, whose proofs can be found in [32]. Then, we prove one theorem, implying Proposition 4.1.

Definition A.1. M -matrices are real nonsingular matrices C such that $c_{ij} \leq 0$ for all $i \neq j$ and $C^{-1} \geq 0$ (each entry of C^{-1} is nonnegative).

Proposition A.1. *Let C be a real square matrix. Then the eigenvalues of C are the same as the eigenvalues of C^T .*

Proposition A.2. (Gerschgorin) *Let C be an $n \times n$ real matrix, and let*

$$r_i = \sum_{j=1, j \neq i}^n |c_{ij}|, \quad 1 \leq i \leq n.$$

Then, all the eigenvalues of λ of C lie in the union of the n Gerschgorin disks

$$|z - c_{ii}| \leq r_i, \quad 1 \leq i \leq n.$$

Proposition A.3. *All matrices with nonpositive off-diagonal entries whose spectrums are in the right-hand half-plane are M -matrices.*

Theorem A.1. *The matrix B is an M -matrix, and hence $B^{-1} \geq 0$.*

Proof. Note that regardless the size of Δt , the off-diagonal entries of B are nonpositive. Furthermore, the set of the union of all the Gerschgorin discs of B^T are in the right-hand halfplane, in fact, in the set $\{z = x + iy : x \geq 1\}$. So by Propositions A.1, A.2, and A.3, we see that B is an M -matrix, and hence $B^{-1} \geq 0$. \square

Remark A.1. Since $B^{-1} \geq 0$ and $p(\theta_j, i) \geq 0$, $p(\theta_j, i + 1) = B^{-1}p(\theta_j, i)$ implies $p(\theta_j, i + 1) \geq 0$, which is exactly Proposition 4.1.

Received 24 July 2013

REFERENCES

- [1] AÏT-SAHALIA, Y. (1996). Nonparametric pricing of interest rate derivative securities. *Econometrica* **64** 527–560.
- [2] AÏT-SAHALIA, Y., CACHO-DIAZ, J., and LAEVEN, R. J. A. (2013). Modeling financial contagion using mutually exciting jump processes. Working Paper, Princeton University.
- [3] AÏT-SAHALIA, Y. and MYKLAND, P. A. (2008). An analysis of Hansen-Scheinkman estimators for randomly sampled diffusions. *Journal of Econometrics* **144** 1–26. [MR2439920](#)
- [4] BANDI, F. M. and RUSSELL, J. R. (2006). Separating microstructure noise from volatility. *Journal of Financial Economics* **79** 655–692.
- [5] BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A., and SHEPHARD, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica* **76** 1481–1536. [MR2468558](#)
- [6] BUNDICK, B. and ZENG, Y. (2013). Real-time stochastic volatility estimation via filtering for a partially-observed Heston model. Working Paper, University of Missouri at Kansas City.
- [7] CHEN, X., HANSEN, L. P., and CARRASCO, M. (2010). Nonlinearity and temporal dependence. *Journal of Econometrics* **146** 155–169. [MR2607192](#)
- [8] CVITANIC, J., LIPTSER, R., and ROZOVSKY, B. (2006). A filtering approach to tracking volatility from prices observed at random times. *Annals of Applied Probability* **16** 1633–1652. [MR2260076](#)
- [9] DUFFIE, D. and GLYNN, P. (2004). Estimation of continuous-time markov processes sampled at random time intervals. *Econometrica* **72** 1773–1808. [MR2095532](#)
- [10] ELLIOTT, R. and MALCOLM, W. P. (2005). Nonlinear filtering with counting observations. *IEEE Transactions on Automatic Control* **50** 1123–1134. [MR2156041](#)
- [11] ENGLE, R. (2000). The econometrics of ultra-high-frequency data. *Econometrica* **68** 1–22.
- [12] ENGLE, R. and RUSSELL, J. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica* **66** 1127–1162. [MR1639411](#)
- [13] ENGLE, R. and RUSSELL, J. (2009). Analysis of high frequency data. In: *Handbook of Financial Econometrics*, Y. Aït-Sahalia and L. P. Hansen, eds, North-Holland, Amsterdam.
- [14] ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence*. Wiley, New York. [MR0838085](#)
- [15] FAN, J. and WANG, Y. (2007). Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of American Statistical Association* **102** 1349–1362. [MR2372538](#)
- [16] FREY, R. and RUNGALDIER, W. J. (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. [MR1831267](#)
- [17] HANSEN, L. P. and SCHEINKMAN, J. (1995). Back to the future: Generating moment implications for continuous-time markov processes. *Econometrica* **63** 767–804. [MR1343081](#)
- [18] HASBROUCK, J. (2007). *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading*. Oxford University Press, USA.
- [19] HESTON, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies* **6** 327–343.
- [20] HU, G. X., KUIPERS, D., and ZENG, Y. (2013). Bayesian inference via filtering equation for ultra-high frequency data (I): Model and estimation. Under review.
- [21] HU, G. X., KUIPERS, D., and ZENG, Y. (2013). Bayesian inference via filtering equation for ultra-high frequency data (II): Hypotheses testing and model selection. Under review.
- [22] KOURITZIN, M. and ZENG, Y. (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. [MR2122692](#)
- [23] KUSHNER, H. J. and DUPUIS, P. G. (2001). *Numerical Methods for Stochastic Control Problems in Continuous Time*. Springer-Verlag, New York, second edition. [MR1800098](#)
- [24] LARGE, J. (2007). Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets* **10** 1–25.
- [25] LI, Y. and MYKLAND, P. A. (2007). Are volatility estimators robust with respect to modeling assumptions? *Bernoulli* **13** 601–622. [MR2348742](#)
- [26] MERTON, R. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics* **3** 125–144.
- [27] PACURAR, M. (2008). Autoregressive conditional duration models in finance: A survey of the theoretical and empirical literature. *Journal of Economic Surveys* **22** 711–751.
- [28] SCOTT, L. C. (2006). Bayesian inference via filtering of micro-movement multivariate stock price models with discrete noises. Doctoral Dissertation, University of Missouri at Kansas City, Mathematics and Statistics Department (Unpublished). [MR2708675](#)

- [29] SCOTT, L. C. and ZENG, Y. (2008). A class of multivariate micro-movement models of asset price and their Bayesian model selection via filtering. In: *Markov Processes and Related Topics – A Festschrift for Thomas G. Kurtz*, S. Ethier J. Feng and R. Stockbridge, eds, Vol. 4, IMS Collection, Beachwood Ohio, pp. 123–136. [MR2574228](#)
- [30] SPALDING, R., TSUI, K. W., and ZENG, Y. (2006). A micro-movement model with bayes estimation via filtering: Applications to measuring trading noises and trading cost. *Nonlinear Analysis: Theory, Methods and Applications* **64** 295–309. [MR2188460](#)
- [31] STRIKWERDA, J. C. (1989). *Finite Difference Schemes and Partial Differential Equations*. Wadsworth and Brooks/Cole, Pacific Grove, California. [MR1005330](#)
- [32] VARGA, R. S. (2000). *Matrix Iterative Analysis*. Springer, New York, 2 edition. [MR1753713](#)
- [33] ZENG, Y. (2003). A partially-observed model for micro-movement of asset prices with Bayes estimation via filtering. *Mathematical Finance* **13** 411–444. [MR1995285](#)
- [34] ZENG, Y. (2004). Estimating stochastic volatility via filtering for the micro-movement of asset prices. *IEEE Transactions on Automatic Control* **49** 338–348. [MR2062247](#)
- [35] 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. [MR2177317](#)
- [36] ZENG, Y. and SCOTT, L. C. (2002). Bayes estimation via filtering equation for O-U process with discrete noises: Application to the micro-movement of stock prices. In: *Stochastic Theory and Control*, B. Pasik-Duncan, ed., Lecture Notes in Control and Information Sciences, Vol. 280, Springer-Verlag, Berlin Heidelberg, pp. 533–548. [MR1931675](#)
- [37] ZHANG, L. (2006). Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli* **12** 1019–1043. [MR2274854](#)
- [38] ZHANG, L., MYKLAND, P. A., and AÏT-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. [MR2236450](#)

Brent Bundick
 Department of Economics
 Boston College, Boston, MA
 USA
 E-mail address: bundickb@bc.edu

Noah Rhee
 Department of Mathematics and Statistics
 University of Missouri at Kansas City
 Kansas City, MO 64110
 USA
 E-mail address: rheen@umkc.edu

Yong Zeng
 Department of Mathematics and Statistics
 University of Missouri at Kansas City
 Kansas City, MO 64110
 USA
 E-mail address: zengy@umkc.edu