Bayesian Parameter Learning via Filtering Equation for a Partially-Observed Merton's Model With Ultra-High Frequency Data: an Explicit Algorithm

Jamila Kridan[1] and Yong Zeng[2]
[1]Program of Mathematics and Statistics School of Science and Engineering
University of Missouri at Kansas City
[2]Division of Mathematical Sciences, National Science Foundation

Abstract

Merton's model, which extends the Black-Scholes model by adding a jump component to model abnormal price changes, is widely used in financial derivative markets. This paper proposes a partially observed Merton's model for ultra-high-frequency financial data, accommodating two features: random trading times and trading noises. We employ the Bayesian estimation via filtering equation approach to develop a recursive algorithm for parameter estimation/updating, which has the potential for real-time parameter learning.

Specifically, the evolution of the joint posterior distribution of the parameters of interest and the intrinsic value process (which is the Merton model) is characterized by the normalized filtering equation. This filtering equation is a stochastic partial differential equation (SPDE) and adding a jump component makes the SPDE significantly more complicated. Based on the normalized filtering equation, we develop an explicit recursive algorithm for the Bayes estimation of the model parameters and the intrinsic value process.

Key Words: Bayes estimation, marked point process, market microstructure noise, Markov chain approximation method, nonlinear filtering, partially observed model, ultrahigh frequency data

1. Introduction

With the development of the speed and capacity of computers in recording, storing, and analyzing data, it has become common to collect, study, and analyze data in the financial markets that result from the dynamic of trading. The market data contains prices during trading and gives an image of the size of the change in prices resulting from the operation of supply and demand. The market data has been called ultra-high frequency (UHF) data. UHF data have become increasingly important in the financial market because it provides market participants with a wealth of information that can be used to make better trading decisions.

Financial institutions use UHF data to develop sophisticated trading strategies that take advantage of market inefficiencies and other opportunities. This data is often used in algorithmic trading, where complex mathematical models are used to analyze market data and make trades based on the insights gained automatically. There are different views on how to model UHF data.

One view was introduced by Engle (2000), which was natural from a time series standpoint. By modeling the duration between trades as a stochastic phenomenon, UHF was viewed as irregularly-spaced time series data. This approach recognizes that traders transact at random times due to information and liquidity-based motives. The autoregressive conditional duration (ACD) model was proposed by Engle & Russell (1998) to analyze this phenomenon. Variants of the ACD model and joint models of duration and price have been developed to account for market and economic factors that impact trade timing. Pacurar (2008), Russell and Engle (2010), and Bhogal & Variyam (2019) provided surveys of these models.

Another view was from the standpoint of stochastic processes and treats the transaction observations as an observed sample path of a Marked Point Process (MPP). Zeng (2003) suggested this view, where the data were regarded as a collection of counting process points, a special case of MPP observations. In this model, there is an asset's intrinsic value process, linking to the usual models in option pricing and the empirical econometric literature for price series. The intrinsic value process cannot be observed directly and is assumed to be partially observed at random trading times through the prices, corrupted by microstructure noise. The model was then formulated into a filtering framework with counting process observations. The stochastic nonlinear filtering technique was employed to develop the Bayesian inference via filtering equations in Zeng (2003) and Kouritzin and Zeng (2005).

Hu et al. (2018a) further developed a more general model, unifying the two views in Engle (2000) and Zeng (2003). The intrinsic value process is extended to a vector Markov process, allowing the influence of other observable economic or market factors. The observation times are governed by a general point process (instead of a conditional Poisson one) with both endogenous and exogenous samplings, including but not limited to proportional hazard models, and ACD models. Other observable factors can also influence the observation times, and the noise, and the mark space is generic allowing both discrete and continuous prices.

Ultrahigh frequency data can be converted to equally-spaced high frequency data in, such as 1 or 5 minutes etc. Volatility is a key concept in financial economics, and realized volatility has been an important topic, see the textbook Aït-Sahalia & Jacod (2014). The realized volatility in the presence of market microstructure noise, which began in Zhang, Mykland, & Aït-Sahalia (2005) and Ait-Sahalia, Mykland, & Zhang (2005), has been an active research area. See recent works, for instance, Li & Mykland (2015), Chen & Mykland (2017), Aït-Sahalia & Xiu (2019), Mykland, Zhang, & Chen (2019), Chen, Mykland, & Zhang (2020), Chen, Mykland, and Zhang (2024). When the intrinsic value process has a stochastic volatility (SV) component which may include Lévy jumps, the framework in Hu et al. (2018a) is closely related to that in two-scale realized volatility framework.

In this paper, we propose a partially observed Merton's model, accommodating two features in ultrahigh frequency data: random trading times and trading noises. Merton's model is a jump-diffusion process, proposed by Rober C. Merton, a Nobel Prize laureate (1997) to further represent abnormal changes in price generated by "news." After incorporating random trading times and noises, the partial-observed Merton's model can be framed as a filtering problem with marked point process under the framework of Hu et al. (2018a). To conduct Bayes estimation, the key is the joint posterior distribution process of the parameters of interest and the intrinsic value process (which is the Merton model). The normalized filtering equation derived in Hu et al. (2018a) characterizes the evolution of the joint posterior distribution, including the parameters of interest, with the generator associated with the Merton model as the intrinsic value process. The goal of this paper is to convert the normalized filtering equation to a recursive algorithm, providing an approximate of the updated joint posterior and the updated Bayes estimates as the data streamed in.

This normalized filtering equation is a stochastic partial differential equation (SPDE) for an intrinsic value process, assuming to be a Markov process, each is characterized by a (infinitesimal) generator. Previously studied models included Black-Scholes model, a geometric Brownian motion. Merton's model adds a jump component, making the generator and the SPDE significantly more complicated. Hu et al. (2018a) proves a convergence theorem, providing a recipe for constructing consistent algorithms to numerically approximate the joint posterior distribution process. Following the recipe and based on the approximate normalized filtering equation, we construct a consistent, explicit recursive algorithm for the Bayesian learning of the model parameters and the intrinsic value process with streaming prices. The advantage of the approach is the potential real-time updates, and the disadvantage is the high computational burden, which can be mitigated by accelerated manycore GPU computing.

The rest of the paper is organized in the following manner. Section 2 presents the partially-observed Merton's model in two representations, one by constructing the prices from the intrinsic value process, and one by formulated it as a filtering problem with marked point process. Section 3 illustrates the methodology for the model (which includes the simulation, Bayes Estimation via

Filtering Equation (BEFE), and explicit recursive algorithm). Section 4 develops the conclusion. In this study we will construct a computer program that can simulate data from the model. This involves specifying the underlying assumptions and parameters of the model and using these to generate artificial data that reflects the behavior of the model. After creating the program, the next step is to validate its accuracy by comparing the simulated data with the known properties of the model. This can help to identify any errors or limitations in the program and refine its implementation accordingly. Once the program has been successfully validated, it can be applied to real asset price ultra-high frequency (UHF) data to generate predictions or insights based on the model. By using the program to analyze real-world data, we can test the validity and usefulness of the model in practical settings and potentially identify areas for further refinement or improvement. The application of our model demonstrates its potential for providing useful insights in empirical studies where complex, non-linear data is present, and where traditional statistical models may fall short.

2. The Partially Observed Merton's Model

The proposed model builds upon the work of Hu et al. (2018 a) and Merton (1976), leading to a partially observed Merton's model. There are two representations.

2.1. Construction of Price from an Intrinsic Value Process

The construction consists of modeling the components: intrinsic value process, trading times and market microstructure noises.

2.1.1. Intrinsic Value Process. Merton's original 1976 model introduced a jump-diffusion framework, where the stock price is assumed to follow a continuous diffusion process, punctuated by random jumps. He derived an explicit formula to calculate the fair price of an option under these conditions, demonstrating that it simplifies to the well-known Black-Scholes formula when jumps are absent. The Merton's model consists of three components after taking the logarithm of the intrinsic value process X(t):

1) Linear Drift represents the average expected return of the stock over time.
2) Brownian Motion (BM) is for normal price fluctuations, such as those caused by momentary mismatches between supply and demand, or other new information that cause small changes in the value of a share. In essence, the effect of this information on the share price per unit of time results in a small change in the price.
3) Jump Component (Poisson process) for the behavior of abnormal price changes captures the impact of news on security prices. News that cause upward jump in prices are called "good news" and news that cause downward jump in prices – "bad news." Upon the arrival of news, jump magnitudes are determined by sampling from an independent and identically distributed (IID) random variable.

The dynamics of the stock price X in the Merton model can be described by the following stochastic differential equation:

$$\frac{dX_t}{X_t} = d\log X_t = \mu dt + \sigma\, dB_t + \left(\xi_{N_t} - 1\right)dN_t \tag{1}$$

where
dX/X: represents the relative change in the stock price.
X(t): represents the price of the financial asset at time t.
dX: represents the change in the stock price (price of the asset).
μ: is the instantaneous expected return of the stock (asset) per unit time (excluding jump component).
σ: is the instantaneous volatility (standard deviation of returns) of the asset, conditional on no news occurs

B_t: represents a Wiener process (standard Brownian Motion).
ξ_{N_t}: represents jump magnitudes, and they are assumed to be i.i.d. with lognormal distribution:

$$\xi_i \sim \text{ lognormal or } \eta_i = \log \xi_i \sim N(\alpha_J , \beta_J^2)$$

N (t): represents a Poisson process (counting process) which is the number of jumps up to time t with intensity of jumps λ_J .(frequency of jumps per unit time). N (t) and B_t are assumed to be independent.

With the notation

$$\xi_i(N_t) = \begin{cases} 1 & if \quad N(t) = 0 \\ \prod_{i=1}^{N(t)} \xi_i & if \quad N(t) = 1,2,.. \end{cases}$$

Applying Ito's formula, we can obtain the explicit expression of X(t)

$$X_t = X_0 \; exp\left\{\left(\mu - \frac{1}{2}\sigma^2 \right)t + \sigma B_t\right\} \xi_i(N_t)$$

The construction of price from an intrinsic value process is likened to a state-space model. To conduct Bayes estimation, we include θ, a vector of the parameters in the state. Let us recall the Assumption 1 in Hu et al. (2018a).

Assumption 1: The solution of a martingale problem for a generator A is (θ, X), which is a $p + m$ dimension vector Markov process in which

$$\mathrm{M}_f(t) = f(\theta(t), X(t)) - \int_0^t \boldsymbol{A} \; f(\theta(s), X(s))ds$$

Is a $\mathcal{F}_t^{\theta,X}$ -martingale, which is the $\sigma -$ algebra generated by $(\theta, X)_{0\le s\le t}$ and f is in the domain of $\boldsymbol{A}$.
Remark 1: For the partially-observed Merton's model, $m = 1$ and $p = 6$. $\theta = (\mu, \sigma, \rho, \mu_J, \sigma_J, \rho_J)$ and the corresponding generator $\boldsymbol{A}$ is given by

$$\boldsymbol{A}f(\theta, x) = \boldsymbol{A}_1 f(\theta, x) + \lambda \int [f(xz) - f(x)] \; p_\xi(z) \; dz$$

with

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

Note that $\boldsymbol{A}_1 f(\theta, x)$ is the generator for Black-Scholes model (geometric Brownian motion). The additional integral part describes the jump activities in Merton's model.

2.1.2. Trading times

We assume $T_1, T_2, \ldots, T_i, \ldots$ to be a general non-explosive point process, characterized by a non-negative F_t-predictable stochastic with a stochastic intensity $\lambda(t) \equiv \overline{\lambda}\,(\theta(t), X(t), \Phi^{t-}, t)$ where $\Phi = \{T_i, Y_i\}_{i>1}$ is a sequence of mark points and $\Phi^{t-} = \{(T_i, Y_i),\; i < t\}$ denotes the sample path of Φ up to time t-.

2.1.3. Market Microstructure Noise

The asset price at trading time T_i, denoted as $Y_i = Y(T_i)$, takes a value in the mark space $\mathbb{Y}$ and can written as $Y(T_i) = F(X(T_i))$, where y = F (x) represents a stochastic transformation from $x = X(T_i)$ to $y = Y_i$, and is defined by a transition probability distribution $p(Y|X(T_i))$. The stochastic transformation $F(\cdot)$ incorporates market microstructure noise. Specifically, the conditional distribution of Y_i given $X(T_i) = x$ is expressed as $p(y|x;t) = p(y|X(T_i);\ \theta(t), \Phi^{t-}, t)$. In essence, the observation set $\{T_i, Y_i\}\}$ constitutes a marked point process (MPP) governed by the predictable stochastic intensity kernel $\overline{\lambda}\,(\theta(t), X(t), \Phi^{t-}, t)\ p(y|X(T_i);\ \theta(t), \Phi^{t-}, t)$.

This representation can be regarded as a state-space model with random arrival times, or a hidden Markov model with random arrival times in a general mark space.

In the financial market when use Ultra-High Frequency (UHF) data, the process X represents the intrinsic value vector of m assets, emphasizing the straightforward economic idea that $Y(T_i)$ (the observed price at trading time T_i) is derived from $X(T_i)$ (the intrinsic value), but is effected by trading noise modeled by F (the random transformation).

In trading there are three types of market microstructure noise that occur: discreet, clustering, and non-clustering noise.
First, the discreet micromovement noise appears when the price moves discreetly intraday (tick-by-tick). which is determined by a random transition function $Y(T_i) = F(X(T_i))$
such that function $Y(T_i) = Round\left[X(T_i), \frac{1}{M}\right]$ where M is the size of the tick.
Second, the clustering noise, which occurs when the prices do not appear evenly on all ticks, which is determined by $Y(T_i) = Rooound\left[X(T_i) + V_i, \frac{1}{M}\right]$ where V_i double geometric distribution with mass function.

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

Third, the non-clustering noise involves undesignated outliers, whish is added by biasing function b (.) such that

$$Y(T_i) = b\left(Round\left[X(T_i), \frac{1}{M}\right]\right)$$

In the stock price the biasing function for the tick 1/100 occurs multiples of ten cents are more common than odd multiples of five cents, odd multiples of five cents are second more common and other fraction are rarely common. the biasing function construes the biasing rules: if the fraction part of y′ is a multiple of five cents then y stays on y′ withe probability 1; if the fractional part of y′ other than a multiple of five cents then y stays on y′ withe probability $1 - \alpha - \beta$; y moves to the closest multiple of ten cents with probability β; y moves to the closest odd multiple of five cents with probabilityα. Then we can calculate $p(y|x)$ such that:

$$p(y|x) = \sum_{\grave{y}} p(y|y')\, p(y'|x)$$

2.1.4. Simulation The simulated data of Partially Observed Merton Model was generated using

$$X_{t_i} = X_{t_{i-1}} \exp\,(\mu - \frac{1}{2}\sigma^2(t_i - t_{i-1}) + \sigma z\sqrt{(t_i - t_{i-1})}\ (\xi_i)$$
$$Y(t_i) = b\left(Round\left[X(t_i), \frac{1}{M}\right]\right)$$

The simulation involves the generation of various components:
Standard Normal (z), Time Generation (t) Jump Magnitudes (ξ_i) and non-clustering Noise. We assume that the trade durations follow an exponential distribution. The simulation data reflects the properties of the actual data for Stock price. Figure 1 below provides a simulated sample path of the model.

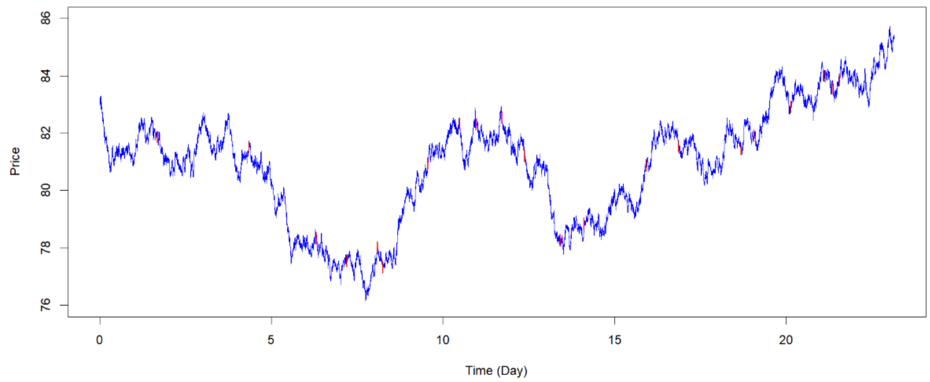

Figure 1: Simulated data of daily stock price with jump size (red) over one month

2.2. Filtering with MPP Observations

In a filtering framework, there are signal and observations corrupted by noise. Here, (θ, X) is the signal process that is not directly observable but partially observed through the marked point process (MPP), $\Phi = \{T_i, Y_i\}_{i>1}$, representing a sequence of marked points at event times T_i with mark Y_i describing the event. Y_i lives in a mark space $\mathbb{Y}$, where $(\mathbb{Y}, \mu)$ is a measure space with $\mu(\mathbb{Y}) < \infty$.

Under this setup, there are five assumptions made in Hu et al. (2018a), guaranteeing the second representation is the same as the first in the sense of having the same probability measure. Assumption 1 is given in 2.1.1, and the other four are given below.

Assumption 2: Under the physical measure P, the stochastic intensity kernel of Φ is given by $\lambda(t, y) = \overline{\lambda}(t)p(y|X(t))$, where $p(y|X(t))$ is the transition probability from the value *x* to price *y*.

Assumption 3: $\exists$ a reference measure Q such that $P \ll Q$ and under Q, (θ, X) and $\Phi = \{T_i, Y_i\}_{i>1}$, are independent. the compensator of MPP Φ is $\gamma_Q\big(d(t, y)\big) = \mu(dy)dt$, namely, the corresponding stochastic intensity kernel is $\mu(dy)$.

Define the ratio of the mark distributions under P and Q as $r(y) = p(y|X(t))/\mu(dy)$.
Define the ratio of the compensators under P and Q:

$$\omega(t, y) = \frac{\gamma P\big(d(t, y)\big)}{\gamma Q\big(d(t, y)\big)} = \frac{\lambda(t, dy)}{\mu(dy)} = \overline{\lambda}(t)r(y)$$

The Continuous-time Joint Likelihood

Let $L(t) = \frac{dP}{dQ}(t)$ be the Radon-Nikodym derivative given in (2) below.

$$L(t) = L(0)exp\left\{\int_0^t \int_{\mathbb{Y}} \log \omega(s-.y)\Phi(d\big(d(s, y)\big) - \int_0^t \int_{\mathbb{Y}} [\omega(s-.y) - 1]\,\mu(dy)ds\right\} \qquad (2)$$

Assumption 4: $\omega(t, y)$ satisfies: $E^Q[L(t)] = 1\, for\ \ T > 0$.

Assumption 5: $\int_0^t E^P[\overline{\lambda}(s)]ds\ < \infty\, for\ t > 0$.
Assumption 4 guarantees the existence of the reference measure Q and Assumption 5 ensure the non-explosiveness of the point process of trading.

3. Bayes Estimation via Filtering Equation

Bayesian Inference via Filtering Equations is a powerful method for parameter and model learning. A flowchart from Hu et al. (2018 a) below depicts the proposed methodology. The filtering model with MPP observation is described in Section 2.2. A related filtering equation, a weak convergence theorem, and a general recipe to construct a consistent recursive algorithm are reviewed in this section.

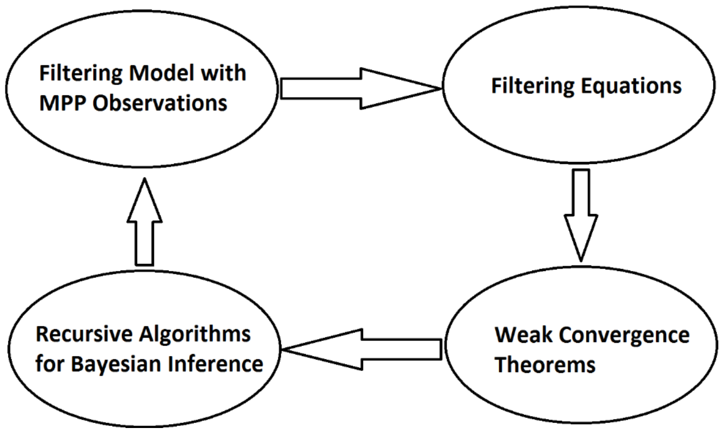

Figure 2: A flow chart of the proposed methodology

3.1. The Posterior Distribution Process of $(\theta(t), X(t))$ and the Normalized Filtering Equation

We first recalls related definitions and then present the normalized filtering equation.

Definition 1: Let π_t be the conditional distribution of $(\theta(t), X(t))$ given $\mathcal{F}_t^{\Phi}$ under the physical measure P. Then, π_t is the conditional distribution of $(\theta(t), X(t))$ given $\mathcal{F}_t^{\Phi}$. π_t becomes the posterior after a prior is assigned.

$$\pi(f,t) = E^P\left[f\big(\theta(t), X(t)\big)\middle|\mathcal{F}_t^{\Phi}\right] = \int f(\theta, X)\, \pi_t(d\theta, d\,X)$$

Definition 2: Let ρ be the conditional expectation of $((\theta(t), X(t))$ given $\mathcal{F}_t^{\Phi}$ under the reference measure Q. Let

$$\rho(f,t) = E^Q\left[f\big(\theta(t), X(t)\big)L(t)\middle|\mathcal{F}_t^{\Phi}\right] = \int f(\theta, X)\, \rho_t(d\theta, d\,X)$$

The Bayes Theorem, also called the Kallianpur-Striebel formula, gives $\pi(f,t) = \frac{\rho(f,t)}{\rho(1,t)}$.

Theorem 1: Under assumptions 1 – 5, π_t is the unique measure-valued solution of the following SPDE, called the normalized filtering equation:

$$\pi(f,t) = \pi(f,0) + \int_0^t \pi(Af,s)ds$$
$$+ \int_0^t \int_Y \left[\frac{\pi(r(y)f, s-)}{\pi(r(y), s-)} - \pi(f, s-)\right] \Phi(ds, dy) - \pi(\omega(y), s)\, \mu(dy)ds)$$

where $\omega(y)) = \omega(y,t) = \overline{\lambda}(t) r(y)$ the ratio of compensators under P and Q. Hu et al. (2018 a).

Separation of the Normalized Filtering Equation

Remark 2: When $\overline{\lambda}(t)$ is $\mathcal{F}_t^{\Phi}$ -predictable, that is the case here, the filtering equation can be simplified as:

$$\pi(f,t) = \pi(f,0) + \int_0^t \pi(Af,s)ds + \int_0^t \int_Y \left[\frac{\pi(r(y)f,s-)}{\pi(r(y),s-)} - \pi(f,s-)\right] \Phi(d(s,y))$$

where *r(y)* is the ratio of mark distributions under P and Q.

Separate the above into propagation equation:

$$\pi(f,t_{i+1}) = \pi(f,t_i =) + \int_{t_i}^{t_{i+1}-} \frac{\pi(r(y)f,t_{i+1}-)}{\pi(r(y),t_{i+1}-)} \quad (3)$$

and the updating equation:

$$\pi(f,t_{i+1}) = \frac{\pi(r(y)f,t_{i+1}-)}{\pi(r(y),t_{i+1}-)} \quad (4)$$

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

3.2. A Weak Convergence Theorem

Theorem 1 provides the normalized filter that plays a pivotal role in characterizing posterior measure-valued processes. This equation is of importance given the inherent infinite-dimensionality of these processes. However, the practical computation of such processes necessitates a conversion of this problem from an infinite-dimensional to a finite-dimensional context. This conversion, in turn, leads to the development of a recursive algorithm, underpinned by the fundamental requirement of consistency. This implies that the computed approximate posterior distribution-valued process must converge to the true one.

The weak convergence theorem, reviewed below from Hu et al (2018), furnishes a framework for constructing consistent recursive algorithms using Kushner's Markov Chain approximation methods. See Kushner & Dupuis (2001).

Recall that $X_\varepsilon \Longrightarrow X$ denotes that X_ε converges weakly to X in the Skorohod topology as $\epsilon \to 0$. Such weak convergence is uniform in time.

Theorem 2: Suppose that (θ, X, Φ) on the probability space $(\Omega, \mathcal{F}, P)$ with assumptions 1-5. Suppose that for any $(\theta_\epsilon, X_\epsilon, \Phi_\epsilon)$ have on $(\Omega_\epsilon, \mathcal{F}_\epsilon, P_\epsilon)$ also with Assumptions $1-5$. If $(\theta_\varepsilon, X_\varepsilon) \Longrightarrow (\theta, X)$ as $\epsilon \to 0$, then, for any bounded continuous functions, f, we have, under the physical measures, as $\epsilon \to 0$,
(1) $\Phi_\epsilon \Rightarrow \Phi$, and
(2) $\pi_\varepsilon(f,\cdot) = \pi(f,\cdot)$.

3.3. A Recipe to Construct Consistent Recursive Algorithms

Following Theorem 2, there are three primary steps involved in constructing a consistent recursive algorithm.

Step1: Markov chain approximation $(\theta_\epsilon,\ X_\epsilon)$
Generate a Markov chain approximation $(\theta_\epsilon,\ X_\epsilon)$ to approximate (θ, X).
Define $p_\epsilon(y) = p\big(y_\epsilon | x_\epsilon(t)\big)$ as an approximation to $p(y) = p\big(y|x(t)\big)$.

Step 2: Approximate filter of $\pi_\epsilon(f,t)$:

Utilize theorem 1 to obtain an approximation filter corresponding to $(\theta_\epsilon, X_\epsilon, \Phi_\epsilon, \mathrm{p}_\epsilon)$, namely, $\pi_\epsilon(f,t)$. Similarly, this approximation filter can be broken down into the propagation equation (5) and the updating equation (6).

The propagation equation approximate filter $(\theta_\epsilon, X_\epsilon, \Phi_\epsilon, \mathrm{p}_\epsilon)$ is

$$\pi_\epsilon(f, t_{i+1^-}) = \pi_\epsilon(f, t_i) + \int_{t_i}^{t_{i+1}-} \pi_\epsilon(\boldsymbol{A}_\epsilon f, s)\, ds \tag{5}$$

The updating equation assuming y_{i+1} mark occurring at time t_{i+1} is

$$\pi_\epsilon(f, t_{i+1}) = \frac{\pi_\epsilon(f\mathrm{p}_\epsilon(y_{i+1}), t_{i+1^-})}{\pi_\epsilon(\mathrm{p}_\epsilon(y_{i+1}), t_{i+1^-})} \tag{6}$$

Step 3: Recursive Algorithm on the Discrete Grid of State Space: Converting (5) and (6) into a recursive algorithm is carried out through two distinct substeps:

(1) Represent $\pi_\epsilon(.,t)$ with the component correspond to $\pi_\epsilon(f,t)$ for a lattice-point indicator f.
(2) Approximate the time integral in (6) with q Euler scheme or other numerical schemes.

4. Construct A Recursive Algorithm for the Partially Observed Merton's Model

We follow the three-step recipe to construct a consistent recursive algorithm using the explicit method in numerical partial differential equation (PDE). Before that, we discretize the seven-dimension state space of $(\theta, X(t))$ to form a state grid.

4.1. A Markov Chain Approximation to Signal

We first constructing an approximate state process, $(\theta_\epsilon, X_\epsilon(t))$. This involves the construction of a Markov chain approximation to $(\theta, X(t))$. This approximation can subsequently be leveraged to formulate a Markov chain generator, denoted as A_ϵ, with the property that A_ϵ converges pointwise to A as ϵ tends to 0 point-wisely. This convergence implies that $(\theta_\epsilon,\ X_\epsilon(t)) \Rightarrow (\theta, X(t))$ according to Corollary 4.8.5 on Ethier & Kurtz (1986).

Recall that the generator of Merton's model (1) is

$$\boldsymbol{A}f(\theta,x) = \boldsymbol{A}_1 f(\theta,x) + \lambda \int [f(\theta,xz) - f(\theta,x)]\ p_\xi(z)\, dz$$

where $\theta = (\mu, \sigma, \rho, \mu_J, \sigma_J, \rho_J)$, $p_\xi(z)$ is the density of of ξ_i which is log-normal, and

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

For the diffusion generator, $\boldsymbol{A}_1$, we approximate the first and second order differentials by the first and second order central differences as below.

$$\begin{aligned}\boldsymbol{A}_{1,\varepsilon} f(\theta,x) = \mu X &\left(\frac{f(\theta,x+\epsilon_x) - f(\theta,x-\epsilon_x)}{2\,\epsilon_x}\right) \\ &+ \frac{1}{2}\sigma^2 X^2 \left(\frac{f(\theta,x+\epsilon_x) + f(\theta,x-\epsilon_x) - 2f(\theta,x)}{\epsilon_x^2}\right)\end{aligned} \tag{7}$$

$$= a_x(\theta,x)\big(f(\theta,x+\epsilon_x) - f(\theta,x)\big) + b_x(\theta,x)\big(f(\theta,x-\epsilon_x) - f(\theta,x)\big)$$

where

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

and

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

Then, the approximate generator becomes

$$\boldsymbol{A}_\varepsilon f(\theta,x)=\boldsymbol{A}_{1,\varepsilon} f(\theta,x)+\ \lambda\int f(\theta,xz)\,p_\xi(z)\,dz-f(\theta,x)$$

We will approximate the integral part of the generator for jumps in the approximate filtering equation.

4.2. The Filter Equation of the Approximate Model

With the above expression of $\boldsymbol{A}_\varepsilon$, $\pi_\epsilon(f,t)$, as expressed in equations (5) becomes

$$\pi_\epsilon(\boldsymbol{A}_\varepsilon f(\theta,x_t),t)=\pi_\epsilon\big(\boldsymbol{A}_{1,\varepsilon} f(\theta,x_t),t\big)+\lambda\,\pi_\epsilon\Big(\int f(\theta,x_t\,z)\;p_\xi(z)\,dz,t\Big)$$
$$-\lambda\pi_\epsilon\big(f(\theta,x_t),t\big)\qquad(8a)$$

The posterior mass function of the approximate model, represented as $(\theta_\epsilon,X_\epsilon,\Phi_\epsilon)$, at time t is denoted by:

$$\mathrm{p}_\epsilon(\theta_\epsilon,X_\epsilon;t)=P\big[\big(\theta_\epsilon=\theta,X_t(t)=x_J\big)|\mathcal{F}_t^{\Phi_\epsilon}\big]$$

We are going to take f as the lattice-point indicator, which is defined below.

Definition 3: The indicator function is defined as

$$I_{\{\theta_j,x_l\}}\big(\theta_\epsilon,X_\epsilon(t)\big)=\begin{cases}1 & if\quad \theta_\epsilon=\theta_j,X_\epsilon=x_l\quad iff\quad X(t)\in\left\{x_l-\frac{1}{2}\epsilon_x,x_l+\frac{1}{2}\epsilon_x\right\}\\ 0 & \qquad o.w\end{cases}$$

and we can express it as

$$I_{\{\theta_\varepsilon=\theta_j,\,X_\epsilon(t)=x_l\}}\big(\theta_\epsilon,X_\epsilon(t)\big)\overset{def}{=}I_{\{\theta_j,x_l\}}\big(\theta_\epsilon,X_\epsilon(t)\big)$$

Note that

$$\pi_\epsilon\left(I_{\{\theta_j,x_l\}}(\theta_\epsilon,X_\epsilon)t\right)=p_\epsilon(\theta_j,x_l;t)$$

Next, we approximate the integral component in Equation (8):

$$\begin{aligned}\pi_\epsilon\left(\int I_{\{\theta_j,x_l\}}(\theta_\epsilon,X_\epsilon z)p_\xi(z)dz,t\right)&=\int\pi_\epsilon\;\left(I_{\{\theta_j,x_l\}}(\theta_\epsilon,X_\epsilon z),t\right)p_\xi(z)dz\\&=\int\sum_{l',j'}I_{\{\theta_j,x_l\}}\big(\theta_{j'},x_{l'}z\big)p_\epsilon\big(\theta_{j'},x_{l'};t\big)p_\xi(z)dz\\&=\sum_{l'}p_\epsilon\big(\theta_j,x_{l'};t\big)\int I_{\{\theta_j,x_l\}}\;\big(\theta_j,x_{l'}z\big)\;p_\xi(z)dz\\&=\sum_{l'}p_\epsilon\big(\theta_j,x_{l'};t\big)\,\hat{p}\;(\xi\in D_{l',l})\end{aligned}$$

where $\hat{p}\;(\xi\epsilon D_{l',l})$)is the probability of the set of ξ jump size, to make $x_{l'}z$ into the neighborhood of x_l. Then, Equation (8a) with f as the indicator becomes

$$\pi_\epsilon\left(\boldsymbol{A}_\varepsilon I_{\{\theta_j,x_l\}}(\theta_\epsilon, X_\epsilon), t\right) = \pi_\epsilon\left(\boldsymbol{A}_{1,\varepsilon} I_{\{\theta_j,x_l\}}(\theta_\epsilon, X_\epsilon), t\right) + \lambda \sum_{l'} p_\epsilon(\theta_j, x_{l'}; t)\, \hat{p}(\xi \in D_{l',l}) - \lambda\, p_\epsilon(\theta_j, x_l; t)$$

Recall, from Bundick, Rhee, & Zeng (2013), the diffusion part is expressed as

$$\begin{aligned}\pi_\epsilon(\boldsymbol{A}_{1,\varepsilon} I, t) = a(\theta_j, x_{l-1}) p_\epsilon(\theta_j, x_{l-1}; t) - \left(a(\theta_i, x_l) + b(\theta_j, x_l)\right) p_\epsilon(\theta_j, x_l; t) \\ + b(\theta_j, x_{l+1}) p_\epsilon(\theta_j, x_{l+1}; t) + \lambda \sum_{l'} p_\epsilon(\theta_j, x_{l'}; t)\, \hat{p}(\xi \in D_{l',l}) \\ - \lambda\, p_\epsilon(\theta_j, x_l; t) \qquad (8b)\end{aligned}$$

By Propagation equation (5), we have

$$p_\epsilon(\theta_j, x_l; t_{i+1^-}) = p_\epsilon(\theta_j, x_l; t_i) + \int_{t_i}^{t_{i+1^-}} \pi_\epsilon(\boldsymbol{A}_\varepsilon I, s)\, ds \qquad (8c)$$

where $\pi_\epsilon(\boldsymbol{A}_\varepsilon \boldsymbol{I}, s)$ is given by Equation (8b).

Assuming y_{i+1} mark occurring at time t_{i+1}, Equation (6) becomes

$$p_\epsilon(\theta_j, x_l, t_{i+1}) = \frac{p_\epsilon(\theta_j, x_l, t_{i+1^-}) p(y_{j+1} | x_l, \theta_j)}{\sum_{l',j'} p_\epsilon(\theta_{j'}, x_{l'}; t_{i+1^-})\, p(y_{j+1} | x_{l'}, \theta_{j'})} \qquad (9)$$

Equations (8c) with (8b) and (9) are the core equations in further deriving an efficient recursive algorithm.

4.3. An Explicit Recursive Algorithm

We approximate the time integral in Equation (8c) with an Euler scheme. Such an algorithm is called explicit algorithm in numerical PDE, and we can such one explicit recursive algorithm. After ruling out the probability-zero event that two or more jumps take place simultaneously, there are two possible cases for the inter-trading time.

First case: If $t_{i+1} - t_i \leq LL$, where LL is a length controller, from equation (9) we can approximate $p_\epsilon(\theta_j, x_l; t_{i+1^-})$ as

$$p_\epsilon(\theta_j, x_l; t_{i+1^-}) \approx p_\epsilon(\theta_j, x_l; t_i) + \pi_\epsilon(\boldsymbol{A}_{1,\varepsilon} I, t_i)(t_{i+1^-}\ t_i) \qquad (10)$$

Substituting $\pi_\epsilon(\boldsymbol{A}_{1,\varepsilon} I, t)$ as Equation (8b), we obtain

$$\begin{aligned}&p_\epsilon(\theta_j, x_l; t_{i+1^-}) \\ &\approx p_\epsilon(\theta_j, x_l; t_i) + a(\theta_j, x_{l-1}) p_\epsilon(\theta_j, x_{l-1}; t_i)(t_{i+1^-}\ t_i) \\ &- (a(\theta_j, x_l)\_b(\theta_j, x_l))\, p_\epsilon(\theta_j, x_l; t_i)(t_{i+1^-}\ t_i) b(\theta_j, x_{l+1}) p_\epsilon(\theta_j, x_{l+1}; t_i)\, (t_{i+1^-}\ t_i) \\ &- \lambda p_\epsilon(\theta_j, x_l; t_i)(t_{i+1^-}\ t_i) \\ &+ \lambda \sum_{l'} p_\epsilon(\theta_j, x_{l'}; t)\, \hat{p}(\xi \in D_{l',l})(t_{i+1^-}\ t_i) \qquad (11a)\end{aligned}$$

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

$$
\begin{aligned}
p_\epsilon(\theta_j, x_l; t_{i+1}-) \approx\ & p_\epsilon(\theta_j, x_l; t_i) \\
& + \Big[a(\theta_j, x_{l-1}) p_\epsilon(\theta_j, x_{l-1}; t_i) - \big(a(\theta_j, x_l) - b(\theta_j, x_l)\big)\, p_\epsilon(\theta_j, x_l; t_i)\big) \\
& + b(\theta_j, x_{l+1}) p_\epsilon(\theta_j, x_{l+1}; t_i) - \lambda p_\epsilon(\theta_j, x_l; t_i) \\
& + \lambda \sum_{l'} p_\epsilon(\theta_j, x_{l'}; t)\, \hat{p}(\xi \in D_{l',l})\Big] \Delta \qquad (11b)
\end{aligned}
$$

Second case: If $t_{i+1} - t_i > LL$, we can select a fine partition $\{\{t_{i,0} = t_i, t_{i,1} \ldots t_{i,n} = t_{i+1}\}$ of $[t_i, t_{i+1}]$ such that max $|t_{i,j+1} - t_{i,j}| <$ < LL and approximate $p_\epsilon(\theta_j, x_l; t_{i+1}-)$ by repeatedly applying the recursive algorithm described by equation (11b) from $t_{i,0}$ *to* $t_{i,1}$, then $t_{i,2}$ ... until $t_{i,n} = t_{i+1}$ important to note that $p_\epsilon(\theta_j, x_l; t_{ij}-) = p_\epsilon(\theta_j, x_l; t_{ij})$ for all j except j = n, as $p_\epsilon(\theta_j, x_l; t)$ is continuous in t within the interval $[t_i\ , t_{i+1})$.

Equations (9) and (11b) together form for the explicit recursive algorithm to compute the joint posterior of the approximate model, $(\theta_\epsilon, X_\epsilon(t))$, over time. With the joint posterior of the approximate model, we can easily compute the marginal means and standard deviations, which are the Bayes estimates and their standard errors.

5. Conclusion and Future Works

In this paper, we propose a partially observed Merton's model for ultra-high-frequency financial data, study the Bayes parameter learning via filtering equation, and derive an explicit recursive algorithm for updating an approximate joint posterior distribution, allowing the computation of the Bayes estimates and related uncertainty quantification.

An advantage of the explicit algorithm is that it is easily parallelizable. There is a drawback for the explicit recursive algorithm when the stock price tick size becomes 1/100. Namely, the explicit recursive algorithms become highly inefficient from a computational perspective. That is due to the posterior masses are required to be nonnegative. Keeping the nonnegativity requires an impractical number of steps to compute the joint posterior before the next trade, making real-time computation impossible. This inefficiency is the primary limitation of the explicit recursive algorithm. To address this curse of step-size, we will employ the implicit method in a future work to derive an implicit recursive algorithm, which is designed to operate efficiently with small tick size such as 1/100. We plan to develop a computer program for the algorithm, validate it on simulated data sets, and apply it to real UHF financial data.

Hu et al. (2018b) studies the Bayesian model selection via filtering equations for two or more models. Moreover, we plan to do the same to conduct Bayesian model selection for partially-observed Merton's and Black-Scholes' models.

Acknowledgments. Yong Zeng acknowledges that this material is based upon work supported by (while serving at) the National Science Foundation.

References

Aït-Sahalia, Y., & Xiu, D. (2019). A Hausman test for the presence of noise in high frequency data. Journal of Econometrics, 211, 176-205.

Bhogal, S. K., & Thekke Variyam, R. (2019). Conditional duration models for high-frequency data: a review on recent developments. Journal of Economic Surveys, 33 (1), 252–273.

Bundick, B., Rhee, N., & Zeng, Y. (2013). Bayes estimation via filtering equation through implicit recursive algorithms for financial ultra-high frequency data. Statistics and Its Interface, 6, 487 – 498.

Chen, D., Mykland, P. A., & Zhang, L. (2020). The five trolls under the bridge: Principal component analysis with asynchronous and noisy high frequency data. Journal of the American Statistical Association, 115, 1960–1977.

Chen, D., Mykland, P.A., and Zhang, L. (2024). Realized regression with asynchronous and noisy high frequency and high dimensional data. Journal of Econometrics, 239, article 105446.

Chen, R. Y., & Mykland, P. A. (2017). Model-free approaches to discern non-stationary market microstructure noise and time-varying liquidity in high frequency data. Journal of Econometrics, 200 (1) 79-103.

Engle, R. (2000). The econometrics of ultra-high-frequency data, Econometrica, 68, 1–22.

Engle, R., & Russell, J. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. Econometrica, 66, 1127–1162.

Ethier, S. N. & Kurtz, T. G. (1986). *Markov Processes: Characterization and Convergence*, Wiley, New York.

Figueroa-López, J. E., Lancette, S. R., Lee, K., & Mi, Y. (2011). Estimation of NIG and VG models for high frequency financial data. Handbook of Modeling High-Frequency Data in Finance, 1–25.

Fleming, Thomas R., and David P. Harrington. (2013). Counting processes and survival analysis. Vol. 625. John Wiley & Sons.

Hu, G. X., Kuipers, D. R., & Zeng, Y. (2018a). Bayesian inference via filtering equations for ultrahigh frequency data (I): Model and estimation. SIAM/ASA Journal on Uncertainty Quantification, 6 (1), 34–60.

Hu, G. X., Kuipers, D. R., & Zeng, Y. (2018b). Bayesian inference via filtering equations for ultrahigh frequency data (II): Model selection. SIAM/ASA Journal on Uncertainty Quantification, 6 (1), 61–86.

Kouritzin, M. A., & 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 (1), 97–121.

Kushner, H. J., & Dupuis, P. (2001). Numerical methods for stochastic control problems in continuous time,2nd ed. Springer.

Li, Y., & Mykland, P. A. (2015). Rounding errors and volatility estimation. Journal of Financial Econometrics, 13 (2), 478–504.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. Journal of Financial Economics, 3, pp. 25-144.

Mykland, P. A., Zhang, L., & Chen, D. (2019). The algebra of two scales estimation, and the s-tsrv: High frequency estimation that is robust to sampling times. Journal of Econometrics, 208 (1), 101–119.

Pacurar, M. (2008). Autoregressive conditional duration models in finance: A survey of the theoretical and empirical literature. Journal of Economic Surveys, 22 (4), pp. 711–751.

Ramezani, C. A., & Zeng, Y. (2007). An empirical assessment of the double exponential jump-diffusion process. Annals of Finance, 3, 487 – 507.

Russell, J. R. and Engle, R. F. (2010), Analysis of high frequency data, in Handbook of Financial Econometrics, Aït-Sahalia, Y., and L. P. Hansen, eds., North-Holland, Amsterdam, 2010, 383-426.