SIAM/ASA J. UNCERTAINTY QUANTIFICATION
Vol. xx, pp. x

Bayesian Inference via Filtering Equations for Ultra-High Frequency Data (I): Model and Estimation

Grace X. Hu*, David R. Kuipers †, and Yong Zeng ‡

Abstract. We propose a general partially-observed framework of Markov processes with marked point process observations for ultra-high frequency (UHF) data. The model well fits the stylized facts of UHF data in both macro- and micro-movements, subsumes important existing models, and incorporates the influence of other observable economic or market factors. We develop the corresponding Bayes estimation via filtering equation to quantify parameter uncertainty. Namely, we derive the normalized filtering equation to characterize the evolution of the posterior distribution, present a weak convergence theorem, and construct consistent, easily-parallelizable, recursive algorithms to calculate the joint posteriors and the Bayes estimates for streaming UHF data. Moreover, an easy-to-check sufficient condition for the consistency of the Bayes estimators is provided. The general estimation theory is illustrated by four specific models built for U.S. Treasury Notes transactions data from GovPX. We show that in this market, both information-based and inventory management-based motives are significant factors in the trade-to-trade price volatility.

Key words. Bayes estimation, marked point process, market microstructure noise, Markov chain approximation method, nonlinear filtering, partially observed model, ultra-high frequency data

AMS subject classifications. Primary: 60H35, 62F15, 62M05, 62P05, 93E11; Secondary: 60F05, 60G55, 65C40, 65C60.

1. Introduction.

Electronic trading in all major world financial markets has routinely generated streams of transactions data. Such time-stamped ("tick") data, containing the most detailed information for the price evolution, are referred to as *ultra-high frequency* (UHF) data by Engle in [24]. The direct modeling and analysis of UHF data is essential for insight concerning market microstructure theory, for monitoring and regulating markets and for conducting risk management.

There are two stylized facts in UHF data. First, the arrival times are irregular and random. Second, UHF data contain microstructure (or trading) noise due to price discreteness, price clustering, bid-ask bounce and other market microstructure issues. UHF data can be represented by two kinds of random variables: the transaction time and a vector of characteristics observed at the transaction time. Hence, UHF data are naturally modeled as a *Marked Point Process* (MPP): a *point process* describes the transaction times and *marks* represent, for example, price and volume associated with a trade. However, there are two different views on how to treat the observations, leading to different formulations of the statistical foundation for analyzing UHF data.

*School of Economics and Finance, University of Hong Kong, Hong Kong (gracexhu@hku.hk).

†Department of Finance, Henry W. Bloch School of Management, University of Missouri at Kansas City, Kansas City, MO 64110 (kuipersd@umkc.edu).

‡Department of Mathematics and Statistics, University of Missouri at Kansas City, Kansas City, MO 64110 (zengy@umkc.edu). The author's work was supported in part by National Science Foundation under DMS–0604722, TG-DMS100019, and DMS-1228244, and by University Missouri Research Board.

The first view is well expressed in [24] and is natural from the viewpoint of time series. Because traders choose to transact at random times during the trading day, due to both information-based and liquidity-based motives ([49]), this school of econometricians models the duration between trades as a stochastic phenomena resulting in *an irregularly-spaced time series*. An Autoregressive Conditional Duration (ACD) model is proposed in [25] for analysis. Variants of the ACD model and joint models of duration and return (or price) have been developed with the ability to account for the impact of market and economic factors on the timing of trades (See [50] and [26] for surveys).

The second view, normally perceived from the standpoint of stochastic process, is to treat the transaction observations as *an observed sample path of an MPP*. This view was advocated in [56], where the data is treated as a collection of counting process points, a special case of MPP observations. In this framework, an asset's intrinsic value process, which connects to the usual models in option pricing and the empirical econometric literature for price series, is assumed to be *partially observed* at random trading times through the prices, which are distorted by microstructure noise. Then, the model can be formulated into a *filtering framework with counting process observations* and the link to stochastic nonlinear filtering is established. [56] and [43] develop the Bayesian inference via filtering equations for this model to accommodate the computational burden imposed by this approach.

Motivated by unifying these two views and combining their strengths, we propose a general filtering framework for ultra-high frequency data with two equivalent representations. Just as ARIMA time series models have state-space representations, the first representation can be viewed as a *random observation time state-space model*. The latent state process is a continuous-time multivariate Markov process. The observation times are modeled by a generic point process and the noise is described by a generic transition probability distribution. Other observable economic or market factors are permitted to influence the state process, the observation times, the noise and the observations.

Our proposed framework generalizes the model of [24] by adding a latent continuous-time vector Markov process, which permits the influence of other observable economic or market factors and has a clear economic interpretation. Compared to [56], our framework is a generalization in several respects. Notably, the intrinsic latent value process becomes a correlated multivariate process in our model. The observation times are driven by a general point process with both endogenous and exogenous samplings, including but not limited to conditional Poisson processes, proportional hazard models and ACD models. Other observable factors are allowed to influence the latent process, the observation times and the noise, and the mark space is generic including both discrete and continuous cases. Moreover, the proposed framework connects to several recent literatures.

Just as a state-space model has a filtering formulation, our model also has a second representation as a nonlinear *filtering framework with MPP observations*. Based on this formulation, we develop Bayesian Inference via Filtering Equations (BIFE) for uncertainty quantification in this paper and a companion one [39]. This paper focuses on quantifying parameter uncertainty and the other one on quantifying model uncertainty. Namely, in this paper, we develop the Bayes Estimation via Filtering Equation (BEFE).

To achieve this, we use stochastic nonlinear filtering technique and, under very mild assumptions, we derive the normalized filtering equation, which is a *stochastic partial differential equation* (SPDE). The filtering equation stipulates the evolution of the joint posterior distribution process, which is at the center of quantifying the uncertainty of the state process including the related parameters. The evolving posterior measure-valued process is of continuous time and of infinite dimension. This presents enormous computational challenge for application.

Fortunately, like the Kalman filter, the normalized filter preserves *recursiveness*, which, if properly utilized, provides the tremendous computational advantage. Then, we present a weak convergence theorem, which gives a blueprint for constructing efficient algorithms. Namely, we employ the Markov chain approximation method to construct consistent, easily-parallelizable, recursive algorithms. The algorithms can carry out real-time parameter uncertainty quantification for streaming UHF data. Furthermore, an easy-to-verify sufficient condition for the consistency of the Bayes estimators is given.

Finally, to illustrate the application of the general model and the BEFE, we provide an example from the finance literature involving intradaily prices observed in the market microstructure. The specific data set we examine possesses well-known UHF properties and is particularly amenable to the flexible form of our general model. We build four specific models based on competing theories regarding the evolution of intraday volatility in the U.S. Treasury note market. We exemplify how to develop a recursive algorithm for propagating and updating the joint posterior of the state process. After calibrating the algorithm for correct model identification using simulated data, we use the algorithms to obtain the Bayes estimates for the four models. The estimation results show that both information-based and inventory management-based motives are significant factors in the trade-to-trade price volatility.

The paper is organized as follows. The next section presents the two equivalent representations of our proposed model from the view of stochastic processes with old and new examples. Section 3 develops the BEFE: namely, defines the joint posterior, derives the normalized filter and present a weak convergence theorem for constructing recursive algorithms. Section 4 provides an explicit example for our on-the-run U.S. Treasury 10-year note transactions price data with simulation results. We conclude in Section 5.

2. The Model.

After a brief review on MPP with proper assumptions and notations used throughout this paper and [39], we present our framework in two equivalent representations. The first is more intuitive and the second is a direct generalization of counting process observations. Then, we briefly discuss the generality of the proposed model.

2.1. A Brief Review on MPP.

All stochastic processes are defined on a complete probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$ with the usual hypotheses ([51]), and have with right continuous paths with left limits (cadlag).

Recall that a marked point process $\Phi = \{(T_i, Y_i)\}_{i\geq 1}$ is a sequence of mark point (T_i, Y_i), where the event occurrence times $\{T_i\}$ form a point process with $0 \leq T_1 < T_2 < \cdots < T_i < \cdots < T_\infty$, and the mark $Y_i = Y_i(T_i)$ describes additional information of the event that occurs at time T_i. Y_i is a random element and takes its values in a set $\mathbb{Y}$, called a *mark space*. $(\mathbb{Y}, \mathcal{Y}, \mu)$ is a measure space with a finite measure μ ($\mu(\mathbb{Y}) < \infty$) and $(\mathbb{Y}, d_{\mathbb{Y}})$

is a complete, separable metric space.

There are two general characterizations for a marked point process via the dynamic (martingale) approach: compensators and stochastic (or conditional) intensity kernels. The two notions are closely related. [46] and [19, 20] provide book-length studies on the topics. While the stochastic intensity approach is used extensively in [24], we use both approaches in this paper.

It is convenient to view $\Phi = \{(T_i, Y_i)\}$ as a *random counting measure* defined by

$$\Phi(dt, dy) = \sum_{i \geq 1} \delta_{\{T_i, Y_i\}}(t, y) d(t, y) \tag{2.1}$$

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

For $B \in \mathcal{Y}$, let $\Phi(t, B) := \Phi([0, t] \times B)$, which counts the number of marked points that are in B up to time t. Then, $\Phi(t, B) = \int_0^t \int_B \Phi(dt, dy)$. Let $\Phi(t) := \Phi([0, t] \times \mathbb{Y}) = \Phi(t, \mathbb{Y})$, which is the counting process of event occurrence. Then, $\Phi(t) = \int_0^t \int_{\mathbb{Y}} \Phi(dt, dy)$.

Definition 2.1. *An $\mathcal{F}_t$-predictable random measure γ mapping from $\Omega \times \mathcal{B}(\mathbb{R}^+) \times \mathcal{Y}$ to $[0, +\infty]$ is referred to as a* compensator *of Φ if $\Phi(t, B) - \gamma(t, B)$ is an $\mathcal{F}_t$-martingale for all $t \geq 0$ and $B \in \mathcal{Y}$.*

Definition 2.2. *A stochastic intensity kernel λ of Φ (with respect to a measure η) is a finite $\{\mathcal{F}_t\}$-predictable kernel from $\Omega \times \mathbb{R}^+$ to $\mathbb{Y}$ such that for all predictable $f : \Omega \times \mathbb{R}^+ \times \mathbb{Y} \to \bar{\mathbb{R}}$,*

$$E \int \int f(t, y) \Phi(d(t, y)) = E \int \int f(t, y) \lambda(t, dy) \eta(dt).$$

The connection between the compensator and stochastic intensity kernel is given by

$$\gamma(d(t, y)) = \lambda(t, dy) \eta(dt).$$

When η is a Lebesgue measure $\eta(dt) = dt$, $\lambda(t, B)$ represents the potential for another point at time t with mark in B given $\mathcal{F}_{t-}$, then $P - a.s.$,

$$\lambda(t, B) = \lim_{h \downarrow 0} h^{-1} P(\Phi([t, t+h) \times B) > 0 | \mathcal{F}_{t-}). \tag{2.2}$$

As for $\Phi(t) = \Phi(t, \mathbb{Y})$, its stochastic (or conditional) intensity is a $\mathcal{F}_t$-predictable $\bar{\lambda}(t)$ such that $\bar{\lambda}(t) = \lambda(t, \mathbb{Y})$ and the compensator of $\Phi(t)$ is $\bar{\lambda}(t) dt$.

Finally, we define a few filtrations: $\mathcal{F}_t^{\Phi,V}$ (the observed filtration), $\mathcal{F}_t^{\theta,X,V}$ (the filtration for the latent Markov process), and $\mathcal{F}_t$ (the whole filtration). Let $\mathcal{F}_t^{\Phi,V} = \sigma\{(\Phi(s, B), V(s)) : 0 \leq s \leq t, B \in \mathcal{Y}\}$, $\mathcal{F}_t^{\theta,X,V} = \sigma\{(\theta(s), X(s), V(s)) : 0 \leq s \leq t\}$, and $\mathcal{F}_t = \sigma\{(\theta(s), X(s), \Phi(s, B), V(s)) : 0 \leq s \leq t, B \in \mathcal{Y}\}$, where V is a vector of observable economic or market factors, which can be deterministic or random as seen in Section 4. We let $\mathcal{F} = \mathcal{F}_\infty$ and define $\mathcal{F}_{t-} = \sigma\{(\theta(s), X(s), \Phi(s, B), V(s)) : 0 \leq s < t, B \in \mathcal{Y}\}$. Note that a $\mathcal{F}_t$-predictable process is essentially $\mathcal{F}_{t-}$-measurable.

2.2. Representation I: Random Arrival Time State-Space Model. Like a state-space time series model, this representation has similar components: a state process, observation times, the observations themselves and noise. Unlike the usual state-space models, our model has random arrival times and a continuous-time state process.

2.2.1. The State Process. Applying a modern Bayesian idea for parameter estimation, we extend the state space to include parameter space, and allows the components of θ to potentially change in continuous time. A mild assumption is made on (θ, X).

Assumption 2.1. *(θ, X) is a $p+m$-dimension vector Markov process that is the unique 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, and f is in the domain of $\mathbf{A}_v$.*

The generator and martingale problem approach (see, for example, [27]) furnishes a powerful tool for the characterization of Markov processes, which includes usual deterministic ODE. Obviously, Assumption 2.1 is a multivariate version of Assumption 2.1 in [56], allowing for correlation and subsuming most popular stochastic processes such as multivariate diffusion, jump and Levy processes employed in asset pricing theory. Because of the inclusion of θ, our model further allows unobserved stochastic volatility and regime switching. Moreover, we do not assume stationarity of X, which makes our assumption less restricted than that in the literature for estimating Markov processes sampled at discrete or random times via the operator approach ([36], [1], [22], [3] and [15] among others). Furthermore, other observable factors, represented by a vector process V, are incorporated in the state process of X through the generator, $\mathbf{A}_v$.

2.2.2. Observation times. $T_1, T_2, \ldots, T_i, \ldots$, 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 $\Phi^{t-} = \{(T_i, Y_i) : T_i < t\}$. Namely, we allow $\bar{\lambda}(t)$ to be Markovian in (θ, X, t) and *non-anticipating* in V and Φ (depending on the sample paths of V and Φ up to time $t-$).

Clearly, a conditional Poisson process is an example. However, this form is very general including point processes generated from *endogenous sampling* and *exogenous sampling*. Endogenous sampling is when $\bar{\lambda}(t)$ depends on $(\theta(t), X(t))$. Some related estimation problems are studied in [22], [15] and [52] by different methods. Exogenous sampling is when $\bar{\lambda}(t)$ does not depend on $(\theta(t), X(t))$, but is only $\mathcal{F}_t^{V,\Phi}$-predictable. Cox model and the more general proportional hazard formulations, the Hawkes process ([45]), and ACD models are popular examples of exogenous sampling. We note that the case of exogenous sampling significantly simplifies the filtering equation for posteriors and the evolution system equations for Bayes factors, and greatly reduces the related numerical computation.

2.2.3. The Observations. The noisy observation at event time T_i, $Y(T_i)$, takes a value in the mark space $\mathbb{Y}$ and is modeled by $Y(T_i) = F(X(T_i))$, where $y = F(x)$ is a random transformation from $x = X(T_i)$ to $y = Y(T_i)$ specified by a transition probability distribution $p(dy|x)$. Specifically, the conditional distribution of $Y(T_i)$ given $X(T_i) = x$ is given by $p(dy|x;t) = p(dy|X(t); \theta(t), V^{t-}, \Phi^{t-}, t)$ when $t = T_i$.

In many papers on realized volatility estimators in the presence of noise such as [4], $Y(T_i) - X(T_i)$ has a normal distribution with mean zero and constant variance. The noise in [47] is additive normal with rounding. The noise in [56] includes additive doubly-geometric, rounding and clustering noises. These are a few examples of $F(x)$ with $p(dy|x)$.

In summary, the observation $\{T_i, Y(T_i)\}$ forms a MPP with the predictable stochastic intensity kernel $\bar{\lambda}(\theta(t), X(t), V^{t-}, \Phi^{-}, t)\ p(Y(t)|X(t); \theta(t), V^{t-}, \Phi^{t-}, t)$. This representation can be viewed as a *random arrival time state-space model*, or a *random arrival time hidden Markov model* in a generic mark space.

When applied to financial UHF data, X becomes the intrinsic value vector process of m assets. When $m = 1$, this representation is a generalization of [56], and shares the simple economic notion that $Y(T_i)$, the observed price at trading time T_i, is based upon $X(T_i)$, the intrinsic value, but is distorted by the trading noise modeled by F, the random transformation. The proposed framework connects to the literature of Vector AutoRegressive (VAR) structural models in market microstructure theory surveyed in [38], and to the literature of realized volatility with microstructure noise in [58], [6], [57], [28], [47], and [8] among many others.

2.3. Representation II: Filtering with MPP Observations. Under this representation, (θ, X) becomes the signal, which is not directly observable, but can be partially observed through a MPP, $\Phi = \{T_i, Y_i\}_{i\geq 1}$. The observation is expanded to (Φ, V), including an auxiliary predictable process V, which can be thought of as observable economic or market nuisance variables. The inclusion of V in the observation enables the proposed model to study the interaction of the MPP with other observable variables. With four more assumptions, we formulate (θ, X, Φ, V) into a filtering representation with MPP observations.

Let $\lambda(t, dy)$ be in the form $\lambda(t, dy; \theta(t), X(t), V^{t-}, \Phi^{t-}, t)$.

Assumption 2.2. *Under P, the stochastic intensity kernel of $\Phi = \{(T_n, Y_n)\}_{n\geq 1}$ is given by $\lambda(t, dy) = \bar{\lambda}(t)p(dy|X(t), t)$, namely,*

(2.3)
$$\lambda(t, dy; \theta(t), X(t), V^{t-}, Y^{t-}, t-) = \bar{\lambda}(\theta(t), X(t), V^{t-}, Y^{t-}, t-)p(dy|X(t); \theta(t), V^{t-}, Y^{t-}, t-).$$

Note that $\bar{\lambda}$ is a hazard rate that determines the conditional odds for the next event occurrence, and $p(dy|X(t))$ specifies the conditional distribution for the next mark occurrence given $X(t)$. Then, the compensator under P is $\gamma_P(d(t, y)) = \lambda(t, dy)dt$. Assumption 2.2 implies that two representations have the same stochastic intensity kernel and hence, they are equivalent in the sense of having the same probability law by Theorems 8.3.3 and 8.2.2 in [46]. This guarantees that statistical inferences based on the latter representation are consistent with the former.

In the next assumption, we assume the existence of a unit Poisson random measure with mark distribution $\mu(dy)$ as a reference measure and with the crucial property under the reference measure that the signal and the observation are independent. This assumption is useful in deriving the filtering equations and in constructing the approximate filters in this paper and [39].

Assumption 2.3. *There exists a* reference measure *Q, with respect to which P is absolutely continuous, namely, $P << Q$ so that under Q*

- *(θ, X) and V are independent of $\Phi = \{(T_n, Y_n)\}_{n\geq 1}$;*

- *The compensator of MPP Φ is $\gamma_Q(d(t,y)) = \mu(dy)dt$, or the corresponding stochastic intensity kernel is $\mu(dy)$.*

We use $E^Q[X]$ or $E^P[X]$ to indicate that the expectation is taken with respect to a specific probability measure throughout the paper and [39].

Next is a mild assumption ensuring the existence of a reference measure Q. First, we define the ratio of the mark distributions under P and Q as

$$r(y) = r(y;\theta(t), X(t), V^{t-}, Y^{t-}, t-) = \frac{p(dy|X(t);\theta(t), V^{t-}, Y^{t-}, t-)}{\mu(dy)}. \tag{2.4}$$

and the ratio of the compensators under P and Q as

$$\zeta(t,y) = \frac{\gamma_P(d(t,y))}{\gamma_Q(d(t,y))} = \frac{\lambda(t,dy)}{\mu(dy)} = \bar{\lambda}(\theta(t), X(t), V^{t-}, Y^{t-}, t-) r(y;\theta(t), X(t), V^{t-}, Y^{t-}, t-). \tag{2.5}$$

Let $L(t) = \frac{dP}{dQ}(t)$ be the Radon-Nikodym derivative given below with more explanation in [39].

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

Assumption 2.4. *The $\zeta(t,y)$ defined in (2.5) satisfies the condition that $E^Q[L(T)] = 1$ for all $T > 0$.*

Sufficient conditions for the above assumption can be found in Section 3.3 of [9]. Our final assumption is a technical condition that ensures that $\{T_i\}$ is *P-nonexplosive*, namely, $P(T_\infty = +\infty) = 1$ and that the filtering equations are well-defined.

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

2.4. Generality of the Model. The section illustrates the richness of the proposed framework, unifying the existing important models (Classes I and II), and providing new interesting models (Class III).

Our model encompasses 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. Class I models can be further classified based on those that model event times alone, compared to models that jointly account for event times and their marks. Representative examples of the first sub-class are mentioned in Section 2.2.2. The general framework in [24] is representative of the second sub-class.

As an alternative, Class II contains the model of [56] and its multi-asset version [53] and [54]. A Heston stochastic volatility model with normal noise at random times modified from [2] is also an example. Class II also subsumes additional models in the literature. For example, it generalizes models that provide for the estimation of Markov processes sampled at random time intervals, with or without confounding microstructure noise ([3], [22]. In addition, filtering models that estimate market volatility from UHF data are special cases of our model. Recent papers in this area include [33], [34], [17], and [18]. As a third example, models in many classical filtering problems with MPP observations are contained within our general framework; a partial list of reference papers includes [55], [21], [9], Chapter 19.3 of [48] (the first edition in 1978), Chapter 6.3 of [12], [42], and [23].

Finally, the potential for the latent continuous-time process X to depend on confounding observable factors V is a central element for the models in Class III, a feature which is also built into the model proposed in this paper. One simple example is provided in Section 4, where V is modeled using two variables incorporated into the volatility of a geometric Brownian motion intrinsic value process X. The components in V can change at random (trades) or deterministic (scheduled macroeconomic news releases) times, and can be modeled for a desired set of conditions in the market microstructure including the trade sign (direction), volume, or other trading indicators encountered in high-frequency financial markets trading. Moreover, the flexibility of the model allows for the marks $\{Y_j\}$ to represent trade price, bid-ask quotes, or some combination of price and quote data. Taken together, these features create the opportunity for new insights in empirical studies of market microstructure theory.

3. Bayes Estimation via Filtering Equation.

The partially-observed model offers a typical framework for Bayes estimation. In this section, we derive the normalized filtering equation and point out its two important features. Then, we present a convergence theorem, describe the blueprint for constructing recursive algorithm, and prove a lemma for the consistency of the Bayes estimators.

3.1. The Posterior and its Filtering Equation.

We present the posterior distribution process and derive the normalized filtering equation.

3.1.1. The Posterior Distribution Process of $(\theta(t), X(t))$.

First, we define the conditional distribution.

Definition 3.1. *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*

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

If a prior distribution is assumed on $(\theta(0), X(0))$ as in Bayesian paradigm, then π_t becomes the continuous-time posterior distribution, which is determined by $\pi(f,t)$ for all continuous and bounded f. When prior information is accessible, Bayesian approach provides a scientific way to incorporate the prior information. If no prior information is available, uniform noninformative priors can be employed.

The posterior summarizes Bayesian uncertainty quantification about $(\theta(t), X(t))$ given $\mathcal{F}_t^{\Phi,V}$ and is at the center of Bayesian point and interval estimation. A posterior allows us to determine the different kind of estimators from Bayesian viewpoint such as posterior mean, posterior median, or other summary quantities, which minimize the posterior expectation of a corresponding loss function according to statistical decision theory. Moreover, a posterior allows us to construct Bayesian forms of interval, or more general, set estimation such as the credible set and the highest posterior density region.

3.1.2. The Normalized Filtering Equation. The infinite dimensional posterior distribution process $\{\pi_t\}_{t\geq 0}$ is uniquely characterized by the following normalized filtering equation, which is a SPDE.

Theorem 3.1. *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,*

$$\begin{aligned} \pi(f,t) = \pi(f,0) &+ \int_0^t \pi(\mathbf{A}_v f, s) ds \\ + \int_0^t \int_{\mathbb{Y}} &\Big[\frac{\pi(\zeta(y)f, s-)}{\pi(\zeta(y), s-)} - \pi(f, s-)\Big]\big(\Phi(d(s,y)) - \pi(\zeta(y), s)\mu(dy)ds\big). \end{aligned} \tag{3.1}$$

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

$$\pi(f,t) = \pi(f,0) + \int_0^t \pi(\mathbf{A}_v f, s) ds + \int_0^t \int_{\mathbb{Y}} \Big[\frac{\pi(r(y)f, s-)}{\pi(r(y), s-)} - \pi(f, s-)\Big]\Phi(d(s,y)), \tag{3.2}$$

where $r(y) = r(y; \theta(s), X(s), \Phi^{s-}, V^{s-}, s-)$ is defined in (2.4).

Proof. Let $\rho(f,t) = E^Q[f(\theta(t), X(t))L(t)|\mathcal{F}_t^{\Phi,V}]$. By Theorem 3.1 in [39], $\rho(f,t)$ follows the *unnormalized filtering equation* below:

$$\rho(f,t) = \rho(f,0) + \int_0^t \rho(\mathbf{A}_v f, s) ds + \int_0^t \int_{\mathbb{Y}} \rho((\zeta(y)-1)f, s-)(\Phi(d(s,y)) - \mu(dy)ds), \tag{3.3}$$

for $t > 0$ and $f \in D(\mathbf{A}_v)$, the domain of generator $\mathbf{A}_v$, where $\zeta(y) = \zeta(s-, y)$ defined in (2.5).

Bayes Theorem[1] ([12], page 171) provides the relationship between $\rho(f,t)$ and $\pi(f,t)$:

$$\pi(f,t) = \frac{\rho(f,t)}{\rho(1,t)}. \tag{3.4}$$

Note that $\rho(1,t)$ follows the simple stochastic differential equation (SDE) obtained from (3.3) by taking $f = 1$. To determine the SDE for $\pi(f,t)$, we apply Itô's formula for semimartingale (see [51], page 78) to $f(X,Y) = \frac{X}{Y}$ with $X = \rho(f,t)$ and $Y = \rho(1,t)$. After some simplification, we have

$$\begin{aligned} \pi(f,t) =& \pi(f,0) + \int_0^t \pi(\mathbf{A}_v f, s) ds - \int_0^t \int_{\mathbb{Y}} [\pi(\zeta(y)f, s) + \pi(f,s)\pi(\zeta(y), s)]\mu(dy)ds \\ &+ \int_0^t \int_{\mathbb{Y}} (\pi(f,s) - \pi(f,s-))\Phi(d(s,y)). \end{aligned} \tag{3.5}$$

A last step remains to make $\pi(f,s)$ in the integrand of the last integral predictable. Assume a trade at price y occurs. Then,

$$\pi(f,s) = \frac{\rho(f,s)}{\rho(1,s)} = \frac{\rho(f,s-) + \rho((\zeta(y)-1)f, s-)}{\rho(1,s-) + \rho(\zeta(y)-1, s-)} = \frac{\rho(\zeta(y)f, s-)}{\rho(\zeta(y), s-)} = \frac{\pi(\zeta(y)f, s-)}{\pi(\zeta(y), s-)}.$$

[1] Bayes Theorem implies that the conditional probability measure π is obtained by normalizing the conditional measure ρ by its total measure. For this reason, the equation governing the evolution of $\rho(f,t)$ is called the *unnormalized filtering equation*, and that of $\pi(f,t)$ the *normalized filtering equation*.

Hence, Equation (3.5) implies Equation (3.1).

Recall $\zeta(y) = \zeta(t,y) = \bar{\lambda}(t)r(y|X(t),t)$. When $\bar{\lambda}(t)$ is $\mathcal{F}_t^{\Phi,V}$-measurable, $\bar{\lambda}(t)$ in $\zeta(y)$ can be pulled out of the conditional expectation. Two additional observations further simplify Equation (3.1) to Equation (3.2). First, $\int_{\mathbb{Y}} \pi(\zeta(y)f,t)\mu(dy) = \bar{\lambda}(t)\pi(f,t)$, and $\int_{\mathbb{Y}} \pi(\zeta(y),t)\mu(dy) = \bar{\lambda}(t)$. Then, the two terms in the integrand of ds in Equation (3.1) cancel out. Second, $\frac{\pi(\zeta(y)f,s-)}{\pi(\zeta(y),s-)} = \frac{\pi(fr(y),s-)}{\pi(r(y),s-)}$.

For the uniqueness of the normalized filtering equation, see Appendix A.4 in [39]. ■

Under P, it can be shown that $\pi(\zeta(y),t)\mu(dy)dt$ is a $\mathcal{F}_t^{\Phi,V}$-predictable compensator of Φ. Hence, the normalized filter has a semimartingale representation.

Note that in the case of exogenous sampling, such as when the event times follow an ACD model or a Cox model, $\{T_i\}$'s stochastic intensity $\bar{\lambda}(t)$ is $\mathcal{F}_t^{\Phi,V}$-predictable. Then, Equation (3.1) is simplified to (3.2). The advantage is the reduction of the computation of the posterior distribution. Moreover, observe that the stochastic intensity disappears in the simplified filter. Hence, the parameters for the model of $\{T_i\}$ can be estimated separately via other approaches such as maximum likelihood procedure for an ACD model. The uncertainty of the rest parameters that are related to X or the noise can be quantified by the Bayesian approach via filtering equation to be developed, and the Bayes estimates are *model-free* of the assumptions on durations. The example to be given in Section 4 takes such advantages. The tradeoff is the exclusion of the relationship between durations and the state (θ, X).

3.1.3. Two Features of the Filtering Equation. First, the normalized filters can be split up into the *propagation* and the *updating* equations. We illustrate such separation by Equation(3.2). Let the arrival times be $t_1, t_2, \ldots$, then, Equation (3.2) can be dissected into the *propagation equation*, describing the evolution without event occurrence:

$$\pi(f, t_{i+1}-) = \pi(f, t_i) + \int_{t_i}^{t_{i+1}-} \pi(\mathbf{A}_v f, s)ds, \tag{3.6}$$

and the *updating equation*, describing the update when an event occurs:

$$\pi(f, t_{i+1}) = \frac{\pi(r(y)f, t_{i+1}-)}{\pi(r(y), t_{i+1}-)}, \tag{3.7}$$

where the mark at time t_{i+1} is assumed to be y. Note that the propagation equation has no random component, implying that π_t evolves deterministically when there are no event arrival. The updating equation is random because the observation mark y is random.

Second, the normal filters have recursiveness. Again, we illustrate this by (3.2), where the goal is to find the expressions for π_t, the posterior in Bayesian paradigm. The normalized filter (3.2) possesses a desirable *recursive form* as Kalman filter. In nonmathematical terms, Equation (3.2) supplies a device H such that for all t,

$$\pi_{t+dt} = H(\pi_t, \Phi(d(t,y))), \tag{3.8}$$

where $\Phi(d(t,y))$ denotes the innovation during $(t, t+dt)$. The device H can be regarded as a *computer program* not depending on data. π_t is all that has to be stored in *memory*, and $\Phi(d(t,y))$ is a*new datum* fed into the program H. If no new datum arrives during $(t, t+dt)$, H is determined by the propagation equation (3.6). If a new datum arrives, H is specified by the updating equation (3.7). Note that π_{t+dt} only depends on π_t, but not the whole past $\{\pi_s : 0 \leq s \leq t\}$. This enormous dimension reduction saves memory space. In statistical jargons, we may say that $(\pi_t(X,\theta), d\Phi_t)$ is a *sufficient statistic* for the problem of "estimating" $\pi_{t+dt}(X,\theta)$ on the basis of $\mathcal{F}^{\Phi,V}_{t+dt}$. In summary, the practical implication of recursiveness is to make it possible the real time parameter uncertainty quantification for streaming UHF data.

3.2. A Convergence Theorem. Theorem 3.1 provides the normalized filter to characterize the posterior measure-valued process, which is infinite dimensional. For computation, one needs to reduce the infinite dimensional problem to a finite dimensional problem and to construct a recursive algorithm, whose basic requirement is consistency. Namely, the computed approximate posterior distribution-valued process converges to the true one. The weak convergence theorem below provides a blueprint for constructing consistent recursive algorithms through Kushner's Markov chain approximation methods ([44]).

Due to the page limit, we only give a brief description of weak convergence and state the theorem. A formal detailed description and the proof is contained in [39].

Following the literature of weak convergence (see, for example, Chapter 3 of [27]), we use the notation, $X_\epsilon \Rightarrow X$, to mean X_ϵ converges weakly to X in the Skorohod topology as $\epsilon \to 0$. Such weak convergence is uniform in time for the process X.

Let $(\theta_\epsilon, X_\epsilon)$ be an approximation of the signal, (θ, X). $(\theta_\epsilon, X_\epsilon)$ can be naturally taken as a sequence of pathwise approximating Markov chain index by ϵ as seen in Section 4.3. Let the observations for an approximate signal $(\theta_\epsilon, X_\epsilon)$ be $\Phi_\epsilon = \{(T_{i,\epsilon}, Y_{i,\epsilon})\}_{i\geq 1}$ and let the approximation of $\pi(f,t)$ be $\pi_\epsilon(f,t) = E^{P_\epsilon}[f(\theta_\epsilon, X_\epsilon(t)) \,|\, \mathcal{F}^{\Phi_\epsilon,V}_t]$.

Theorem 3.2. *Suppose that (θ, X, Φ, V) on $(\Omega, \mathcal{F}, P)$ 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 $\epsilon \to 0$, then, for any bounded continuous functions, f,*
(1) $\Phi_\epsilon \Rightarrow \Phi$ under the physical measures; and
(2) $\pi_\epsilon(f,\cdot) \Rightarrow \pi(f,\cdot)$, as $\epsilon \to 0$.

This theorem states that as long as the approximate signal weakly converges to the true signal, we have the weak convergence of (1) the observation of an approximate model to that of the true model, and (2) the posterior measure-valued process of the approximate model to that of the true one.

3.3. A Blueprint for Constructing Recursive Algorithms. Theorem 3.2 provides a three-step blueprint for constructing consistent recursive algorithms based on Markov chain approximation method. Moreover, the algorithms are easily parallelizable according to the grid space of time-invariant parameters. The algorithms compute the current joint posterior distribution and the marginal posterior means, which are the usual Bayes estimates under squared error loss. As in numerical solution to a PDE, we discretize the state and the mark spaces into two grids, respectively. Then, there are three main steps in constructing a recursive algorithm, namely, $H_\epsilon(\cdot,\cdot)$ as the approximation of $H(\cdot,\cdot)$ in (3.8). Step 1 is to construct $(\theta_\epsilon, X_\epsilon)$, a Markov chain approximation to (θ, X), and let $r_\epsilon(y) = r(y_\epsilon|\theta_\epsilon, x_\epsilon)$

be an approximation to $r(y) = r(y|\theta, x)$, where $(\theta_\epsilon, x_\epsilon)$ and y_ϵ are restricted to the grids of $(\theta_\epsilon, X_\epsilon)$ and $\mathbb{Y}_\epsilon$, respectively. Step 2 is to obtain the approximate filter for $\pi_\epsilon(f,t)$ corresponding to $(\theta_\epsilon, X_\epsilon, Y_\epsilon, r_\epsilon)$ by applying Theorem 3.1. Recall that the approximate filter can also be dissected into the propagation equation:

$$\pi_\epsilon(f, t_{i+1}-) = \pi_\epsilon(f, t_i) + \int_{t_i}^{t_{i+1}-} \pi_\epsilon(\mathbf{A}_{\epsilon,v} f, s) ds, \tag{3.9}$$

and the updating equation (assuming that a mark at y_{i+1} occurs at time t_{i+1}):

$$\pi_\epsilon(f, t_{i+1}) = \frac{\pi_\epsilon(f r_\epsilon(y_{i+1}), t_{i+1}-)}{\pi_\epsilon(r_\epsilon(y_{i+1}), t_{i+1}-)}. \tag{3.10}$$

If no new mark point arrives during $(t, t+dt)$, H_ϵ is governed by the propagation equation (3.9). If a new mark point arrives, H_ϵ is specified by the updating equation (3.10). Step 3 converts Equations (3.9) and (3.10) to a recursive algorithm for the state grid and in discrete time by two substeps: (a) we represent $\pi_\epsilon(\cdot, t)$ as a finite array with the components being $\pi_\epsilon(f,t)$ for lattice-point indicators f and (b) we approximate the time integral in (3.9) with an Euler scheme or other higher order numerical schemes.

3.4. Consistency of the Bayes Estimators. The following lemma provides an easily-verified sufficient condition for the consistency of the Bayes estimators for the time-invariant parameters in the proposed model.

Lemma 3.1. *Suppose that (θ, X, Φ, V) satisfies Assumptions 2.1 - 2.5 with a vector of parameter $\theta = (\theta_1, \theta_2)$ where θ_1 and θ_2 are the time-invariant and time-variant vector of parameter with some priors. If there exists a L^P - consistent estimator for θ_1 based on (X, Φ, V) for some $p \geq 1$, namely, there exists $\hat{\theta}_t = \hat{\theta}\big((X_s, \Phi_s, V_s) : 0 \leq s \leq t\big)$, such that $E|\hat{\theta}_{1,t} - \theta_1|^p \to 0$ as $t \to \infty$, then, $E^P[f(\theta_1)|\mathcal{F}_t^{\Phi,V}] \to f(\theta_1)$ a.s. as $t \to \infty$ for any bounded continuous function f.*

Proof. Martingale convergence theorem implies that $E^P[f(\theta_1)|\mathcal{F}_t^{\Phi,V}] \to E^P[f(\theta_1)|\mathcal{F}_\infty^{\Phi,V}]$ a.s. If we can prove that θ_1 is $\mathcal{F}_\infty^{\Phi,V}$-measurable, then $E^P[f(\theta_1)|\mathcal{F}_\infty^{\Phi,V}] = f(\theta_1)$ and the consistency is obtained.

In order to show that θ_1 is $\mathcal{F}_\infty^{\Phi,V}$-measurable, it is enough to show the existence of an $\mathcal{F}_\infty^{\Phi,V}$-measurable and a.s. consistent estimator of θ_1. Since $|x|^p$ is a convex function for $p \geq 1$, we use Jensen's inequality to obtain

$$E|\bar{\theta}_{1,t} - \theta_1|^p = E\big|E[\hat{\theta}_{1,t} - \theta_1|\mathcal{F}_t^{\Phi,V}]\big|^p \leq E\big[E|\hat{\theta}_{1,t} - \theta_1|^p|\mathcal{F}_t^{\Phi,V}\big] = E|\hat{\theta}_{1,t} - \theta_1|^p,$$

which goes to zero by the assumption. Markov inequality implies that $\bar{\theta}_{1,t} \to \theta_1$ in probability. Then, we can select an almost surely convergent subsequence of $\{\bar{\theta}_{1,t}\}$ to θ_1, which is the needed $\mathcal{F}_\infty^{\Phi,V}$-measurable and a.s. consistent estimator of θ_1. ■

This lemma implies that we have the strong consistency of the Bayes estimator for a time-invariant parameter as long as we can construct a L^p-consistent estimator based on (X, Φ, V). This condition is much easier to check and meet than to construct a L^2-consistent estimator based solely on Φ as in the proof of Theorem 5.1 of [56]. Many previous consistency results based on X such as those in [22] can be applied to find a L^p-consistent estimator and then obtain the consistency. Furthermore, L^2-consistency is relaxed to L^p-consistency for some $p \geq 1$, allowing the application to Levy jump processes proposed asset price (see examples in Chapter 4 of [16]) where the second moment does not exist.

4. Modeling and Estimation via Filtering: An Application to Bond Transactions Data.

This section provides an example for the method of building a specific filtering model and that of constructing a recursive algorithm. We provide simulation and estimation examples drawn from the microstructure of trading in U.S. Treasury notes - a market with well-known UHF properties and, as we illustrate below, one where theories regarding market behavior are particularly amenable to the flexible form of our model.

4.1. Data Description.

The data consists of a complete record of quote revisions and trades in the voice-brokered 10-year U.S. Treasury note market during the period August - December 2000 as transacted through GovPX, Inc.[2] During this time period, GovPX was the leading interdealer broker for trades in the U.S. Treasury note market ([7]), acting as a clearinghouse for market activity transacted through five of the six largest Treasury interdealer brokers ([41]). The GovPX tick-by-tick data has been widely used in the empirical finance literature in recent years in studies of market microstructure effects, including [29], [40], [11], and [35], among others. The GovPX data is also the primary source used by the Center for Research in Security Prices (CRSP) to construct the Daily U.S. Government Bond Database.

The specific data we examine concerns quotes and trades for the "on-the-run" 10-year U.S. Treasury note during the sample period. On-the-run (hereafter, "OTR") notes are those of a given initial maturity that have been most-recently auctioned by the U.S. Treasury; trading activity on any given day is extremely active for OTR notes ([31]). During the sample period, the 10-year note was auctioned on a quarterly cycle, with occasional skipped auctions due to government surpluses at that time. As a result, the data contains a tick-by-tick record for the OTR 10-year note issued on August 15, 2000, and follows this note through the end of calendar year 2000, during which time it remained OTR in the Treasury market. [3]

Due to their superior liquidity, OTR notes are the preferred trading vehicle for participants in the Treasury market with both liquidity-based and information-based motives for trading. For this reason, studies that examine market efficiency and microstructure behavior in the Treasury market almost exclusively focus on the same data we use here for our filtering model examples. Given the richness of our model's functional form, we are able to conduct estimation and inferential tests regarding competing microstructure

---

[2] Research that relies on the GovPX data typically ends with calendar year 2000, as GovPX changed its data reporting methods in early 2001 in a way that impedes a detailed study of market microstructure effects. See [32] and [11] for details.

[3] The GovPX data itself consists of a raw data feed in a flat-file text format for the real-time, unfiltered record updates observed for all Treasury securities throughout the trading day. Substantial data preprocessing is necessary to re-construct the original quotes and trades from this raw data; see [32] in particular for details.

hypotheses in the Treasury market using the OTR note transactions data in this paper and [39]. A distinct advantage of estimating our model using this specific data is that stale trading noise is minimized, while the information content of trades is maximized.

4.2. A Filtering Model for UHF Transactions Data in the Treasury Market.

We illustrate how to build a filtering model to incorporate market behavior by following Representation I because it is more intuitive.

4.2.1. The Intrinsic Value Process.

First, the state process becomes the intrinsic value process X_t, of which we consider a general form of the geometric Brownian motion(GBM) process for illustration. We assume the intrinsic value process follows the SDE given by:

$$\frac{dX_t}{X_t} = \mu dt + (\sigma_0 + \sigma_{en} V_{en}(t) + \sigma_{bs} V_{bs}(t)) dB_t \tag{4.1}$$

where B_t is a standard Brownian Motion, and $(\mu, \sigma_0, \sigma_{en}, \sigma_{bs})$ are model parameters. The observable factors $V(t) = \{V_{en}(t), V_{bs}(t)\}$ comprise two processes examined in the finance empirical literature on Treasury tick data.

$V_{en}(t)$ indicates whether trades observed at time t are coincident with significant macroeconomic news announcements. [40] and others find that volatility in the Treasury market increases significantly upon release of macroeconomic news, and [11] provide evidence that information-based trading is the predominant source of the observed change in volatility. The other process $V_{bs}(t)$ indicates whether a trade is buyer or seller initiated. [10] and [30] argue that Treasury dealers, as the primary buyers of new OTR notes from the U.S. government, must manage the resulting inventory in a manner that creates volatility effects in the auction aftermarket. An advantage of the proposed framework we use here is that we can easily incorporate different theories. Specifically, we examine the four specifications of the process $V(t)$ described above in terms of the implied parameter constraints:

Model 1: Simple GBM, $\sigma_{bs} = \sigma_{en} = 0$
Model 2: GBM with an economic news dummy, $\sigma_{bs} = 0$.
Model 3: GBM with a buyer-seller initiation dummy, $\sigma_{en} = 0$.
Model 4: GBM with both dummies.

Let $\theta = (\mu, \sigma_0, \sigma_{en}, \sigma_{bs}, \rho)$, where ρ is used to model non-clustering noise to be defined later. The generator of (4.1) is:

$$\mathbf{A}_v f(x, \theta) = \mu x \frac{\partial}{\partial x} f(\theta, x) + \frac{1}{2}(\sigma_0 + \sigma_{en} V_{en} + \sigma_{bs} V_{bs})^2 x^2 \frac{\partial^2}{\partial x^2} f(\theta, x). \tag{4.2}$$

4.2.2. Trading Durations and Noise.

We assume the trade duration follows an exponential ACD or Wiebull ACD model, which is exogenous sampling. Hence, the simplified version of the normalized filtering equation for the posterior can be employed.

We incorporate trade noise onto the intrinsic values at observed trading times to produce the model price process. Three sources of noise are identified and studied in the finance literature (e.g., [37]): Price discreteness, trade clustering, and nonclustering noise.

To model discreteness, the mark space for the tick price (the minimum price variation in a trading market) becomes a discrete space $\mathbf{Y} = \{\frac{a+1}{M}, \frac{a+2}{M}, \cdots, \frac{a+m}{M}\}$. For the OTR 10-year Treasury note, M is 64 by market convention. In addition, market trades tend to occur on coarse ticks such as $\frac{2}{M}$ and $\frac{4}{M}$, resulting in price clustering of the data. Finally, we assume that nonclustering noise captures all other unspecified noise in the market.

Given this framework, 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 defined as the non-clustering noise. We construct a random transformation $y = F(x)$ in three steps and calculate the corresponding transition probability $p(y|x)$.

Step (i): 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:

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

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 conditionally independent given $\{y'_i\}$. To be consistent with the bond price data analyzed, we construct a simple random biasing function for the tick of 1/64 percentage (i.e. $M = 64$). The data to be fitted has the clustering occurrence: odd sixteenths are most likely; odd thirty-seconds are the second most likely; odd sixty-fourths are least likely; and each tick of its kind has about the same frequencies.

To produce such clustering, a biasing function $b(\cdot)$ is constructed based on these biasing rules: if the fractional part of y' is even sixty-fourths, then y stays on y' with probability one; if the fractional part of y' is odd sixty-fourths, then y stays on y' with probability $1 - \alpha - \beta$; y moves to the closest odd thirty-seconds with probability α, and moves to the closest odd sixteenth with probability β.

To sum up, for consistency with market convention in the OTR 10-year note market, we construct a random transformation F for ticks with $M = 64$:

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

where F is specified by $p(y|x)$ given in Appendix A. The parameters related to clustering noise (α, β) can be estimated through the method of relative frequency.

As for Representation II, the price process $\Phi = \{t_i, Y(t_i)\}$ can be viewed as a random counting measure $\Phi(t, B)$ for $B \subset \mathbb{Y}$. with the predictable stochastic intensity kernel for $y \in \mathbb{Y}$:

$$\lambda(t, y) = \bar{\lambda}(t)p(y|X(t); \theta, t) \tag{4.3}$$

where $p(y|X(t); t) = p(y|x)$ is the transition probability given in Appendix A.

The focus of this example is to quantify the uncertainty of the parameter $\theta = (\mu, \sigma_0, \sigma_{en}, \sigma_{bs}, \rho)$ for the four models, via the Bayesian approach using the recursive algorithm constructed in the following subsection.

4.3. The Recursive Algorithm for Posterior. We exemplify the three-step blueprint of Markov chain approximation method in Section 3.3 to construct a consistent and easily-parallelized recursive algorithm using Model 4. The algorithms for Models 1-3 can be developed similarly. The goal is to compute π_t, the joint posterior distribution of $(\theta, X(t))$ given $\mathcal{F}_t^{\Phi,V} = \sigma(((T_i, Y_i), V(s)) : T_i \leq t, s \leq t)$, the observations up to time t with observable market factors. Recall Definition 3.1. 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 $\pi_\epsilon(f,t) = \sum_{\theta,x} f(\theta,x)\pi_{\epsilon,t}(\theta,x)$ where the sum is over the state grid.

In preparation for constructing the recursive algorithm, we discretize the six-dimension state space of $(\theta, X(t))$ where $\theta = (\mu, \sigma_0, \sigma_{en}, \sigma_{bs}, \rho)$ to form a state grid. Mark space $\mathbb{Y}$ is already a discrete space for the models, and $r(y)$ in (3.2) becomes $p(y|X(t);\theta)$ for $y \in \mathbb{Y}$. Suppose there are $n_\mu + 1$, $n_{\sigma_0} + 1$, $n_{\sigma_{en}} + 1$, $n_{\sigma_{bs}} + 1$, $n_\rho + 1$, and $n_x + 1$ lattices in the discretized spaces of $\mu, \sigma_0, \sigma_{en}, \sigma_{bs}, \rho$ and X respectively. For example, the lattice space of μ becomes

$$\mu : [\alpha_\mu, \beta_\mu] \to \{\alpha_\mu, \alpha_\mu + \epsilon_\mu, \dots, \alpha_a + (n_\mu - 1)\epsilon_\mu, \beta_\mu\},$$

where $\beta_\mu = \alpha_\mu + n_\mu \epsilon_\mu$, and the number of lattices is $n_\mu + 1$. Define $\mu_i = \alpha_\mu + i\epsilon_\mu$, the ith element in the discretized parameter space of μ, and define $\sigma_{0,j}, \sigma_{en,k}, \sigma_{bs,l}, \rho_m$ and X_w similarly. Let

$$\theta_{\vec{v}} = (\mu_i, \sigma_{0,j}, \sigma_{en,k}, \sigma_{bs,l}, \rho_m)$$

where $\vec{v}$ is (i,j,k,l,m).

Let $\theta_\epsilon = (\mu_\epsilon, \sigma_{\epsilon,0}, \sigma_{\epsilon,en}, \sigma_{\epsilon,bs}, \rho_\epsilon)$ to denote an approximate discretized parameter signal, which is random in the Bayesian framework.

Step 1: Construct $(\theta_\epsilon, X_\epsilon)$, an approximate signal where $\epsilon = \max(\epsilon_\mu, \epsilon_{\sigma_0}, \epsilon_{\sigma_{en}}, \epsilon_{\sigma_{bs}}, \epsilon_\rho, \epsilon_x)$. 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 Markov processes. We construct a birth and death generator $\mathbf{A}_{\epsilon,v}$, such that $\mathbf{A}_{\epsilon,v} \to \mathbf{A}_v$ in (4.2). Namely, we construct a Markov chain $(\theta_\epsilon, X_\epsilon)$, where X_ϵ is a birth and death process, to approximate $(\theta, X(t))$ through the approximation of the generator, which consists the first and second order derivatives. We employ finite difference method. Specifically, we apply the central difference approximation to the differentials to construct the approximate generator as below:

$$\begin{aligned}
&\mathbf{A}_{\epsilon,v} f(\theta_{\vec{v}}, x_w) \\
=&\mu_i x_w \frac{f(\theta_{\vec{v}}, x_w + \epsilon_x) - f(\theta_{\vec{v}}, x_w - \epsilon_x)}{2\epsilon_x} \\
&+ \frac{1}{2}(\sigma_{0,j} + \sigma_{en,k} V_{en} + \sigma_{bs,l} V_{bs})^2 x_w^2 \frac{f(\theta_{\vec{v}}, x_w + \epsilon_x) + f(\theta_{\vec{v}}, x_w - \epsilon_x) - 2f(\theta_{\vec{v}}, x_w)}{(\epsilon_x)^2} \\
=&\beta_v(\theta_{\vec{v}}, x_w)(f(\theta_{\vec{v}}, x_w + \epsilon_x) - f(\theta_{\vec{v}}, x_w)) + \delta_v(\theta_{\vec{v}}, x_w)(f(\theta_{\vec{v}}, x_w - \epsilon_x) - f(\theta_{\vec{v}}, x_w)),
\end{aligned} \tag{4.4}$$

where

$$\beta_v(\theta_{\vec{v}}, x_w) = \frac{1}{2}\left(\frac{(\sigma_{0,j} + \sigma_{en,k} V_{en} + \sigma_{bs,l} V_{bs})^2 x_w^2}{\epsilon_x^2} + \frac{\mu_i x_w}{\epsilon_x}\right),$$

and

$$\delta_v(\theta_{\vec{v}}, x_w) = \frac{1}{2}\left(\frac{(\sigma_{0,j} + \sigma_{en,k}V_{en} + \sigma_{bs,l}V_{bs})^2 x_w^2}{\epsilon_x^2} - \frac{\mu_i x_w}{\epsilon_x}\right).$$

Note that $\beta_v(\theta_{\vec{v}}, x_w)$ and $\delta_v(\theta_{\vec{v}}, x_w)$ are the birth and death rates of X_ϵ respectively, and should be nonnegative. If necessary ϵ_x can be made smaller to ensure the nonnegativity.

Clearly, $\mathbf{A}_{\epsilon,v} \to \mathbf{A}_v$ and we have $(\theta_\epsilon, X_\epsilon) \Rightarrow (\theta, X)$ as $\epsilon \to 0$ by Corollary 4.8.5 on page 230 of [27].

Step 2: Obtain the normalized filtering equation of the approximate model. For the approximate signal $(\theta_\epsilon, X_\epsilon(t))$, the observations $\Phi = \{(T_i, Y_i)\}$ become $\Phi_\epsilon = \{(T_{i,\epsilon}, Y_{i,\epsilon})\}$, the observation corresponding to $(\theta_{\vec{\epsilon}}, X_{\epsilon_x}(t))$ with the stochastic intensity kernel $\lambda_\epsilon(t, y) = \bar{\lambda}_\epsilon(t) p(y|X_\epsilon(t); \theta_\epsilon)$. Then, $(\theta_\epsilon, X_\epsilon, \Phi_\epsilon, V)$ consists of the approximate model on $(\Omega_\epsilon, \mathcal{F}_\epsilon, P_\epsilon)$, and satisfies Assumptions 2.1 - 2.5.

Definition 4.1. *Let $\pi_{\epsilon,t}$ be the conditional probability mass function of $(\theta_{\vec{\epsilon}}, X_{\epsilon_x}(t))$ on the discrete state space given $\mathcal{F}_t^{\Phi_\epsilon, V}$. Let*

$$\pi_\epsilon(f, t) = E^{P_\epsilon}[f(\theta_\epsilon, X_\epsilon(t)) \,|\, \mathcal{F}_t^{\Phi_\epsilon, V}] = \sum_{\theta_{\vec{v}}, x_w} f(\theta_{\vec{v}}, x_w)\pi_{\epsilon,t}(\theta_{\vec{v}}, x_w),$$

where the summation goes over the whole state grid.

Then, the normalized filtering equation for the approximate model is given by Equations (3.9) and (3.10).

Step 3: Convert (3.9) and (3.10) to a Recursive Algorithm, which computes an approximate joint posterior.

Definition 4.2. *The posterior mass of the approximate model at $(\theta_{\vec{v}}, x_w)$ for time t is denoted by*

$$p_\epsilon(\theta_{\vec{v}}, x_w; t) = \pi_{\epsilon,t}\{\theta_\epsilon = \theta_{\vec{v}}, X_\epsilon(t) = x_w\}.$$

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

$$\mathbf{I}_{\{\theta_{\vec{v}}, x_w\}}(\theta_\epsilon, X_\epsilon(t)), \tag{4.5}$$

which takes value one when $\theta_\epsilon = \theta_{\vec{v}}$ and $X_\epsilon(t) = x_w$, and value zero otherwise. Then, the following fact emerges:

$$\pi_\epsilon\Big(\beta_v(\theta_\epsilon, X_\epsilon(t))\mathbf{I}_{\{\theta_{\vec{v}}, x_w\}}(\theta_\epsilon, X_\epsilon(t) + \epsilon_x)\Big) = \beta_v(\theta_{\vec{v}}, x_{w-1})p_\varepsilon(\theta_{\vec{v}}, x_{w-1}; t).$$

Along with similar results, Equation (3.9) becomes

$$\begin{aligned} p_\epsilon(\theta_{\vec{v}}, x_w; t_{i+1}-) = p_\epsilon(\theta_{\vec{v}}, x_w; t_i) + \int_{t_i}^{t_{i+1}-} \Big(\beta_v(\theta_{\vec{v}}, x_{w-1})p_\epsilon(\theta_{\vec{v}}, x_{w-1}; t) \\ -(\beta_v(\theta_{\vec{v}}, x_w) + \delta_v(\theta_{\vec{v}}, x_w))p_\epsilon(\theta_{\vec{v}}, x_w; t) + \delta_v(\theta_{\vec{v}}, x_{w+1})p_\epsilon(\theta_{\vec{v}}, x_{w+1}; t)\Big) dt \end{aligned} \tag{4.6}$$

If a trade at y_j, jth price level, occurs at time t_{i+1}, the updating Equation (3.10) can be written as,

$$p_\epsilon(\theta_{\vec{v}}, x_w; t_{i+1}) = \frac{p_\epsilon(\theta_{\vec{v}}, x_w; t_{i+1}-)p(y_j|x_w, \rho_m)}{\sum_{\vec{v}', w'} p_\epsilon(\theta_{\vec{v}'}, x_{w'}; t_{i+1}-)p(y_j|x_{w'}, \rho_{m'})}, \tag{4.7}$$

where the summation goes over the whole state grid, and $p(y_j|x_w,\rho_m)$, the transition probability from x_w to y_j, is specified by Equation (A.2) in Appendix A.

In the second substep, we approximate the time integral in Equation (4.6) 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 $t_{i+1}-t_i \le LL$, the length controller in the Euler scheme, then we approximate $p_\epsilon(\theta_{\vec{v}}, x_w; t_{i+1}-)$ as

$$\begin{aligned} p_\epsilon(\theta_{\vec{v}}, x_w; t_{i+1}-) &\approx p_\epsilon(\theta_{\vec{v}}, x_w; t_i) + \Big[\beta_v(\theta_{\vec{v}}, x_{w-1})p_\epsilon(\theta_{\vec{v}}, x_{w-1}; t_i) \\ -\big(\beta_v(\theta_{\vec{v}}, x_w) + \delta_v(\theta_{\vec{v}}, x_w)\big)&p_\epsilon(\theta_{\vec{v}}, x_w; t_i) + \delta_v(\theta_{\vec{v}}, x_{w+1})p_\epsilon(\theta_{\vec{v}}, x_{w+1}; t_i)\Big](t_{i+1}-t_i). \end{aligned} \tag{4.8}$$

Case 2, if $t_{i+1}-t_i > LL$, then we can choose a finer partition $\{t_{i,0}=t_i, t_{i,1}, \ldots, t_{i,n} = t_{i+1}\}$ of $[t_i, t_{i+1}]$ such that $\max_j |t_{i,j+1}-t_{i,j}| < LL$ and then approximate $p_\epsilon(\theta_{\vec{v}}, x_l;\, t_{i+1}-)$ by applying repeatedly Equation (4.8) from $t_{i,0}$ to $t_{i,1}$, then $t_{i,2}, \ldots$, until $t_{i,n}=t_{i+1}$.

Equations (4.7) and (4.8) consist of the recursive algorithm we employ to calculate the approximate posterior at time t_{i+1} for $(\theta, X(t_{i+1}))$ based on the posterior at time t_i. At time t_{i+1}, the Bayes estimates of θ and $X(t_{i+1})$ are the corresponding marginal means.

To complete the algorithm for posterior, we need to choose a reasonable prior. Assume independence between $X(0)$ and θ. The prior for $X(0)$ can be set by $P\{X(0)=Y(t_1)\}=1$ where $Y(t_1)$ is the first observed price because they are very close. For other parameters, we can simply take a uniform prior to the discretized state space of θ. At $t=0$, we select the prior as below:

$$p_\epsilon(\theta_{\vec{v}}, x_w;\, 0) = \begin{cases} \frac{1}{(1+n_\mu)(1+n_{\sigma_0})(1+n_{\sigma_{en}})(1+n_{\sigma_{bs}})(1+n_\rho)} & \text{if } x_w = Y(t_1) \\ 0 & \text{otherwise} \end{cases}.$$

We need to further find out an appropriate range and step size for a prior of each parameter. UHF data set is often large and the marginal posterior of a parameter commonly converges to a small interval around the true value because of the consistency of Bayes estimators. After spotting the small interval, we can set the uniform prior there. Then, we identify a suitable step size, which typically outputs a unique modal bell-shaped posterior distribution.

4.3.1. Consistency of the Recursive Algorithm. There are two approximations in the construction of the above recursive algorithm. One is the approximation of the time integral in (4.6) by a Euler scheme, whose convergence is well-established. The other is the approximation of the filtering equation in (3.6) and (3.7) of the true model by that in (3.9) and (3.10) for the approximated model. The weak convergence result (2) of Theorem 3.2 guarantees the consistency of the second approximation.

To be specific, let

$$p(\theta_{\vec{v}}, x_w;\, t) = \pi_t\{\theta \in N_{\theta_{\vec{v}}}, X(t) \in N_{x_w}\},$$

where $N_{\theta_{\vec{v}}}$ and N_{x_w} are appropriate neighborhoods. For example, $N_{x_w} = (x_w - \frac{1}{2}\epsilon_x, x_w - \frac{1}{2}\epsilon_x)$. Then, Theorem 3.2 implies that $p_\epsilon(\theta_{\vec{v}}, x_w;\, \cdot) \Rightarrow p(\theta_{\vec{v}}, x_w;\, \cdot)$ for all $(\theta_{\vec{v}}, x_w)$ in the state grid as $\epsilon \to 0$.

4.4. A Simulation Demonstration. The recursive algorithm for calculating the joint posterior and the Bayes estimates are fast enough to be real time. We develop a Fortran computer program for the recursive algorithms, and in this section, we demonstrate that the program works well using simulated data.

Using Model 4, we generate a simulated data series with economic news and buyer-seller initiation dummies. We choose parameter values of $\mu = 4.0E{-}8$, $\sigma_0 = 3E{-}8$, $\sigma_{en} = 2E{-}5$, and $\sigma_{bs} = 1E{-}5$ (per seconds). Based on an approximate trading year of 252 business days and 9.5 hours of trading per business day ([31]), these parameters correspond to annualized values of $\mu = 35\%$, $\sigma_0 = 8.8\%$, $\sigma_{en} = 5.9\%$, and $\sigma_{bs} = 2.9\%$. Because the trade intensity $\bar{\lambda}(t)$ is not critical in the recursive algorithms due to the exogenous sampling, we use a constant value of 0.0058 seconds/trade, which is consistent with observed trade frequencies in the Treasury market. For similar reasons, we use $\rho = 0.06$, $\alpha = 0.057$, and $\beta = 0.083$ to model the trade noise.

Given these choices, we generate a simulated time series of 2000 transactions in the OTR 10-year Treasury note market. We use a random buyer-seller initiation dummy for all the simulated observations. This choice results from the null hypothesis regarding inventory effects in the intraday Treasury note market [30]: in the absence of inventory management, the arrival time for buyer- and seller-initiated trades should be completely random. In addition, between the 1000^{th} and 1001^{st} transaction, a news announcement impulse occurs. Given the results in [35] and elsewhere, we model the 50 subsequent trades with the news dummy equal to one, after which time we assume all relevant information is reflected in the market price.

Table 4.1
Final Bayes estimates for the simulated dataset.

Parameter	True Value	Bayes Estimate	St.Error
μ	4.00E-08	4.37E-08	1.10E-08
σ_0	3.00E-05	3.03E-08	7.71E-07
ρ	0.06	0.07	0.01
σ_{en}	2.00E-05	2.27E-05	4.77E-06
σ_{bs}	1.00E-05	8.69E-06	1.25E-06

With uninformative uniform priors, at each trading time t_i, the algorithm calculates the joint posterior of $(\mu, \sigma_0, \rho, \sigma_{en}, \sigma_{bs})$, the marginal posteriors, the Bayes estimates, and the standard errors (SE). Table 4.4 provides the final results for the simulated dataset. The Bayes estimates (final posterior means) are close to their true values, each within two standard error (SE, standard deviation of the final posterior) bounds. This is in line with the consistency of the Bayes estimators for the models, which can be proved using the same method described in Appendix B of [56]. While the estimation error for μ is much larger than the other parameters, μ corresponds to the overall price trend, which will not be estimated with precision using shorter time series. Since the volatility function is the primary area of interest in financial market microstructure, μ is essentially a nuisance parameter for our purposes.

Given the simulation result, there is reason to believe the recursive algorithm can provide model identification when applied to actual market data, and provide insight into whether information-based trading, inventory management effects, or both are significant factors in the observed microstructure within the OTR 10-year Treasury note market. We explore such tests via estimation in the next subsection.

4.5. Empirical Results. We apply the estimation recursive algorithm to the tick transaction data for the OTR 10-Year Treasury note during the period 8/15/2000 to 12/31/2000, where Figure 4.1 is the time series plot with the color-distinctive shaded areas marking different news period ($V_{en} = 1$). Table 4.2 reports summary statistics for the complete set of 10006 trades in our data. Based on the relative frequencies for observed tick values and recorded trades, the estimated clustering parameters are $\alpha = .057$ and $\beta = 0.083$.

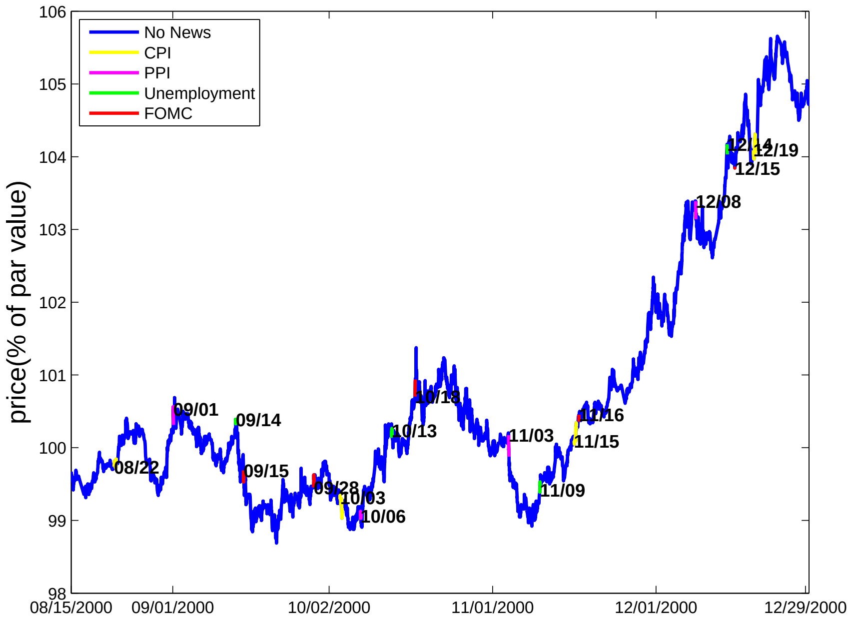

Figure 4.1. *Times Series of Prices for 10-Year Note.*

Table 4.2
Summary Statistics for the OTR 10-Year Note, 8/15/2000 - 12/31/2000

	# trades	Min	Max	Median	Mean	Std	Skewness	Kurtosis
10-Year Note	10006	98.6875	105.6563	100.0625	100.3787	1.3777	1.9014	6.3022

For V_{bs}, we use a simple dummy variable equal to one for buyer-initiated trades and 0 for seller-initiated trades; this indicator can be constructed directly from an examination of the raw GovPX data feed. For V_{en}, we follow [35] and capture the four regularly-scheduled macroeconomic news releases found to be most significant in the U.S. Treasury market: CPI, PPI, the Labor Department's unemployment report, and the Federal Reserve's rate

Table 4.3

Final Annualized Bayes Estimates for the OTR 10-Year Note, 8/15/2000 - 12/31/2000

Model	μ	σ_0	ρ	σ_{en}	σ_{bs}
Model 1	14.48%	5.01%	3.49%		
	(7.77%)	(0.046%)	(0.30%)		
Model 2	13.95%	4.93%	3.42%	5.37%	
	(8.22%)	(0.073%)	(0.40%)	(0.62%)	
Model 3	50.33%	5.30%	3.33%		-0.60%
	(6.76%)	(0.06%)	(0.48%)		(0.07%)
Model 4	50.07%	5.28%	3.21%	4.99%	-0.80%
	(7.39%)	(0.022%)	(0.60%)	(0.38%)	(0.13%)

meetings. One of these four items occurs on 18 of the 94 days in our sample period. The first three announcements are scheduled for release at 8:30AM Eastern time; for consistency with prior research in this area, we set $V_{en} = 1$ for the thirty minute period that follows. While Fed announcements are usually disclosed near 2:00PM Eastern time, the actual release time tends to vary somewhat. Following other research, we classify trades during the one-hour period 1:30PM to 2:30PM Eastern time as announcement period trades ($V_{en} = 1$) on Fed announcement days. In combination, there are 191 transactions in the data that occur with $V_{en} = 1$.

For the Models 1 - 4, we use the recursive algorithm to generate the Bayes estimates and parameter values for application to the OTR 10-year Treasury note data; Table 4.5 provides the final Bayes estimation results. The volatility parameter σ_0 is in the range of $4.9\% - 5.5\%$, consistent with historical 10-year Treasury note volatility. The dummies V_{en} and V_{bs} are both statistically significant. σ_{en} is positive and significant, consistent with the evidence that volatility in the Treasury market increases when new public information arrives ([29]). The magnitude of σ_{en} is also economically significant. In Models 2 and 4, market volatility is estimated as 95% to 110% higher during the time periods immediately following significant news releases, similar to the results found using an entirely different approach in [11].

For parameter σ_{bs}, the estimates are negative and significant for both Models 3 and 4. Given that the dummy V_{bs} is set to one for buyer-initiated trades, our finding confirms the supposition in [30] that Treasury dealers induce market volatility on the sell-side of the market due to their role as market intermediaries on behalf of the government. To clear out the inventory of their recently-acquired OTR notes and distribute them to the public, dealers are net sellers and significant inventory effects are the natural result. However, while significant, this effect is estimated to be a smaller component of the information contained in the market's observed order flow compared to macroeconomic news releases; the point estimate for σ_{bs} is just 15% of σ_0.

Overall, our empirical results strongly suggest that both information-based and inventory-based effects can explain the observed volatility microstructure in the OTR 10-year Treasury note market. Perhaps of more immediate importance, our example here shows that the filtering model and computational algorithm described in this paper can provide a new approach for statistical analysis of UHF financial market data.

5. Conclusions. In this paper, we develop a general nonlinear filtering framework for UHF data that uses MPP observations with confounding observable factors. Under our framework, complete information of UHF price data is used for statistical inference. Using stochastic filtering, we derive the normalized filtering equation to characterize the posterior measure-valued process, quantifying the parameter uncertainty. With a weak convergence theorem, we develop a general method to numerically solve the SPDE and conduct Bayes estimation. The general theory is used to study the microstructure of trading in 10 year U.S. Treasury notes, where we find that both information-based and inventory management-based motives are significant factors in the observed trade-by-trade price volatility.

Together with the model uncertainty studied in [39], this paper opens up an exciting new research area with many research topics. For Markov processes more complicated than GBM such as important benchmark models of Stochastic Volatility (SV) models and exponential Lévy models, which has tremendous financial applications in asset pricing and risk management, the current recursive algorithm based on explicit method become bottlenecked. However, the easily-parallelized recursive algorithms based on implicit methods develop in [13] and the newly-available low-cost and green GPU supercomputing power make it possible to provide real-time track and feed of SV. This can provide timely valuable market information for critical decision-making, especially, in the time most needed such as during financial turmoils. This is under investigation and one work in such direction is [14].

Besides parameter and model uncertainty, one interesting topic is to quantify algorithmic uncertainty of the recursive algorithms. Other than consistency of the Bayes estimators, we can study the related asymptotic normality and efficiency. Besides Markov chain approximation methods, other numerical methods such as particle Markov chain Monte Carlo ([5]), EM algorithms, Poisson chaos expansion methods, and their combination could be useful. Related mathematical finance problems such as (American) option pricing, portfolio selection, consumption-investment problems, optimal selling rules, quickest detection of regime switching and others are interesting but challenging, because this market not only is incomplete but also possesses incomplete information. Some of the above issues are currently under investigation.

Acknowledgements: The previous version of this paper and [39] was a working paper "Econometric Analysis via Filtering for Ultra-High Frequency Data". Part of the works were done when Yong Zeng visited and taught a special topic graduate course on *Statistical Analysis of Ultra-high Frequency Financial Data - An Overview and A New Filtering Approach* in Department of Operations Research and Financial Engineering (ORFE) in Princeton University in Spring 2007. He especially thanks Savas Dayanik for the invitation and the ORFE Department for the hospitality. We thank Yacine Aït-Saharia, Rene Carmona, Jianqing Fan, Tom Kurtz, Per Mykland, Ronnie Sircar, Wei Sun, Yazhen Wang, Jie Xiong, and Thaleia Zariphopoulou for helpful discussions. The works were presented at several seminars and conferences and we thanks participants for comments and suggestions.

Appendix A. More on Clustering Noise of Section 4. To formulate the biasing rule, we first define a classifying function $r(\cdot)$,

$$r(y) = \begin{cases} 2 & \text{if the fractional part of } y \text{ is odd } 1/64 \\ 1 & \text{if the fractional part of } y \text{ is odd } 1/32 \\ 0 & \text{if the fractional part of } y \text{ is odd } 1/16 \text{ or coarser} \end{cases}. \tag{A.1}$$

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 $p(y'|x) = P\{U = y' - R[x, \frac{1}{64}]\}$. Suppose $D = 64|y - R[x, \frac{1}{64}]|$. Then, $p(y|x)$ can be calculated as, for example, when $r(y) = 2$,

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

REFERENCES

[1] Y. Aït-Sahalia. Nonparametric pricing of interest rate derivative securties. *Econometrica*, 64:527–560, 1996.

[2] Y. Aït-Sahalia and L. Mancini. Out of sample forecasts of quadratic variation. *Journal of Econometrics*, 147:17–33, 2008.

[3] Y. Aït-Sahalia and P. A. Mykland. An analysis of Hansen-Scheinkman estimators for randomly sampled diffusions. *Journal of Econometrics*, 144:1–26, 2008.

[4] Y. Aït-Sahalia, P. A. Mykland, and L. Zhang. How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies*, 18:351–416, 2005.

[5] C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov chain Monte Carlo methods. *Journal of Royal Statistical Soceity, Series B*, 72:269 342, 2010.

[6] F. M. Bandi and J. R. Russell. Separating microstructure noise from volatility. *Journal of Financial Economics*, 79:655–692, 2006.

[7] M.J. Barclay, T. Hendershott, and K. Kotz. Automation versus intermediation: Evidence from Treasuries going off the run. *Journal of Finance*, 61:2395–2414, 2006.

[8] O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde, and N. Shephard. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76:1481–1536, 2008.

[9] R. Boel, P. Varaiya, and E. Wong. Martingales on jump processes. II: Applications. *SIAM Journal of Control*, 13:1022–1061, 1975.

[10] L. Boni and C. Leach. Expandable limit order markets. *Journal of Financial Markets*, 7:145–185, 2004.

[11] M. Brandt and K.A. Kavacejz. Price discovery in the U.S. Treasury market: The impact of order flow and liquidity on the yield curve. *Journal of Finance*, 59:2623–2654, 2004.

[12] P. Bremaud. *Point Processes and Queues: Martingale Dynamics*. Springer-Verlag, New York, 1981.

[13] B.H. Bundick, N. Rhee, and Y. Zeng. Bayes estimation via filtering equation through implicit recursive algorithms for financial ultra-high frequency data. *Statistical Inference and Interface*, 6, 2013. (in press).

[14] B.H. Bundick and Y. Zeng. Real-time stochastic volatility estimation via filtering for a partially-observed Heston model, 2013. University of Missouri at Kansas City, Working Paper.

[15] X. Chen, L.P. Hansen, and M. Carrasco. Nonlinearity and temporal dependence. *Journal of Econometrics*, 146:155–169, 2010.

[16] R. Cont and P. Tankov. *Financial Modeling with Jump Processes*. Hall/CRC, London, 2004.

[17] J. Cvitanic, R. Liptser, and B. Rozovsky. A filtering approach to tracking volatility from prices observed at random times. *Annals of Applied Probability*, 16:1633–1652, 2006.

[18] J. Cvitanic, B. Rozovsky, and I. Zaliapin. Numerical estimation of volatility values from discretely observed diffusion data. *Journal of Computational Finance*, 4:1–36, 2006.

[19] D.J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes - Vol. I: Elementary Theory and Methods*. Springer, New York, 2 edition, 2003.

[20] D.J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes - Vol. II: General Theory and Structure*. Springer, New York, 2 edition, 2008.

[21] M. H. A. Davis, A. Segall, and T. Kailath. Nonlinear filtering with counting observations. *IEEE Transactions on Information Theory*, 21:143–149, 1975.

[22] D. Duffie and P. Glynn. Estimation of continuous-time markov processes sampled at random time intervals. *Econometrica*, 72:1773 – 1808, 2004.

[23] R. Elliott and W. P. Malcolm. Nonlinear filtering with counting observations. *IEEE Transactions on Automatic Control*, 50:1123–1134, 2005.

[24] R. Engle. The econometrics of ultra-high-frequency data. *Econometrica*, 68:1–22, 2000.

[25] R. Engle and J. Russell. Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, 66:1127–1162, 1998.

[26] R. Engle and J. Russell. Analysis of high frequency data. In Y. Aït-Sahalia and L. P. Hansen, editors, *Handbook of Financial Econometrics*. North-Holland, Amsterdam, 2010.

[27] S.N. Ethier and T.G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley, New York, 1986.

[28] J. Fan and Y. Wang. Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of American Statistical Association*, 102:1349–1362, 2007.

[29] M. J. Fleming and E.M. Remolona. Price formation and liquidity in the U.S. Treasury market: The response to public information. *Journal of Finance*, 54:1901–1915, 1999.

[30] M. J. Fleming and J.V. Rosenberg. How do Treasury dealers manage their positions?, 2007. Working paper (Federal Reserve Bank of New York, Staff Papers).

[31] M.J. Fleming. The round-the-clock market for U.S. Treasury securities. *Economic Policy Review (Federal Reserve Bank of New York)*, pages 9–32, 1997.

[32] M.J. Fleming. Measuring Treasury market liquidity. *Economic Policy Review (Federal Reserve Bank of New York)*, 9:83–108, 2003.

[33] R. Frey. Risk-minimization with incomplete information in a model for high-frequency data. *Mathematical Finance*, 10:215–225, 2000.

[34] R. Frey and W. J. Runggaldier. A nonlinear filtering approach to volatility estimation with a view towards high frequency data. *International Journal of Theoretical and Applied Finance*, 4:199–210, 2001.

[35] T.C. Green. Economic news and the impact of trading on bond prices. *Journal of Finance*, 59:1201–1233, 2004.

[36] L. P. Hansen and J. Scheinkman. Back to the future: Generating moment implications for continuous-time Markov processes. *Econometrica*, 63:767–804, 1995.

[37] L. Harris. Stock price clustering and discreteness. *Review of Financial Studies*, 4:389–415, 1991.

[38] J. Hasbrouck. *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading*. Oxford University Press, USA, 2007.

[39] G.X. Hu, D. Kuipers, and Y. Zeng. Bayesian inference via filtering equation for ultra-high frequency data (II): Model selection, 2013. under review.

[40] R. D. Huang, J. Cai, and X. Wang. Information-based trading in the Treasury Note interdealer broker market. *Journal of Financial Intermediation*, 11:269–296, 2002.

[41] S.D. Jordan and D.R. Kuipers. End-of-day pricing in the U.S. Treasury market: A comparison of GOVPX and the Federal Reserve Bank of New York. *Journal of Financial Research*, 28:97–113, 2005.

[42] W.H. Kliemann, G. Koch, and F. Marchetti. On the unnormalized solution of the filtering problem with counting process observations. *IEEE Transactions on Information Theory*, 36:1415–1425,

1990.

[43] M. Kouritzin and Y. Zeng. 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, 2005.

[44] H.J. Kushner and P.G. Dupuis. *Numerical Methods for Stochastic Control Problems in Continuous Time*. Springer-Verlag, New York, second edition, 2001.

[45] J. Large. Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets*, 10:1 – 25, 2007.

[46] G. Last and A. Brandt. *Marked Point Processes on the Real Line: The Dynamic Approach*. Springer, New York, 1995.

[47] Y. Li and P. A. Mykland. Are volatility estimators robust with respect to modeling assumptions? *Bernoulli*, 13:601–622, 2007.

[48] R.S. Liptser and A.N. Shiryayev. *Statistics of Random Processes*, volume 2. Springer-Verlag, New York, 2 edition, 2002.

[49] M. O'Hara. *Market Microstructure Theory*. Blackwell, Oxford, 1997.

[50] Maria Pacurar. Autoregressive conditional duration models in finance: A survey of the theoretical and empirical literature. *Journal of Economic Surveys*, 22:711–751, 2008.

[51] P. Protter. *Stochastic Integration and Differential Equations*. Springer-Verlag, New York, 2 edition, 2003.

[52] E. Renault and B. J.M. Werker. Causality effects in return volatility measures with random times. *Journal of Econometrics*, page (in press), 2010.

[53] L. C. Scott. Bayesian inference via filtering of micro-movement multivariate stock price models with discrete noises, 2006. Doctoral Dissertation, University of Missouri at Kansas City, Mathematics and Statistics Department. (unpublished).

[54] L. C. Scott and Y. Zeng. A class of multivariate micromovement models of asset price and their bayesian model selection via filtering. In S. Ethier J. Feng and R. Stockbridge, editors, *Markov Processes and Related Topics - a Festschrift for Thomas G. Kurtz*, volume 4, pages 123–136. IMS Collection, Beachwood Ohio, 2008.

[55] D.L. Synder. Filtering and detection for doubly stochastic poisson processes. *IEEE Transactions on Information Theory*, 18:91–102, 1972.

[56] Y. Zeng. A partially-observed model for micro-movement of asset prices with Bayes estimation via filtering. *Mathematical Finance*, 13:411–444, 2003.

[57] L. Zhang. Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli*, 12:10191043, 2006.

[58] L. Zhang, P. A. Mykland, and Y. Aït-Sahalia. A tale of two time scales: Determining integrated volatility with noisy high frequency data. *Journal of the American Statistical Association*, 100:1394–1411, 2005.