ECONOMETRIC ANALYSIS VIA FILTERING FOR ULTRA-HIGH FREQUENCY DATA

XING HU, DAVID R. KUIPERS AND YONG ZENG

We propose a general nonlinear filtering framework with marked point process observations incorporating other observable economic variables for ultra-high frequency (UHF) data. The approach generalizes several existing models and provides extensions in new directions. We derive filtering equations to characterize the evolution of the statistical foundation such as likelihoods, posteriors, Bayes factors and posterior model probabilities. Given the computational challenge, we provide a powerful convergence theorem, enabling us to employ the Markov chain approximation method to construct consistent, easily-parallelizable, recursive algorithms to calculate the fundamental statistical characteristics and to implement Bayesian inference in real-time for streaming UHF data. The general theory is illustrated by a specific model 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.

Keywords: filtering, marked point process, ultra-high frequency data, market microstructure noise, Bayesian analysis, Markov chain approximation method.

First Draft: February 8, 2010
This Version: July 6, 2010

1. INTRODUCTION

Transactions data in financial markets have become widely available in recent years. Such time-stamped (“tick”) data, containing the most detailed information for the price evolution and quote revisions, are referred to as *ultra-high frequency* (UHF) data by Engle (2000). The direct modeling and analysis of these data, which make full use of the information at one’s disposal, is essential for insight concerning market microstructure theory.

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 clus-
tering, 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 the trades. However, there are two different views on how to treat the observations, leading to different formulations of the statistical foundation for analyzing UHF data.

The first view is well expressed by Engle (2000) 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 (O’Hara (1997)), this school of econometricians models the duration between trades as a stochastic phenomena resulting in an irregularly-spaced time series. Engle & Russell (1998) propose an Autoregressive Conditional Duration (ACD) model 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 Pacurar (2008) and Engle & Russell (2009) 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 first advocated in Zeng (2003), 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 as a filtering problem with counting process observations and the link to stochastic nonlinear filtering is established. Zeng (2003) and Kouritzin & Zeng (2005a) develop the Bayesian inference via filtering for this model to accommodate the computational burden imposed by the MPP 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, we call the first representation 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.

This representation inherits the simple economic notion predicated in the filtering model of Zeng (2003): The price is formed from an intrinsic value process by incorporating the noise created by the trading process (See for example, Black (1986) and Stoll (2000)). There is a literature of time series Vector AutoRegressive (VAR) structural models in market microstructure theory, surveyed in Hasbrouck (2007), that uses the same framework. Sharing this structure, our model can separately identify the influence of information and noise on price: Information affects the intrinsic value of an asset, with a permanent influence on the price, while noise has a transitory price impact alone1.

1We develop a general parametric framework with latent process for UHF data. However, in the literature, models with similar structure for the log-price have been used to study realized volatility (RV) estimators, including two-scale realized volatility (TSRV) estimators and their variants. Examples in this exploding literature include Andersen et al. (2003), Barndorff-Nielsen & Shephard (2004), Andersen et al. (2005), Goncalves & Meddahi (2009), and Mykland & Zhang (2009) for cases without microstructure noise. With noise added, several important papers include A¨ıt-Sahalia et al. (2005), Zhang et al. (2005), Zhang (2006), Li & Mykland (2007), and Barndorff-Nielsen et al. (2008) among many others. Additional papers by Oomen (2005), Bandi & Russell (2006), Hansen & Lunde (2006), Fan & Wang (2007), Jacod et al. (2009), Andersen et al. (2010), and Ghysels & Sinko (2010) highlight recent developments in this area.
ECONOMETRIC ANALYSIS VIA FILTERING

The latent state process in our representation is assumed to be a multivariate Markov process characterized by its infinitesimal generator. Thus, our model connects to the literature on the operator approach for continuous-time Markov processes surveyed in A¨ıt-Sahalia et al. (2009). This is especially so for estimating a Markov process sampled at random-time intervals, a setup evolved from the discrete time sampling proposed in Hansen & Scheinkman (1995) via the Generalized Method of Moments (GMM). Recent papers in this area include Conley et al. (1997), Hansen et al. (1998), A¨ıt-Sahalia & Mykland (2003), A¨ıt-Sahalia & Mykland (2004), Duffie & Glynn (2004), A¨ıt-Sahalia & Mykland (2008) and Chen et al. (2010). Instead of directly observing the state process at event times, our observations are corrupted by noise. Here, we systematically develop Bayesian inference via filtering without assuming stationarity of the Markov process, a key assumption in the papers noted above.

Our proposed framework generalizes the model of Engle (2000) by adding a latent continuous-time Markov process, permitting the influence of other observable economic or market factors and with a clear economic interpretation. Compared to Zeng (2003), 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, including but not limited to conditional Poisson processes 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.

Just as a state-space model has a linear Kalman filter formulation, our model also has a second representation as a nonlinear filtering framework with marked point process observations. Based on this formulation, we define the corresponding fundamental statistical characteristics for estimation, hypothesis testing and model selection: the joint likelihood, the marginal (or integrated) likelihood, the posterior, the likelihood ratio or the Bayes factors, and the posterior model probabilities of the class of the proposed models. Each is usually characterized by a conditional measure that changes over time: Namely, by measure-valued processes, which are of continuous time and infinite dimension due to the latent continuous-time state vector process. Hence, the computational challenge necessary for statistical inference is not trivial.

To address this issue, we use stochastic nonlinear filtering and, under very mild assumptions, we derive the necessary filtering equations, which are stochastic partial differential equations (SPDEs), to stipulate the evolution of the continuous-time statistical characteristics. Like the Kalman filter, these filters preserve recursiveness, which provides a computational advantage if properly utilized. Further, besides the desirable properties of Bayesian analysis (Zellner (1985), Efron (1986) and Berger (1986)), it is natural to choose a Bayesian approach and to develop Bayesian inference via filtering for the class of our proposed models to address the computational overhead. Mainly, we prove a weak convergence theorem, enabling us to employ the Markov chain approximation method to construct consistent (or robust), easily-parallelizable, efficient and recursive algorithms for real-time updates of the approximate fundamental statistical characteristics and real-time model identification and selection for streaming UHF data.

Finally, to illustrate the application of the general model for providing inference tests in empirical studies, 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. After calibrating the algorithms for correct model identification and selection using simulated data, we use the model to provide Bayesian inference via filtering to discriminate between competing theories in the finance literature regarding the evolution of intraday volatility in the U.S. Treasury note market. We 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 two characterizations of marked point processes which we use throughout the paper. Section 3 reviews the two different approaches in the literature for modeling UHF data, along with recent developments, and Section 4 introduces the two equivalent representations of our proposed model with examples. Section 5 defines the continuous-time statistical foundations and presents the related filtering equations. Section 6 develops the Bayesian inference via filtering for the class of proposed models. Section 7 provides an explicit example for our on-the-run U.S. Treasury 10-year note transactions price data. We provide some concluding remarks in Section 8, with the mathematical proofs for the paper collected into Appendices².

2. TWO CHARACTERIZATIONS OF MARKED POINT PROCESSES

On a complete probability space \((\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P)\) with the usual hypotheses, we define a marked point process \(\{(T_i, Y_i)\}_{i \geq 1}\). The event occurrence times \(\{T_i\}\) form a point process with \(T_1 < T_2 < \cdots < T_i < \cdots < T_\infty\). 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 \(\mathcal{Y}\), which is called a mark space. The pair \((T_i, Y_i)\) is a mark point and the sequence \(\Phi = \{(T_i, Y_i)\}_{i \geq 1}\) is called a marked point process. Throughout this paper, we restrict that \(\{T_i\}\) are positive continuous random variables and are \(P\)-nonexplosive, namely, \(P(T_\infty = +\infty) = 1\); and that the mark space \(\mathcal{Y}\) and and its corresponding \(\sigma\)-field, \((\mathcal{Y}, \mathcal{B})\) forms a measurable 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. Last & Brandt (1995) and Daley & Vere-Jones (2003, 2008) provide book-length studies on the topics, while Engle (2000) uses the stochastic intensity approach extensively. In this paper, we use both approaches.

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

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

with \(\Phi(\{0\} \times \mathcal{Y}) = 0\) where \(\delta_{[T_i, Y_i]}(t, y)\) is the Dirac delta-function on \(\mathbb{R}^+ \times \mathcal{Y}\) which takes the value one only when \(t = T_i\) and \(y = Y_i\) and takes the value zero otherwise.

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 \(\Phi(t) := \Phi([0, t] \times \mathcal{Y}) = \Phi(t, \mathcal{Y})\), which is the counting process of event occurrence.

DEFINITION 2.1 An \(\mathcal{F}_t\)-predictable³ random measure \(\gamma\) mapping from \(\Omega \times B(\mathbb{R}^+) \times \mathcal{Y}\) to \([0, +\infty]\) is referred to as a compensator of \(\Phi\) if \(\Phi(t, B) - \gamma(t, B)\) is an \(\mathcal{F}_t\)-martingale for all \(t \geq 0\) and \(B \in \mathcal{Y}\).

The martingale property implies

\[
E[\Phi((s, t] \times B)|\mathcal{F}_s] = E[\gamma((s, t] \times B)|\mathcal{F}_s].
\]

²Throughout the paper, we assume all stochastic processes are defined on a probability space \((\Omega, \mathcal{F}, P)\) with right continuous paths with left limits (càdlàg).

³Essentially, the same as \(\mathcal{F}_{t-}\)-measurable.
Letting \( s \downarrow t \) and \( B \to dy \), we write the above equation in a heuristic form:

\[
\gamma(d(t, y)) = P(\Phi(d(t, y)) > 0|\mathcal{F}_t).
\]

**Definition 2.2.** A stochastic intensity kernel \( \lambda \) of \( \Phi \) (with respect to a measure \( \eta \)) is a finite \( \{F_t\} \)-predictable kernel from \( \Omega \times \mathbb{R}^+ \) to \( \mathbb{Y} \) such that for all predictable \( f : \Omega \times \mathbb{R}^+ \times \mathbb{Y} \to \mathbb{R} \),

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

The martingale property of a compensator and the above definition implies

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

The second equality shows the connection between the compensator and stochastic intensity kernel:

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

When \( \eta \) is a Lebesgue measure \( \eta(dt) = dt \), \( \lambda(t, B) \) represents the potential for another point at time \( t \) with mark in \( B \) given \( \mathcal{F}_t \):

\[
(2.2) \quad \lambda(t, B) = \lim_{h \to 0} h^{-1} P(\Phi([t, t+h] \times B) > 0|\mathcal{F}_t).
\]

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

\[
\bar{\lambda}(t) = \lambda(t, \mathbb{Y}) = \lim_{h \to 0} h^{-1} P(\Phi([t, t+h]) > 0|\mathcal{F}_t),
\]

and the compensator of \( \Phi(t) \) is \( \bar{\lambda}(t)dt \).

The stochastic intensity kernel is closely related to the conditional hazard measures of arrival time, duration and marked point, respectively. The conditional hazard measures help to explicitly express the stochastic intensity kernel and the compensator of \( \Phi \).

**Definition 2.3.** The conditional hazard measure of time point \( T_{i+1} \) given \( \mathcal{F}_{T_i} \), denoted by \( R^{T_i}(dt) \), is defined as, for \( t \in [T_i, T_{i+1}) \),

\[
R^{T_i}(dt) = \frac{P(T_{i+1} \in dt|\mathcal{F}_{T_i})}{P(T_{i+1} \geq t|\mathcal{F}_{T_i})}.
\]

If \( R^{T_i}(dt) = r^{T_i}(t)dt \), then \( r^{T_i}(t) \) is called the conditional hazard rate function of \( T_{i+1} \) given \( \mathcal{F}_{T_i} \).

Observe that

\[
R^{T_i}(dt) = \frac{P((T_{i+1} - T_i) \in (dt - T_i)|\mathcal{F}_{T_i})}{P((T_{i+1} - T_i) \geq t|\mathcal{F}_{T_i})}.
\]

Let \( R^{T_i}_d(dt - T_i) \) denote the conditional hazard measure of the duration, \( \Delta T_{i+1} = T_{i+1} - T_i \), given \( \mathcal{F}_{T_i} \). Then, \( R^{T_i}_d(dt) = R^{T_i}_d(dt - T_i) \). If \( R^{T_i}_d(dt - T_i) = r^{T_i}_d(t - T_i)dt \), then \( r^{T_i}_d(t - T_i) \) is called the conditional hazard rate function of the duration \( \Delta T_{i+1} \) given \( \mathcal{F}_{T_i} \).
Remark 2.1 Note that \( r^T_d(t - T_i) \) provides a connection to ACD models and Engle’s framework for UHF data, to be reviewed in the next Section.

Definition 2.4 The conditional hazard measure of marked point \((T_{i+1}, Y_{i+1})\) given \( \mathcal{F}_{T_i} \), denoted by \( R^T_{T_i}(d(t,y)) \), is defined as, for \( t \in [T_i, T_{i+1}) \),

\[
R^T_{T_i}(d(t,y)) = \frac{P((T_{i+1}, Y_{i+1}) \in d(t,y) | \mathcal{F}_{T_i})}{P(T_{i+1} \geq t | \mathcal{F}_{T_i})}.
\]

Observe that \( R^T_{T_i}(dt) = R^{T_i}(dt, \mathcal{Y}) = \int_{\mathcal{Y}} R^{T_i}(d(t,y)) \) and that there exists a \( \mathcal{F}_t \)-predictable \( \kappa_i(t, dy) \), which is the conditional density of \( y \), such that \( R^{T_i}(d(t,y)) = \kappa_i(t, dy) R^T_{T_i}(dt) = \kappa_i(t, dy) R^T_{T_i}(dt - T_i) \).

Theorem 4.1.11 of Last & Brandt (1995) provides the expression for the compensator:

\[
\gamma(d(t,y)) = \sum_{i \geq 1} I\{T_i < t \leq T_{i+1}\} R^{T_i}(d(t,y))
\]

\[
= \sum_{i \geq 1} I\{T_i < t \leq T_{i+1}\} \kappa_i(t, dy) R^T_{T_i}(dt) = \sum_{i \geq 1} I\{T_i < t \leq T_{i+1}\} \kappa_i(t, dy) R^T_{T_i}(dt - T_i).
\]

When \( R^T_{T_i}(dt) = r^{T_i}(t) dt \) and \( R^T_{T_i}(dt - T_i) = r^T_d(t - T_i) dt \), the stochastic intensity kernel and the stochastic intensity are expressed as:

\[
\lambda(t, dy) = \sum_{i \geq 1} I\{T_i < t \leq T_{i+1}\} \kappa_i(t, dy) r^{T_i}(t) = \sum_{i \geq 1} I\{T_i < t \leq T_{i+1}\} \kappa_i(t, dy) r^T_d(t - T_i)(t - T_i)
\]

\[
\lambda(t) = \sum_{i \geq 1} I\{T_i < t \leq T_{i+1}\} r^{T_i}(t) = \sum_{i \geq 1} I\{T_i < t \leq T_{i+1}\} r^T_d(t - T_i)(t - T_i).
\]

3. UHF DATA AND RELATED DEVELOPMENTS

3.1. Irregularly-spaced Time Series

Engle (2000) provides a representative view for irregular-spaced time series in UHF data. In Engle’s framework, \( \mathcal{F}_t = \mathcal{F}_t^\Phi \), which is the natural filtration generated by \( \Phi \). Recall that \( \Delta T_i = t_i - t_{i-1} \) is the observed duration between events. Suppose that the observed mark \( y_i \) is a \( k \times 1 \) vector from a mark space \( \mathcal{Y} \). Then the UHF data is viewed as a “pseudo” time series, which is irregularly spaced in time:

\[
\{(\Delta t_i, y_i), i = 1, ..., n\}
\]

where the \( i \)th observation has the joint density conditional on the past filtration of \((\Delta t, y)\) given by \((\Delta t_i, y_i) | \mathcal{F}_{i-1} \sim f(\Delta t_i, y_i | \Delta t_{i-1}, \tilde{y}_{i-1}; \theta)\), and \( \tilde{y}_i = \{z_i, z_{i-1}, ..., z_n\} \) denotes the past of \( z \) and \( \theta \) as a vector of parameters, whose components are allowed to differ across observations, similar to ARCH-type models.

Without loss of generality, the joint density can be decomposed into the product of the marginal density of the duration multiplied by the conditional density of the marks given the duration, all conditioned upon the past transactions:
The above decomposition has a one-to-one correspondence with the decomposition of the stochastic intensity kernel in (2.3):

\[ \kappa_i(t, dy) = p(y|\Delta t_i, \tilde{\Delta}_{t_i-1}, \tilde{y}_{i-1}; \theta)dy, \]

and

\[ r_d^{T_i-1}(t - T_{i-1}) = \frac{g(t - T_{i-1}|\tilde{\Delta}_{t_i-1}, \tilde{y}_{i-1}; \theta)}{\int_s^t g(s - T_{i-1}|\tilde{\Delta}_{t_i-1}, \tilde{y}_{i-1}; \theta)ds}, \]

which is the conditional hazard rate function with respect to the conditional \( g(x) \), where \( x = t - T_{i-1} \). Conversely, the conditional \( g(x) \) can be obtained from \( r_d^{T_i-1}(x) \):

\[ g(x) = r_d^{T_i-1}(x) \exp \left(-\int_0^x r_d^{T_i-1}(s)ds\right). \]

With the conditional density in (3.2), the log likelihood is simply the sum of the logs of the \( n \) individual joint densities conditional on the past, and can be written as:

\[ l(\Delta, Y; \theta) = \sum_{i=1}^K \log g(\Delta t_i|\tilde{\Delta}_{t_i-1}, \tilde{y}_{i-1}; \theta) + \sum_{i=1}^K \log p(y_i|\Delta t_i, \tilde{\Delta}_{t_i-1}, \tilde{y}_{i-1}; \theta), \]

where \( \Delta \) and \( Y \) are the data. Based on this log likelihood, the associated parametric inference and testing procedures can be developed.

### 3.1.1. ACD Models

An important special case concerns modeling the duration alone, with the simplest mark space \( \mathbb{Y} = \{1\} \). The log likelihood then simplifies to the first sum in Eq.(3.4). Engle & Russell (1998) propose an Autoregressive Conditional Duration (ACD) model, a specification of the conditional density of the durations that needs only a mean function. Let \( \psi \) be the conditional expected duration given by \( \psi_i = \psi(\Delta t_{i-1}, \tilde{y}_{i-1}; \theta) = E_{i-1}(\Delta t_i|\tilde{\Delta}_{t_i-1}, \tilde{y}_{i-1}; \theta) = \int s g(s|\tilde{\Delta}_{t_i-1}, \tilde{y}_{i-1}; \theta)ds \) and suppose that \( \Delta t_i = \psi_i \varepsilon_i \), where \( \varepsilon \sim \text{i.i.d.} \) with mean one, density \( p_0(x) \) and hazard rate \( \lambda_0(x) = p_0(x)/\int_x^\infty p_0(u)du \).

With the iid assumption of \( \varepsilon \), \( g(\Delta t_i|\tilde{\Delta}_{t_i-1}, \tilde{y}_{i-1}; \theta) = g(\Delta t_i|\psi_i; \theta) \), where \( g \) is potentially a known function of \( \psi \). Then an ACD(\( m,q \)) model specifies a GARCH-like structure that depends on the most recent \( m \) durations as well as the most recent \( q \) conditional expectations of the durations:

\[ \psi_i = \omega + \sum_{j=1}^m \alpha_j \Delta t_{i-j} + \sum_{j=0}^q \beta_j \psi_{i-j}. \]

Table I provides the baseline hazards and stochastic intensities when \( \varepsilon_i \) has an exponential or Weibull distribution, the two most common specifications. Other possibilities include the Burr distribution advocated in Grammig & Maures (2000); the generalized gamma distribution employed in Zhang et al. (2001); and a mixture of distributions as considered in Luca & Gallo (2004). Engle (2000) develops semiparametric hazard estimation via quasi-maximum likelihood estimation (QMLE). Drost & Werker (2004) allow the innovations \( \varepsilon_i \) to have dependencies, and illustrate sizable efficiency gains for several semiparametric forms. Testing the distribution of \( \varepsilon_i \) is considered in Bauwens et al. (2004) and Fernandes & Grammig (2005).

The analogy between ACD and GARCH models allows a generalization of ACD models in various directions. Bauwens & Giot (2000) develop a logarithmic ACD model that avoids positivity
TABLE I
Examples of ACD Models

<table>
<thead>
<tr>
<th>ACD Models</th>
<th>Baseline hazard</th>
<th>Stochastic Intensity: ( \lambda(t) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>General ACD</td>
<td>( \lambda_0(x) )</td>
<td>( \sum_{i \geq 1} I(T_{i-1} &lt; t \leq T_i) \lambda_0 \left( \frac{t-T_{i-1}}{\psi_i} \right) \frac{1}{\psi_i} )</td>
</tr>
<tr>
<td>Exponential ACD</td>
<td>1</td>
<td>( \sum_{i \geq 1} I(T_{i-1} &lt; t \leq T_i) \psi_i^{-1} )</td>
</tr>
<tr>
<td>Weibull(( b )) ACD</td>
<td>( b(1 + 1/b)^b \psi_i^{b-1} )</td>
<td>( \sum_{i \geq 1} I(T_{i-1} &lt; t \leq T_i) b \left( \frac{t-T_{i-1}}{\psi_i} \right)^b )</td>
</tr>
</tbody>
</table>

restrictions on the parameters, and Bauwens & Giot (2003) develop an asymmetric ACD model that differentiates between price increases and decreases as a result of news releases. Fernandes & Grammig (2006) introduce augmented ACD models that subsume most previous ACD models, though they do not include the threshold (regime-switching) ACD model proposed in Zhang et al. (2001).

3.1.2. Models for Durations and Marks


The likelihood in (3.4) can easily be generalized to two or more distinct time sequences. For example, Engle & Lunde (2003) employ a bivariate point process to jointly model the trade and quote times. Russell (1999) develops an Autoregressive Conditional Intensity model and applies it to the dynamics of trade times and limit order arrival times. The likelihood function for each models reviewed above is a special case of the likelihood in (3.4), or its extension to higher dimensions. Engle (2000) further describes the conditional joint density for a general model with unobserved latent processes taking values only at the event occurrence times. The marginal density of \( (\Delta t_i, y_i) \) requires multiple integration with respect to the latent processes. Research in this area includes the stochastic duration models in Bauwens & Veredas (2004), Ghysels et al. (2004), Feng et al. (2004), Bauwens & Hautsch (2006), Xu et al. (2009) and Renault et al. (2010). Because of the numerical hurdles associated with the need for multiple integration of continuous-time latent processes in this paper, we use a filtering technique for our proposed model to circumvent these issues.

3.2. A Realized Sample Path of MPP

Due to price discreteness, Zeng (2003) treats the observed transaction prices of an asset, \( \{t_i, y_i\}_{i=1}^n \), as a realized sample path of a collection of counting processes, a special case of a MPP observation. The observed filtration is still \( \mathcal{F}_t^\Phi \), but the filtration of the probability space is extended to
\( F_t = F_{t}^{\theta,X,\Phi} \), where \( \theta(t) \) is a vector of parameters and \( X(t) \) is the latent intrinsic value process of the asset. A nonlinear filtering model is formed, and the mark space is \( Y = \{ \frac{a}{M}, \frac{a+1}{M}, \ldots, \frac{a+m}{M} \} \), where \( \frac{1}{M} \) is a tick, the minimum price variation set by trading regulation. There are three representations of the filtering model. They are equivalent in the sense that they have the same probability distribution.

### 3.2.1. Representation I: Constructing the Transaction Price from the Intrinsic Value

Representation I is based on the simple idea that the transaction price is formed from an intrinsic value by incorporating the noise arising from the trading activity. In general, there are three steps in constructing the price process \( Y \) from the intrinsic value process \( X \). First, we invoke a mild assumption jointly on \((\theta, X)\).

**Assumption 3.1** \((\theta, X)\) is a \( p + 1 \) dimension vector Markov process which is the unique solution of a martingale problem for a generator \( A \) such that for a function \( f \) in the domain of \( A \), \( M_f(t) = f(\theta(t), X(t)) - \int_{0}^{t} A f(\theta(s), X(s)) ds \), is a \( F_{t}^{\theta,X} \)-martingale, where \( F_{t}^{\theta,X} \) is the \( \sigma \)-algebra generated by \((\theta(s), X(s))\) for \( 0 \leq s \leq t \).

With UHF data, prices cannot be observed in continuous-time, neither can they move in space as a continuous-path stochastic process such as geometric Brownian motion (GBM) suggests. Therefore, two more steps are necessary. Steps 2 and 3 specify trading times and transaction prices, respectively.

In Step 2, we assume trading times \( T_1, T_2, \ldots, T_i, \ldots \), are observations of a conditional Poisson process with a stochastic intensity \( a(X(t), \theta(t), t) \).

In Step 3, \( y_i = Y(T_i) \), the price at time \( T_i \), is corrupted observation of \( X(T_i) \), the intrinsic value. Namely, \( Y(T_i) = F(X(T_i)) \), where \( y_i = F(x) \) is a random transformation with the transition probability \( p(y_i|x) = p(y_i|x; \theta(T_i), T_i) \) where \( x = X(T_i) \). The random transformation that models the trading noise is flexible and can accommodate different stylized noise constructs in the UHF data, including discrete and clustering noise.

### 3.2.2. Representation II: Filtering with Counting Process Observations

Because of price discreteness, the price of an asset can be regarded as a collection of counting processes in the following form:

\[
\tilde{Y}(t) = \begin{pmatrix}
N_1(\int_{0}^{t} \lambda_1(\theta(s), X(s), s) ds) \\
N_2(\int_{0}^{t} \lambda_2(\theta(s), X(s), s) ds) \\
\vdots \\
N_m(\int_{0}^{t} \lambda_m(\theta(s), X(s), s) ds)
\end{pmatrix},
\]

where \( Y_j(t) = N_j(\int_{0}^{t} \lambda_j(\theta(s), X(s), s) ds) \) is the counting process recording the cumulative number of trades that have occurred at the \( j \)th price level (denoted by \( y_j \)) up to time \( t \). Note that \( \Phi = \tilde{Y} \) in this setup.

We include four assumptions for the model so that the three representations are equivalent.

**Assumption 3.2** \( \{N_j\}_{j=1}^{m} \) are unit Poisson processes under measure \( \mathbb{P} \).
Then, \( Y_j(t) = N_j(\int_0^t \lambda_j(\theta(s), X(s), s)ds) \) is a conditional Poisson process with the stochastic intensity, \( \lambda_j(\theta(t), X(t), t) \). Given \( \mathcal{F}_t^{\theta,X} \), \( Y_j(t) \) has a Poisson distribution with parameter \( \int_0^t \lambda_j(\theta(s), X(s), s)ds \), and \( Y_j(t) - \int_0^t \lambda_j(\theta(s), X(s), s)ds \) is a \( \mathcal{F}_t^{\theta,X,Y} \) martingale.

**Assumption 3.3** \((\theta, X), N_1, N_2, \ldots, N_m \) are independent under measure \( \mathbb{P} \).

**Assumption 3.4** The total trading intensity \( a(\theta(t), X(t), t) \) is bounded by a positive constant, \( C \). Namely, \( 0 \leq a(\theta, x, t) \leq C \) for all \( t > 0 \) and \((\theta, x)\).

These three assumptions imply that there exists a reference measure \( Q \) and that, after a suitable change of measure to \( Q \), \((\theta, X), Y_1, \ldots, Y_m \) become independent and \( Y_1, Y_2, \ldots, Y_m \) become unit Poisson processes (Bremaud (1981)).

**Assumption 3.5** The stochastic intensities are of the form: \( \lambda_j(\theta, x, t) = a(\theta, x, t)p(y_j|x) \), where \( p(y_j|x) \) is the transition probability from \( x \) to \( y_j \), the \( j \)th price level.

This desirable structure of the intensities is similar to that in (2.3) and ensures the equivalence of the representations. Likewise, \( p(y_j|x) \) models how the trading noise enters the price process.

Under this representation, \((\theta(t), X(t))\) becomes the signal, which cannot be observed directly, but can be partially observed through the counting processes \( \tilde{Y}(t) \), the noisy observations. Hence, \((\theta, X, \tilde{Y})\) is framed as a filtering model with counting process observations.

### 3.2.3. Representation III: An Integral Form of Price

Let \( Y(t) \) be the price of the most recent transaction at or before time \( t \). Then,

\[
Y(t) = Y(0) + \int_{[0,t] \times \mathcal{Y}} (y - Y(s-))\Phi(ds, dy),
\]

where \( \Phi \) is the random counting measure defined in (2.1). Note that \( \Phi(ds, dy) \) is zero most of time, and becomes one only at trading time \( t_i \) with \( y = Y(t_i) \), the trading price. The above expression is but a telescoping sum: \( Y(t) = \tilde{Y}(0) + \sum_{t_i < t} (Y(t_i) - Y(t_{i-1})) \).

This integral form supplies the price evolution, has a Doob-Meyer decomposition, and provides a means for the study of mathematical finance problems related to our model such as hedging and option pricing via the minimum martingale measure (Lee & Zeng (2010)) and the optimal mean-variance portfolio selection problem (Xiong & Zeng (2010)).

### 3.2.4. Related Developments

The Bayes estimation for our model is developed in Zeng (2003) and the Bayesian model selection via filtering is examined in Kouritzin & Zeng (2005a). Applying the theory of Bayes estimation, Zeng (2003) studies a simple model where the intrinsic value process is geometric Brownian motion (GBM) and the noise includes three well-documented types: discrete, clustering and non-clustering. Zeng (2004) examines the case where the value process is a jumping-stochastic-volatility GBM model. Kouritzin & Zeng (2005a) study the discrimination between these two models via Bayes factors.
4. TWO REPRESENTATIONS OF THE MODEL

In this section, we further assume that \((Y, d_Y)\) is a complete, separable metric space. The observed filtration, denoting all available information up to time \(t\), is extended to \(\mathcal{F}^{\Phi,V}_t\), which is defined as

\[(4.1) \quad \mathcal{F}^{\Phi,V}_t = \sigma\{ (\Phi(s, B), V(s)) : 0 \leq s \leq t, B \in \mathcal{Y} \},\]

where \(V\) is a vector of observable economic or market factors. As in Zeng (2003), we incorporate the latency \((\theta, X)\) and assume that \((\theta, X, \Phi)\) is in a probability space \((\Omega, \mathcal{F}, P)\) equipped with a filtration \((\mathcal{F})_{t \geq 0}\) that satisfies the usual conditions (Protter (2003))). We let \(\mathcal{F} = \mathcal{F}_\infty\) where

\[(4.2) \quad \mathcal{F}_t = \sigma\{ (\theta(s), X(s), \Phi(s, B), V(s)) : 0 \leq s \leq t, B \in \mathcal{Y} \}.

Also, we define \(\mathcal{F}_{t-} = \sigma\{ (\theta(s), X(s), \Phi(s, B), V(s)) : 0 \leq s < t, B \in \mathcal{Y} \}\), because \(\mathcal{F}_t\)-predictable is essentially \(\mathcal{F}_{t-}\)-measurable.

With the above background, the first two subsections below present our model framework in two equivalent representations. The first is more intuitive and the second is a direct generalization of counting process observations. After stating the two representations are equivalent, we demonstrate the generality of the proposed model by examples.

4.1. Representation I: Random Arrival Time State-Space Model

As a state-space time series model, this representation has similar components: a state process, observation times, the observations themselves and noise.

4.1.1. State Process

Applying a modern Bayesian idea to prepare for parameter estimation, we extend the state space to include parameter space. The extended state process \((\theta, X)\) allows the components of \(\theta\) to potentially change in continuous time, and we make a very mild assumption on \((\theta, X)\).

**Assumption 4.1** \((\theta, X)\) is a \(p+m\)-dimension vector Markov process that is the unique solution of a martingale problem for a generator \(A_v\), such that \(M_f(t) = f(\theta(t), X(t)) - \int_0^t A_v f(\theta(s), X(s)) ds\) is a \(\mathcal{F}^{\theta,X,V}_t\)-martingale, where \(\mathcal{F}^{\theta,X,V}_t\) is the \(\sigma\)-algebra generated by \((\theta(s), X(s), V(s))_{0 \leq s \leq t}\), and \(f\) is in the domain of \(A_v\).

The generator and martingale problem approach (see, for example, Ethier & Kurtz (1986)) furnishes a powerful tool for the characterization of Markov processes. Obviously, Assumption 4.1 is a multivariate version of Assumption 3.1, 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 \(\theta\), our model further allows stochastic volatility and regime switching. Moreover, we do not assume stationarity of \((\theta, X)\), which makes our assumption less restricted than in the literature for estimating Markov processes sampled at discrete or random times (Hansen & Scheinkman (1995), Ait-Sahalia (1996) and Duffie & Glynn (2004) among others). Furthermore, other observable factors, represented by a vector process \(V\), are incorporated in the state process of \(X\) through the generator, \(A_v\).
4.1.2. 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:

$$\tilde{\lambda}(t) = \tilde{\lambda}(\theta(t), X(t), V(t), \Phi(t), t),$$

where $V^t = V(\cdot \land t)$ denotes the sample path of $V$ up to time $t$ and similarly $\Phi^t = \{(T_i, Y_i) : T_i < t\}$. Namely, we allow $\tilde{\lambda}(t)$ to be Markovian in $(\theta, X, \gamma, t)$ and non-anticipating in $V$ and $\Phi$ (depending on the sample paths of $V$ and $\Phi$ up to time $t$). This form is very general and Table II provides some important examples.

Clearly, a conditional Poisson process is one such example. When $\tilde{\lambda}(t)$ depends on $(\theta(t), X(t))$, this is the endogenous sampling studied in Duffie & Glynn (2004) and Chen et al. (2010) and empirically tested for stochastic volatility in Renault & Werker (2010). When the stochastic intensity is only $\mathcal{F}_t^V, \Phi$-predictable, such as the case of exogenous sampling, our specification accommodates three popular conditional intensity parameterizations: Cox models and their more generalized proportional hazard formulations, the Hawkes process (Large (2007) and Aït-Sahalia, Cacho-Díaz & Laeven (2010)), and ACD models. Theorem 2.2.22 of Last & Brandt (1995) is the key result ensuring the stochastic intensity functions of the proportional hazard and ACD models are $\mathcal{F}_t^V$-predictable. 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.

4.1.3. Observations

The noisy observation at event time $T_i$, $Y(T_i)$, takes a value in the mark space $\mathcal{Y}$ 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 $p(y|x)$, which is $\mathcal{F}_t$-predictable. More specifically, we allow $p(Y(T_i)|X(T_i))$ to be Markovianly dependent on $(\theta(T_i), X(T_i), T_i)$ and non-anticipatingly depending on $(V^{T_i}, \Phi^{T_i})$. In many papers on realized volatility estimators in the presence of noise such as Aït-Sahalia et al. (2005), $Y(T_i) - X(T_i)$ has a normal distribution with mean zero and constant variance. The noise in Li & Mykland (2007) is additive normal with rounding. The noise in Zeng (2003) includes additive doubly-geometric, rounding and clustering noises. These are all examples of $F(x)$ with $p(y|x)$. We do not include the case of dependent noise as studied in Aït-Sahalia, Mykland & Zhang (2010).

In summary, the observation $(T_i, Y(T_i))$ forms a MPP with the predictable stochastic intensity kernel $\lambda(\theta(t), X(t), V^t, \Phi^t, t)$ $p(Y(t)|X(t); \theta(t), V^t, \Phi^t, t)$. This representation is similar to state-space models a literature rich with sequential Monte Carlo methods (Doucet et al. (2001)), and

<table>
<thead>
<tr>
<th>Models</th>
<th>Conditional Hazard Rate $r^{T_i-1} = r^{T_i-1}(t - T_{i-1})$</th>
<th>Stochastic Intensity $\tilde{\lambda}(t) = \sum_{i \geq 1} I_{{T_i-1 &lt; t \leq T_i}}(t)r^{T_i-1}(t - T_{i-1})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conditional Poisson $\lambda(\theta(t), X(t), t) = \tilde{\lambda}(\theta(t), X(t), t)$</td>
<td>$\tilde{\lambda}(\theta(t), X(t), t)$</td>
<td></td>
</tr>
<tr>
<td>Hawkes</td>
<td>$\alpha(t; \theta) + \int_0^t \beta(t - s; \theta) d\Phi(s)$</td>
<td>$\alpha(t; \theta) + \int_0^t \beta(t - s; \theta) d\Phi(s)$</td>
</tr>
<tr>
<td>Proportional hazard $\lambda(t - T_{i-1}) = \exp{-\beta'(V(T_{i-1}))}$</td>
<td>$\sum_{i \geq 1} I_{{T_{i-1} &lt; t \leq T_i}}(t)\lambda(t - T_{i-1}) = \exp{-\beta'(V(T_{i-1}))}$</td>
<td></td>
</tr>
<tr>
<td>General ACD</td>
<td>$\lambda_0 \left( \frac{t - T_{i-1}}{\psi_{i+1}} \right)^{1/\psi_i}$</td>
<td>$\sum_{i \geq 1} I_{{T_{i-1} &lt; t \leq T_i}}(t)\lambda_0 \left( \frac{t - T_{i-1}}{\psi_{i+1}} \right)^{1/\psi_i}$</td>
</tr>
</tbody>
</table>
can also be viewed as a random arrival time state-space model. There are however two significant differences from the state-space models of time series: the random arrival times and the continuous-time feature of the state process. Our representation is also related to hidden Markov models, a large literature surveyed in Elliott et al. (1995), Ephraim & Merhav (2002) and Cappé et al. (2005), and can be viewed as 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 that reviewed in Section 3.2.1, and shares the simple economic notion that the observed price at trading time $T_i$, $Y(T_i)$, is based upon the intrinsic value, $X(T_i)$, but distorted by the trading noise modeled by $F$, the random transformation. Further, and similar to VAR-structure models, our representation can separately identify the impact of information and noise: Information affects $X(t)$, the intrinsic value of an asset, with a permanent influence on the price while noise affects $F(x)$ (or $p(y|x)$) with a transitory impact on price. The finance literature contains many examples of VAR-structure models that rely on this interpretation of price impact, including Roll (1984), Glosten & Harris (1988), Harris (1990), Hasbrouck (1999), George & Hwang (2001) and Hasbrouck (2002), among others. Hasbrouck (1996) and Hasbrouck (2007) provide extensive surveys of this literature.

\section*{4.2. Representation II: Filtering with MPP Observations}

Under this representation, $(\theta, X)$ 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 $(\Phi, V)$, 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. Hence, $(\theta, X, \Phi, V)$ is framed as a filtering model with MPP observations.

Let $(\mathcal{Y}, \mathcal{Y}, \mu)$ be a measure space where $\mathcal{Y}$ is the mark space of observations with $\sigma$-field $\mathcal{Y}$ and $\mu$ is a finite measure ($\mu(\mathcal{Y}) < \infty$). Just as a MPP can be viewed by a collection of counting processes when the mark space is discrete (the case reviewed in Section 3.2.2), a MPP can generally be described by a random counting measure, $\Phi(t, B)$, recording the cumulative number of marks in a set $B \in \mathcal{Y}$ up to time $t$. Recall that $\Phi(t, B) = \int_B \Phi(dt, dy)$ where $\Phi(dt, dy)$ is given by (2.1). Let the stochastic intensity kernel $\lambda(t, dy)$ be in the form $\lambda(t, dy) = \lambda(t, dy; \theta(t), X(t), V^{t-}, \Phi^{t-}, t)$. The assumption below ensures that this representation is equivalent with the previous one.

\textbf{Assumption 4.2} Under $\mathbb{P}$, the stochastic intensity kernel of $\Phi = \{(T_n, Y_n)\}_{n \geq 1}$ is given by

\begin{equation}
\lambda(t, dy) = \lambda(t, dy; \theta(t), X(t), V^{t-}, Y^{t-}, t-) = \tilde{\lambda}(\theta(t), X(t), V^{t-}, Y^{t-}, t-) p(dy|X(t); \theta(t), V^{t-}, Y^{t-}, t-).
\end{equation}

Then, the compensator under $\mathbb{P}$ is $\gamma_p(d(t, y)) = \lambda(t, dy)dt$. This assumption imposes the same structure of (2.3), which is similar to that of (3.2) in Engle (2000): $\tilde{\lambda}$ is similar to $g$ in that both determine the conditional odds for the next event occurrence, and $p$ in both cases determines the conditional distribution for the next mark occurrence. Here, $\tilde{\lambda}$ is a hazard rate instead of a density, and $p$, depending on $X(t)$, explicitly models how the noise interacts with the observation.

\footnote{This representation is also closely connected to the market microstructure literature on TSRV estimators in the presence of noise.}
In the framework of Engle (2000), the reference measure for the likelihood function is the usual Lebesgue product measure. Due to the inclusion of a continuous-time $X$, the Lebesgue product measure is not adequate. In the next assumption, we assume the existence of a simple Poisson random measure as a reference, with the property under the reference measure that the signal and the observation are independent. This assumption is useful for deriving the filtering equations (Section 5.2) and in constructing the approximate filters (Section 6.1) later in the paper.

**Assumption 4.3** There exists a reference measure $Q$, $\mathbb{P} \ll Q$ so that under $Q$

- $(\theta, X)$ and $V$ are independent of $\Phi = \{(T_n, Y_n)\}_{n \geq 1}$;
- The compensator of MPP $\Phi$ is $\gamma_Q (d(t,y)) = \mu(dy)dt$, or the stochastic intensity kernel of 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. A mild assumption ensures that a reference measure $Q$ exists.

**Assumption 4.4** The $\epsilon(t, y)$ defined in (4.5) satisfies the condition that $E^Q[L(\infty)] = 1$.

Assumption 4.4 implies that $E^Q[L(t)] = 1$ for all $t > 0$. Sufficient conditions for $E^Q[L(T)] = 1$ for some fixed $T$ can be found in Section 3.3 of Boel et al. (1975). Our final assumption is a technical condition that ensures that $\{T_i\}$ is non-explosive and that the filtering equations are well-defined.

**Assumption 4.5** $\int_0^t E^P[\bar{\lambda}(s)]ds < \infty$, for $t > 0$.

### 4.3. Equivalence of the Two Representations

The following proposition states that the two representations are equivalent in distribution. This guarantees that statistical inferences based on the latter representation are consistent with the former.

**Proposition 4.1** The two representations of the proposed model in Sections 4.1 and 4.2 have the same probability law.

The central point is that both marked point processes have the same stochastic intensity kernel.
4.4. Examples

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 presented in Table II. The general framework in Engle (2000) is representative of the second sub-class. As an alternative, Class II contains the model of Zeng (2003).

We outline two additional examples from the literature below. The first is a modified version of the parametric models in the TSRV literature, with a continuous mark space (Aït-Sahalia & Mancini (2008)) included by adding random trade times to the model. A second example is the direct multivariate generalization provided in Zeng (2003).

**Example 4.1 Heston stochastic volatility model with normal noise at random times.** The log intrinsic value process $X$ follows

\[
\begin{align*}
    dX_t &= (\mu - \sigma_t^2/2)dt + \sigma_t dW_{1,t} \\
    d\sigma_t^2 &= \kappa(\alpha - \sigma_t^2)dt + \gamma \sigma_t dW_{2,t},
\end{align*}
\]

where $(W_{1,t}, W_{2,t})$ is a bivariate Brownian motion with correlation coefficient $\rho$. Trading times $\{T_i\}_{i \geq 1}$ are driven by a $\mathcal{F}_t$-predictable point process with stochastic intensity $\lambda(t)$. The observable log price at time $t_i$ is given by

\[
Y(T_i) = X(T_i) + \varepsilon_i
\]

where $\varepsilon_i$ are i.i.d. $N(0, \sigma^2)$.

**Example 4.2 An asynchronously observed multi-asset model.** $(\theta, X)$ is assumed to have Assumption 4.1 for $m$ assets. Trading times are driven by conditional Poisson processes. The total trading intensity of $m$ assets is $\bar{\lambda}(t) = \lambda(\theta(t), X(t), t)$. Set $Y = \{y = (j, u_j) : j \in \{1, 2, \cdots, m\}, u_j \in U_j\}$, where $U_j$ is the price space for Asset $j$, which can be discrete or continuous. Suppose that $Y(T_i) = y = (j, u_j)$. We can decompose

\[
\begin{align*}
    p(Y(T_i)|X(T_i)) &= p(j|X(T_i))p(u_j|X_j(T_i)),
\end{align*}
\]

where $p(j|X(T_i))$ specifies the probability that a trade occurs for Asset $j$ and $p(u_j|X_j(T_i))$ further specifies that the price is at $u_j$ given the transaction, when the intrinsic value of Asset $j$ is $X_j(T_i)$. Note that $p(u_j|X_j(T_i))$ incorporates microstructure noise for Asset $j$. When there is only one asset, (4.8) reduces to $p(Y(T_i)|X(T_i)) = p(y|X(T_i))$, the case reviewed in Section 3.2. This model naturally adapts price asynchronicity for multiple assets, as does the related Bayesian inference via filtering (Scott (2006) and Scott & Zeng (2008)).

Moreover, let $Y_j(t)$ be the price of the most recent transaction at or before time $t$ of Asset $j$. Then,

\[
Y_j(t) = Y_j(0) + \int_0^t \int_{(j \times U_j)} (u - Y_j(s-))\Phi(ds, dy).
\]
This integral price evolution form provides a useful framework for many problems in mathematical finance.

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 (A¨ıt-Sahalia & Mykland (2008), Duffie & Glynn (2004), Hansen et al. (1998) and Conley et al. (1997)). In addition, filtering models that estimate market volatility from UHF data are special cases of our model. Recent papers in this area include Frey (2000), Frey & Runggaldier (2001), Cvitanic, Liptser & Rozovsky (2006), and Cvitanic, Rozovsky & Zaliapin (2006). 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 Synder (1972), Davis et al. (1975), Boel et al. (1975), Chapter 19.3 of Liptser & Shiryaev (2002) (the first edition in 1978), Chapter 6.3 of Bremaud (1981), Kliemann et al. (1990), and Elliott & Malcolm (2005).

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 7, 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.

5. STATISTICAL FOUNDATIONS AND THEIR CHARACTERIZATION

We define the continuous-time statistical foundations and characterize them via SPDEs such as the related filtering equations.

5.1. The Continuous-time Statistical Foundations

We study the continuous-time joint likelihood, the marginal (or integrated) likelihood, the posterior of the proposed model for estimation, as well as the continuous-time likelihood ratio, the Bayes factors, and the posterior model probabilities for hypotheses testing and model selection.

5.1.1. The Continuous-time Joint Likelihood

Let $D_U[0, \infty)$ is the space of right continuous with left limit functions with state space $U$. The probability measure $\mathbb{P}$ on $\Omega = D_{\mathbb{R}^{p+m} \times Y}[0, \infty)$ for $(\theta, X, \Phi)$ can be written as $\mathbb{P} = \mathbb{P}_{\theta, X|v} \times \mathbb{P}_{\Phi|\theta, x, v}$, where $\mathbb{P}_{\theta, x|v}$ is the conditional probability measure on $D_{\mathbb{R}^{p+m}}[0, \infty)$ for $(\theta, X)$ given $V$ such that $M_f(t)$ in Assumption 4.1 is a $\mathcal{F}_t^{\theta, X, V}$ martingale, and $\mathbb{P}_{\Phi|\theta, x, v}$ is the conditional probability measure on $D_Y[0, \infty)$ for $\Phi$ given $(\theta, X, V)$. Under $\mathbb{P}$, $\Phi$ relies on $(\theta, X, V)$. With Assumption 4.3, there exists

---

5We also note that the intrinsic value process $X_t$ can have a variety of functional forms in our model, including the O-U process, mean-reverting squared-root processes, stochastic volatility models, jump processes, regime-switching models or some combination of these processes. Computational power alone limits the flexibility of the model framework.
a reference measure \( Q \) such that under \( Q \), \((\theta, X, V)\) and \( \Phi \) become independent, \((\theta, X)\) remains the same probability law and \( \Phi \) become a Poisson random measure with compensator \( \mu(dy)dt \). Therefore, \( Q \) can be decomposed as \( Q = P_{\theta,x} \times Q_\Phi \), where \( Q_\Phi \) is the probability measure for the Poisson random measure on \( D_\mathbb{Y}[0, \infty) \). Then, one can obtain the the Radon-Nikodym derivative guaranteed by Assumption 4.4, (see Proposition 14.4.1 of Daley & Vere-Jones (2008), or Theorem 10.1.3. of Last & Brandt (1995)) and denoted by \( L(t) \) as given below:

\[
L(t) = \frac{dP}{dQ} \bigg|_{\mathcal{F}_t} = \frac{dP_{\theta,x|\mathcal{F}_t}}{dP_{\theta,x|\mathcal{F}_t}} \times \frac{dP_{\Phi|\theta,x|\mathcal{F}_t}}{dP_\Phi} \bigg|_{\mathcal{F}_t} = \frac{dP_{\Phi|\theta,x|\mathcal{F}_t}}{dP_\Phi} \bigg|_{\mathcal{F}_t}
\]

\[= L(0) \exp \left\{ \int_0^t \int_\mathbb{Y} \log \zeta(s, y) \Phi(d(s, y)) - \int_0^t \int_\mathbb{Y} (\zeta(s, y) - 1) \mu(dy)ds \right\}, \tag{5.1} \]

where \( \zeta(t, y) \) is provided by (4.5) and \( L(0) \) is a \( \mathcal{F}_0 \)-measurable random variable with \( E^Q(L(0)) = 1 \). \( L(t) \) is a local martingale and can be written in a SDE form:

\[ L(t) = L(0) + \int_0^t \int_\mathbb{Y} (\zeta(s, y) - 1) L(s-) (\Phi(d(s, y)) - \mu(dy)ds). \tag{5.2} \]

Because of Assumption 4.4, \( L(t) \) is actually a martingale. The SDE form plays an important role in deriving the filtering equations.

Given an observed sample path of \((\theta_1, X, \Phi)\), where \( \theta_1 \) is the time-varying component of \( \theta \) such as stochastic volatility or regime-shifting, \( L(t) \) becomes the joint likelihood of \((\theta_1, X, \Phi)\), which depends on \( V \). In many cases, \( L(0) = 1 \) as in Zeng (2003), and we offer four typical examples of \( L(t) \).

**Example 5.1 ACD Models** with \( \mathbb{Y} = \{1\} \) and observations \( \{t_i\}_{i=1}^n \). Note that \( \mu(\{1\}) = 1 \) and \( \zeta(t, 1) = \bar{\lambda}(t) \), which is given in Table II. There are two different joint likelihoods for ACD Models. The one with respect to a unit Poisson Process \( Y(t) = \Phi(t, \{1\}) \) derived from (5.1) is given below:

\[
L(t_n) = \exp \left\{ \int_0^{t_n} \log \bar{\lambda}(s) dY(s) - \int_0^{t_n} (\bar{\lambda}(s) - 1) ds \right\} = \prod_{i=1}^n \bar{\lambda}(t_i) \exp \left\{ - \int_0^{t_n} (\bar{\lambda}(s) - 1) ds \right\}. \tag{5.3} \]

Observe that \( \bar{\lambda}(t_i) = r_{T_i-1}(t_i) = r_{d-1}(\Delta t_i) \). Then, the likelihood with respect to the usual Lebesgue product measure derived from (3.3) is given by

\[
L^*(t_n) = \prod_{i=1}^n r_{T_i-1}(\Delta t_i) \exp \left\{ - \sum_{i=1}^n \int_{t_{i-1}}^{t_i} r_{d-1}(s - t_{i-1}) ds \right\} = \prod_{i=1}^n \bar{\lambda}(t_i) \exp \left\{ - \int_0^{t_n} \bar{\lambda}(s) ds \right\}. \tag{5.4} \]

The likelihood in (5.4) divided by \( \exp(-t_n) \), the adjustment due to the change of the reference measure, becomes the one in (5.3).
However, it should be noted that there is no measure

The mark space $\mathcal{Y}$ is a bounded subset of $\mathbb{R}$ with $\mu(dy) = dy$, which is Lebesgue measure. Note that $\zeta(t, y) = \lambda(t)q(y)$ where $q(y_i) = q(y_i|\Delta t_i, \Delta t_{i-1}, \tilde{y}_{i-1}; \theta)$ is given in (3.2). Similarly, there are two different joint likelihood functions. The one with respect to a Poisson random measure with the compensator $\gamma_Q(d(t, y)) = dydt$ from (5.1) is:

$$L(t_n) = \exp \left\{ \int_0^{t_n} \int_0^y \log \left( \lambda(s)q(y) \right) \Phi(d(t, y)) - \int_0^{t_n} \int_0^y \left( \lambda(s)q(y) - 1 \right) dyds \right\}$$

$$= \left( \prod_{i=1}^n \lambda(t_i)q(y_i) \right) \exp \left\{ - \int_0^{t_n} \left( \lambda(s) - \mu(\mathcal{Y}) \right) ds \right\},$$

and the one with respect to the usual Lebesgue product measure is given by

$$L^*(t_n) = \left( \prod_{i=1}^n r_d^{T_{i-1}}(\Delta t_i) \right) \exp \left\{ - \sum_{i=1}^n \int_{t_{i-1}}^{t_i} r_d^{-1}(s-t_{i-1})ds \right\} \left( \prod_{i=1}^n q(y_i) \right)$$

$$= \left( \prod_{i=1}^n \lambda(t_i)q(y_i) \right) \exp \left\{ - \int_0^{t_n} \lambda(s)ds \right\}. $$

The likelihood in (5.4) divided by $\exp(-\mu(\mathcal{Y})t_n)$, the adjustment due to the change of the reference measure, becomes the one in (5.5).

Since there is no latent continuous-time $X$ in ACD models and Engle’s framework, the likelihood functions can be obtained in both ways and both joint likelihoods can be computed directly. Because the adjustments are independent of any parameter, the inferences based on maximum likelihood procedure adopted in Engle & Russell (1998) and Engle (2000) using either likelihood coincide. In the next two examples with latent continuous-time $X$, their joint likelihoods only exist with respect to some suitable Poisson random measures.

**Example 5.3** In the model of Zeng (2003), $\mathcal{Y} = \{ y_1, y_2, \ldots, y_m \}$, $\mathcal{Y}$ is all subsets of $\mathcal{Y}$ and $\mu(dy) = 1$ for all $y \in \mathcal{Y}$. Then, $\zeta(t, y_j) = \lambda_j(t) = \lambda_j(\theta(t), X(t), t)$ and $\Phi(t, \{ y_j \}) = Y_j(t)$. With the integration becomes summation in discrete case, $L(t)$ of (5.1) becomes

$$L(t) = \exp \left\{ \sum_{j=1}^m \int_0^t \log \lambda_j(\theta(s^-), X(s^-), s^-)dY_j(s) - \sum_{j=1}^m \int_0^t \left[ \lambda_j(\theta(s^-), X(s^-), s) - 1 \right] ds \right\}.$$

In all the previous three examples, $\mathcal{Y}$ is finite or bounded with $\mu(dy) = 1$ or $dy$. The next example has unbounded $\mathcal{Y}$ and a nontrivial $\mu(dy)$.

---

6In the model of Zeng (2003), given the whole path of $X$, the likelihood with respect to the usual Lebesgue product measure can be written as

$$L^*(t) = \exp \left\{ \sum_{j=1}^m \int_0^t \log \lambda_j(\theta(s^-), X(s^-), s^-)dY_j(s) - \sum_{j=1}^m \int_0^t \lambda_j(\theta(s^-), X(s^-), s)ds \right\}.$$  

However, it should be noted that there is no measure $Q$ such that $E^Q[L^*(t)] = 1$. Hence, $L^*(t)$ is not a joint likelihood in this example.
Example 5.4  For the Heston Stochastic Volatility Model in Example 4.1, \( \mathcal{Y} = \mathbb{R} \), \( \mathcal{Y} = \mathcal{B}(\mathbb{R}) \) and \( \mu(dy) = \frac{1}{\sqrt{2\pi} \nu} \exp\left\{-\frac{1}{2\nu^2}y^2\right\}I_{\mathbb{R}}(y)dy \). Note that \( p(dy|X(t)) = \frac{1}{\sqrt{2\pi} \nu} \exp\left\{-\frac{1}{2\nu^2}(y - X(t))^2\right\}I_{\mathbb{R}}(y)dy \) and

\[
\zeta(t,y) = \frac{\lambda(t)p(dy|X(t))}{\mu(dy)} = \lambda(t)\exp\left\{\frac{1}{2\nu^2}(2X(t)y - X^2(t))\right\}.
\]

Moreover, assuming \( \lambda(t) \) is independent of marks \( Y_i \) but may depend on \( (\theta(t), X(t)) \) and with some simplification, we have \( \int_\mathbb{R}[\zeta(s-, y) - 1]\mu(dy) = \lambda(s-) - 1 \). Hence, \( L(t) \) of (5.1) becomes

\[
L(t) = \exp\left\{\int_0^t \int_\mathbb{R} \left( \log \lambda(s) + \frac{1}{2\nu^2}(2X(s)y - X^2(s)) \right) \Phi(ds, dy) - \int_0^t [\lambda(s) - 1]ds. \right\}.
\]

5.1.2. The Continuous-time Marginal Likelihoods of \( \Phi \)

Since \( (X, \theta) \) is not observable, the joint likelihood is not computable in Examples 5.3 and 5.4. For the statistical analysis, we need the marginal likelihood of \( \Phi \) alone, which can be concisely expressed by conditional expectation.

Definition 5.1  Let \( \rho_t \) be the conditional measure of \( (\theta(t), X(t)) \) given \( \mathcal{F}_t^{\Phi,V} \) defined as

\[
\rho_t \{ (\theta(t), X(t)) \in A \} = E^Q \left[ I_{\{((\theta(t), X(t)) \in A\}}(\theta(t), X(t))L(t)|\mathcal{F}_t^{\Phi,V} \right].
\]

Definition 5.2  Let

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

If \( (\theta(0), X(0)) \) is fixed, then the marginal likelihood of \( \Phi \) for a frequentist is obtained by taking \( f = 1 \): \( E^Q[L(t)|\mathcal{F}_t^{\Phi,V}] = \rho(1, t) \). For the ACD model and Engle’s framework in Examples 5.1 and 5.2, \( L(t_n) \) has no latent continuous-time \( (\theta, X) \) and is \( \mathcal{F}_t^{\Phi,V} \)-measurable. Hence, \( \rho(1, t_n) = E^Q[L(t_n)|\mathcal{F}_t^{\Phi,V}] = L(t_n) \). But, the marginal likelihoods of Examples 5.3 and 5.4 are not that simple and they involve continuous-time pathwise integral, which is of infinite dimension and more challenging in computation than that of the latent discrete-time model also proposed in Engle (2000)[pg. 5]. Using the numerical scheme developed in Section 6.2, the marginal likelihood can be computed and the maximum likelihood procedure can be carried out for inference. However, we choose Bayesian approach in this paper because of the computational advantage delivered by the recursiveness of filters as well as the desirable properties of Bayesian inference.

Table III presents a multi-fold comparison up to marginal likelihoods of the four models: ACD Models, Engle’s framework, the previously-reviewed filtering model and our proposed model. From Table III, we can see pro and con of these four models and how the proposed model provides a more general unifying framework and combines the strengths of the previous three models.

If a prior distribution is assumed on \( (\theta(0), X(0)) \) as in Bayesian paradigm, then the integrated (or marginal) likelihood of \( \Phi \) is also \( \rho(1, t) \). When prior information is accessible, Bayesian approach provides a scientific way to incorporate the prior information. If no prior information is available, noninformative priors can be employed.
<table>
<thead>
<tr>
<th>$n^t \mathbb{L}(\gamma^t_1)<em>{\mathcal{D}} = (\gamma^t_1)</em>{\mathcal{D}}$</th>
<th>$n^t \mathbb{L}(\gamma^t_1)<em>{\mathcal{D}} = (\gamma^t_1)</em>{\mathcal{D}}$</th>
<th>same as above</th>
<th>same as above</th>
<th>same as above</th>
</tr>
</thead>
<tbody>
<tr>
<td>${\exp(\lambda^t f(1-(\lambda^t)^s))} \exp(s \lambda^t f(1-(\lambda^t)^s)) \exp(s \lambda^t f(1-(\lambda^t)^s)) \exp(s \lambda^t f(1-(\lambda^t)^s)) \exp(s \lambda^t f(1-(\lambda^t)^s)) \exp(s \lambda^t f(1-(\lambda^t)^s))$</td>
<td>${\exp(\lambda^t f(1-(\lambda^t)^s))} \exp(s \lambda^t f(1-(\lambda^t)^s)) \exp(s \lambda^t f(1-(\lambda^t)^s)) \exp(s \lambda^t f(1-(\lambda^t)^s)) \exp(s \lambda^t f(1-(\lambda^t)^s)) \exp(s \lambda^t f(1-(\lambda^t)^s))$</td>
<td>measure</td>
<td>measure</td>
<td>measure</td>
</tr>
<tr>
<td>${\exp(\lambda^t f(1-(\lambda^t)^s))} \exp(s \lambda^t f(1-(\lambda^t)^s)) \exp(s \lambda^t f(1-(\lambda^t)^s)) \exp(s \lambda^t f(1-(\lambda^t)^s)) \exp(s \lambda^t f(1-(\lambda^t)^s)) \exp(s \lambda^t f(1-(\lambda^t)^s))$</td>
<td>${\exp(\lambda^t f(1-(\lambda^t)^s))} \exp(s \lambda^t f(1-(\lambda^t)^s)) \exp(s \lambda^t f(1-(\lambda^t)^s)) \exp(s \lambda^t f(1-(\lambda^t)^s)) \exp(s \lambda^t f(1-(\lambda^t)^s)) \exp(s \lambda^t f(1-(\lambda^t)^s))$</td>
<td>random measure</td>
<td>random measure</td>
<td>random measure</td>
</tr>
</tbody>
</table>

**Table III — Comparison of Four Models**

<table>
<thead>
<tr>
<th>Model</th>
<th>Frames model</th>
<th>Future frames model</th>
<th>Endless frames model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proposed</strong></td>
<td><strong>Previous</strong></td>
<td><strong>Earlier</strong></td>
<td><strong>Endless</strong></td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.1.3. The Continuous-time Posterior of \((\theta, X)\)

**Definition 5.3** Let \(\pi_t\) be the conditional distribution of \((\theta(t), X(t))\) given \(\mathcal{F}_{t}^{\Phi, V}\) and let

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

If a prior distribution is assumed on \((\theta(0), X(0))\) as in Bayesian paradigm, then \(\pi_t\) becomes the continuous-time posterior distribution, which is determined by \(\pi(f, t)\) for all continuous and bounded \(f\). The posterior summarizes Bayesian statistical information 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 Bayes estimators 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 confidence intervals: the credible set and the highest posterior density region.

Bayes Theorem (Bremaud (1981), page 171) provides the relationship between \(\rho(f, t)\) and \(\pi(f, t)\):

\[
(5.8) \quad \pi(f, t) = \frac{\rho(f, t)}{\rho(1, t)}.
\]

Namely, the conditional probability measure \(\pi\) is obtained by normalizing the conditional measure \(\rho\) by its total measure. Hence, 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.

5.1.4. Continuous-time Likelihood Ratios, Bayes Factors and Posterior Model Probabilities

With the observation \(\Phi = \{T_i, Y_i\}_{i=1}^{n}\) and the observable factor \(V\), we would like to compare \(K\) models. Denote Model \(k\) by \((\theta^{(k)}, X^{(k)}, \Phi^{(k)}, V^{(k)})\) for \(k = 1, 2, ..., K\) with \(\mathbb{P}^{(k)} = \mathbb{P}_{\theta^{(k)}, x^{(k)}}^{(k)} \times \mathbb{P}_{\Phi^{(k)}, \theta^{(k)}, x^{(k)}, v}^{(k)}\).

With respect to \(Q^{(k)} = \mathbb{P}_{\theta^{(k)}, x^{(k)}, v}^{(k)} \times Q_\Phi\) where \(Q_\Phi\) is common under which the compensator \(\Phi\) is \(\mu(dy)dt\), is the joint likelihood of \((\theta^{(k)}, X^{(k)}, \Phi^{(k)}, V^{(k)})\) denoted by \(L^{(k)}(t)\). Since the available information, \(\mathcal{F}_{t}^{\Phi, V}\), is the same for all models, we denote

\[
(5.9) \quad \rho_k(f_k, t) = E^{Q}[f_k(\theta^{(k)}(t), X^{(k)}(t))L^{(k)}(t)|\mathcal{F}_{t}^{\Phi, V}].
\]

Then, the marginal likelihood of \(\Phi\) is \(\rho_k(1, t)\), for Model \(k\).

Now, we define Bayes factor, which is at the heart of Bayesian hypothesis testing and model selection.

**Definition 5.4** The Bayes factor of Model \(k\) over Model \(j\), \(B_{kj}\), is defined as the ratio of integrated likelihoods of Model \(k\) over Model \(j\):

\[
(5.10) \quad B_{kj}(t) = \frac{\rho_k(1, t)}{\rho_j(1, t)}.
\]

Thus, we usually interpret the Bayes factor as the “odd provided by the data for Model \(k\) versus Model \(j\)”. With the priors being the “weighting functions”, \(B_{kj}\) is sometimes called the “weighted likelihood ratio of Model \(k\) to Model \(j\)”. The Bayes factor can also be written as the model posterior to model prior odds ratio, which desirably brings prior and posterior information into one ratio and
TABLE IV

<table>
<thead>
<tr>
<th>$B_{kj}$</th>
<th>Evidence against Model $j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 to 1</td>
<td>Negative</td>
</tr>
<tr>
<td>1 to 3</td>
<td>Not worth more than a bare mention</td>
</tr>
<tr>
<td>3 to 20</td>
<td>Positive</td>
</tr>
<tr>
<td>20 to 150</td>
<td>Strong</td>
</tr>
<tr>
<td>&gt; 150</td>
<td>Decisive</td>
</tr>
</tbody>
</table>

Interpretation of Bayes Factor

Supplies the evidence of one model specification over another. In summary, the Bayes factor allows us to measure the relative fit of one model versus another one given the observed, or in our case, partially-observed data.

Suppose that $B_{kj}$ has been calculated. Then, we can interpret it using Table IV furnished by Kass & Raftery (1995) as guideline.

Instead of normalizing by its integrated likelihood as in $\pi_t$ or $\pi(f, t)$, when the conditional measure $\rho_k$ is normalized by the integrated likelihood of Model $j$, another useful characterizing conditional measure is obtained.

**Definition 5.5** For $k = 1, 2, \ldots, K$ and $j \neq k$ in the same range, let $q^{(kj)}_t$ be the ratio conditional measure of $(\theta^{(k)}(t), X^{(k)}(t))$ given $\mathcal{F}^P_t$ with respect to Model $j$. $q^{(kj)}_t$ is defined by: for $A \in \mathbb{R}^{p^{(k)} + m^{(k)}}$,

$$ q^{(kj)}_t \left\{ (\theta^{(k)}(t), X^{(k)}(t)) \in A \right\} = \frac{E_Q \left[ I_{\{\theta^{(k)}(t), X^{(k)}(t)) \in A\}}{\rho_j(1, t)} \left( \theta^{(k)}(t), X^{(k)}(t)) L^{(k)}(t) | \mathcal{F}^P_t \right] \right\} $$

where $L^{(k)}(t) = L^{(k)}(t)/\rho_j(1, t)$.

In the second equality, we can move $\rho_j(1, t)$ inside the conditional expectation because $\rho_j(1, t)$ is $\mathcal{F}^P_t$ measurable.

**Definition 5.6** Let the filter ratio processes for $f_k$ of Model $k$ and $f_j$ of Model $j$ be:

$$ q_{kj}(f_k, t) = \frac{\rho_k(f_k, t)}{\rho_j(1, t)}, \quad \text{and} \quad q_{jk}(f_j, t) = \frac{\rho_j(f_j, t)}{\rho_k(1, t)}. $$

Observe that the Bayes factor becomes $B_{kj}(t) = q_{kj}(1, t)$. Moreover, $q_{kj}(f_k, t)$ can be written as

5.11  $$ q_{kj}(f_k, t) = \int f_k(\theta^{(k)}, x^{(k)}) q^{(kj)}_t(d(\theta^{(k)}, x^{(k)})). $$

**Remark 5.1** The integral forms of $\rho(f, t)$, $\pi(f, t)$, and $q_{kj}(f_k, t)$ are important in deriving the recursive algorithms where $f$ (or $f_k$) is taken to be a lattice-point indicator function.
If prior model probabilities \( P(M_k | \mathcal{F}_0^\Phi, V_0) \), \( k = 1, 2, \ldots, K \), are available, then we can calculate, by Bayes formula, the posterior probabilities of the models, \( P(M_k | \mathcal{F}_t^\Phi, V_t) \). They can also be expressed by Bayes factors as below:

\[
P(M_k | \mathcal{F}_t^\Phi, V_t) = \frac{P(M_k | \mathcal{F}_0^\Phi, V_0) \rho_k(1,t)}{\sum_{l=1}^K P(M_l | \mathcal{F}_0^\Phi, V_0) \rho_l(1,t)} = \left[ \sum_{l=1}^K \frac{P(M_l | \mathcal{F}_0^\Phi, V_0) q_{lk}(1,t)}{P(M_k | \mathcal{F}_0^\Phi, V_0)} \right]^{-1}.
\]

In the uninformative case, a typical option of the prior model probabilities is \( P(M_k | \mathcal{F}_0^\Phi, V_0) = 1/K \) for \( k = 1, 2, \ldots, K \), namely, every model has equal initial probability. Then, the posterior model probabilities become:

\[
P(M_k | \mathcal{F}_t^\Phi, V_t) = \frac{\rho_k(1,t)}{\sum_{l=1}^K \rho_l(1,t)} = \left[ \sum_{l=1}^K q_{lk}(1,t) \right]^{-1}.
\]

Given the observations, the posterior model probabilities provide how likely each candidate model is and the needed weights to account for model uncertainty in Bayesian averaging. If one model has to be selected, one reasonable and commonly-used criterion is to select the model with the highest posterior model probability.

Berger & Pericchi (2001) offers seven compelling reasons to use the Bayesian approach for model selection. Most of the reasons are suitable for the proposed models and cited below. As described above, Bayes factors and posterior model probabilities are easy to understand, and the Bayesian approach to model selection is conceptually the same, regardless of the number of models under consideration. Bayesian model selection is consistent. Namely, under very mild conditions, if one of the candidate models is actually the true model, then Bayesian model selection will ensure the selection of the true model with enough observations. Most of classical model selection approaches such as p-values and AIC does not possess consistency. Moreover, even if the true model is not a candidate, Bayesian model selection will select the one among the candidates which is closest to the true model in terms of Kullback - Leibler divergence. Unlike classical statistics using an ad hoc penalty term (as in AIC), Bayesian model selection procedures are automatic Ockham’s razors, preferring simpler models over more sophisticated models when the data provide roughly equivalent fits. Unlike classic model selection, the Bayesian approach does not require nested model and the corresponding probability measures to be absolutely continuous. This is particularly important for the model selection of models involving stochastic process, because absolute continuity is not common among the probability measures of stochastic processes. Finally, the Bayesian approach can account for model uncertainty via Bayesian averaging.

5.2. Filtering Equations

Stochastic partial differential equations (SPDEs) provide an powerful machinery to stipulate the evolution of the infinite dimensional continuous-time conditional measures, determining the likelihoods, the posteriors, the likelihood ratios and Bayes factors and the posterior model probabilities. The following three theorems with proofs in Appendix A summarize all the useful SPDEs.

THEOREM 5.1 Suppose that \((\theta, X, \Phi, V)\) satisfies Assumptions 4.1 - 4.5. Then, \(\rho_t\) is the unique measure-valued solution of the SPDE under \(Q\), the unnormalized filtering equation,

\[
\rho(f,t) = \rho(f,0) + \int_0^t \rho(A_s f, s) ds + \int_0^t \int_\Omega \rho((\zeta(y) - 1)f, s-)\Phi(d(s,y)) - \mu(dy) ds,
\]
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 (4.5).

The unnormalized filtering equation characterizes the evolution of the unnormalized conditional measure, whose total measure is the marginal likelihood. Under \( \mathcal{Q} \), the compensator of \( \Phi \) is \( \mu(dy)dt \) and the double integral is a martingale. Hence, the unnormalized filter has a semimartingale representation. For ACD models and Engle’s framework, no latent \( X \) implies the marginal likelihood \( \rho(1, t) = \mathcal{E}^\mathcal{Q}(L(t)|\mathcal{F}^\mathcal{Q}_t) = L(t) \). Taking \( f = 1 \), Equation (5.12) reduces to Equation (5.2).

**Theorem 5.2** Suppose that \( (\theta, X, \Phi, V) \) satisfies Assumptions 4.1 - 4.5. Then, \( \pi_t \) is the unique measure-valued solution of the SPDE under \( \mathbb{P} \), the normalized filtering equation,

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

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

\[
\pi(f, t) = \pi(f, 0) + \int_0^t \pi(\mathbf{A}_v f, s) ds + \int_0^t \int_\mathcal{Y} \left[ \pi(r(y), s-) - \pi(f, s-) \right] \Phi(d(s, y)),
\]

where \( r(y) = r(y; \theta(s), X(s), \Phi s-, V s-, s-) \) is defined in (4.4).

The normalized filtering equation characterizes the evolution of the conditional distribution, which becomes the posterior when a prior is given. Under \( \mathbb{P} \), it can be shown that \( \pi(\zeta(y), t) \mu(dy) dt \) is a \( \mathcal{F}_t^{\Phi, V} \)-predictable compensator of \( \Phi \). Hence, the normalized filter also has a semimartingale representation.

**Theorem 5.3** Suppose Model \( k (k = 1, 2, ..., K) \) has generator \( \mathbf{A}_v^{(k)} \) for \( (\theta^{(k)}, X^{(k)}) \), the trading intensity \( \lambda_k(t) = \lambda_k(\theta^{(k)}(t), X^{(k)}(t), \Phi^{(k)}(t), V^{(k)}(t)) \), and the transition probability \( p^{(k)}(dy|x) \) from \( x = X(t) \) to \( dy \) for the random transformation \( F^{(k)} \). Suppose that \( (\theta^{(k)}, X^{(k)}, \Phi^{(k)}, V^{(k)}) \) satisfies Assumptions 4.1 - 4.5 with a common reference measure \( Q_\Phi \), under which the compensator of \( \Phi \) is \( \mu(dy) dt \). Then, \( (\pi^{(k)}_t, q_t^{(k)}) \) are the unique measure-valued pair solution of the following system of SPDEs, the ratio filtering equations,

\[
\begin{align*}
q_{kj}(f_k, t) &= q_{kj}(f_k, 0) + \int_0^t q_{kj}(\mathbf{A}_v^{(k)} f_k, s) ds \\
&+ \int_0^t \int_\mathcal{Y} \left[ q_{kj}(\zeta_k(y), s-) - q_{kj}(1, s-) \right] \left( \Phi(d(s, y)) - \frac{q_{kj}(\zeta_k(y), s)}{q_{kj}(1, s)} \mu(dy) ds, \right)
\end{align*}
\]

\[
\begin{align*}
q_{jk}(f_j, t) &= q_{jk}(f_j, 0) + \int_0^t q_{jk}(\mathbf{A}_v^{(j)} f_j, s) ds \\
&+ \int_0^t \int_\mathcal{Y} \left[ q_{jk}(\zeta_k(y), s-) - q_{jk}(1, s-) \right] \left( \Phi(d(s, y)) - \frac{q_{jk}(\zeta_k(y), s)}{q_{jk}(1, s)} \mu(dy) ds, \right)
\end{align*}
\]
for all \( t > 0 \), \( f_k \in D(A^{(k)}) \) and \( f_j \in D(A^{(j)}) \). When \( \bar{\lambda}_k(\theta^{(k)}(t), X^{(k)}(t), \Phi^{t-, V^{t-}}, t) = \bar{\lambda}_j(\theta^{(j)}(t), X^{(j)}(t), \Phi^{t-, V^{t-}}, t) \) and letting \( r_k(y) = p^{(k)}(dy|x)/\mu(dy) \), we can simplify the above two equations as

\[
q_{kj}(f_k,t) = q_{kj}(f_k,0) + \int_0^t q_{kj}(A^{(k)}_{s}, f_k, s) ds \\
+ \int_0^t \int_{\mathcal{Y}} \left[ \frac{q_{kj}(f_k r_k(y), s-)}{q_{kj}(r_j(y), s-)} q_{kj}(1, s-) - q_{kj}(f_k, s-) \right] \Phi(d(s, y)),
\]

\[
q_{jk}(f_j,t) = q_{jk}(f_j,0) + \int_0^t q_{jk}(A^{(j)}_{s}, f_j, s) ds \\
+ \int_0^t \int_{\mathcal{Y}} \left[ \frac{q_{jk}(f_j r_j(y), s-)}{q_{jk}(r_j(y), s-)} q_{kj}(1, s-) - q_{kj}(f_j, s-) \right] \Phi(d(s, y)).
\]

The system of evolution equations for \( q_{kj}(f_k,t) \) and \( q_{jk}(f_j,t) \), characterizes the two conditional measures, whose totals are the likelihood ratios or the Bayes factors with priors. Note that \( \frac{q_{ik}(\zeta_j(y), s)}{q_{ik}(1, s)} \mu(dy) ds = \pi(\zeta_j(y), s) \mu(dy) ds \), which is the compensator of \( \Phi \) under \( P \) for Model \( j \). Hence, the ratio filters of the models have semimartingale representations under suitable measures.

Note that in the case of exogenous sampling, namely, when event times follow an ACD model or a Cox model as described in Table II, \( \{T_i\} \)’s stochastic intensities \( \lambda(t) \) or \( \bar{\lambda}_k(t) \) or \( \lambda_j(t) \) are \( \mathcal{F}_t^{\Phi, V} \)-predictable. Then, Equations (5.13), (5.15) and (5.16) are simplified to (5.14), (5.17) and (5.18). The advantage is the great reduction of the computation in carrying out the Bayes estimation and calculating the Bayes factors. Moreover, observe that the stochastic intensities disappear in the simplified filters. Hence, the parameters of the model of \( \{T_i\} \) can be estimated separately via other approaches such as ML for an ACD model. The rest of parameters related to \( \Phi \) or the noise can be estimated via the Bayesian approach via filtering to be developed in the next section, and their Bayes estimates are model-free of the assumptions on durations. The example to be given in Section 7 takes such advantages. The tradeoff is the exclusion of the endogenous sampling, namely, the exclusion of the relationship between durations and \((\theta_1, X)^7\).

5.2.1. Two Features of Filters

First, all the above three filters can be split up into the propagation equation and the updating equation. We illustrate the separation by the simplified normalized filter (5.14). Let the arrival times be \( t_1, t_2, \ldots \), then, Equation (5.14) can be dissected into the propagation equation, describing the evolution without event occurrence:

\[
\pi(f, t_{i+1-}) = \pi(f, t_i) + \int_{t_i}^{t_{i+1-}} \pi(A_v f, s) ds,
\]

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

\[
\pi(f, t_{i+1}) = \pi(r(y) f, t_{i+1-}) / \pi(r(y), t_{i+1-}),
\]

\(^7\)The time-varying \( \theta_1 \) can be stochastic volatility, which might have an impact on durations as first documented in Engle & Russell (1998) and recently in Renault & Werker (2010) via causality test.
where the mark at time $t_{i+1}$ is assumed to be $y$. Note that the propagation equation has no random component, implying that $\pi_t$ evolves deterministically when there are no event arrival. In the simplified normalized filter, $\pi_t$ actually evolves in the same way as the distribution of $(\theta, X)$ does. The updating equation is random because the observation mark $y$ is random.

Second, all the above filters have the recursiveness. Again, we illustrate this by (5.14), where the goal is to find expressions for $\pi_t$, the posterior in Bayesian paradigm. The normalized filter (5.14) delivers a preferable expression in recursive form as Kalman filters\(^8\). In nonmathematical terms, the expression supplies a device $H$ such that for all $t$,

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

where $\Phi(d(t,y))$ is the innovation. The device $H$ can be regarded as a computer program not depending on data. $\pi_t$ is all that has to be stored in memory, and $\Phi(d(t,y))$ is the new data that is fed into the program $H$. If no new mark point arrives during $[t, t + dt)$, $H$ is determined by the propagation equation (5.19). If a new mark point arrives, $H$ is specified by the updating equation (5.20). Note that $\pi_{t+dt}$ only depends on $\pi_t$, but not the whole past $\{\pi_s : 0 \leq s \leq t\}$. This saves memory space. In statistical jargons, we may say that $(\pi_t(X, \theta), d\Phi_t)$ is a sufficient statistic\(^9\) for the problem of “estimating” $\pi_{t+dt}(X, \theta)$ on the basis of $F_{t}^{t+dt}$. In summary, the practical implication of recursiveness is to provide update of the posteriors, and in general, the implementation of Bayesian model identification and selection, in real time for streaming UHF data.

6. BAYESIAN INFERENCE VIA FILTERING

The filtering formulation furnishes a natural setup for Bayesian inference. Theorems 5.1, 5.2 and 5.3 deliver the filtering equations of the related continuous-time conditional measures, which are defined on $D_{\mathbb{R}^p+m \times Y}[0, \infty)$ and thus are all infinite dimensional. To compute them, one needs to reduce the infinite dimensional problem to a finite dimensional problem so as to construct algorithms. Based on the SPDEs, the algorithms for Bayesian inference are naturally recursive, handling a datum at a time. Moreover, the algorithms are easily parallelizable according to time-invariant parameters. Thus, the algorithm can make real-time updates, do real-time Bayesian inference and handle streaming UHF data.

One basic requirement for the recursive algorithms is consistency (also called robustness): The approximate conditional measures, computed by the recursive algorithms, converge to the true ones. In the following two subsections, we first prove that the weak convergence of the approximate signal guarantees the consistency of all of the useful approximate conditional measures. Built on such result, we provide a blueprint for constructing consistent (or robust) algorithms through Kushner’s Markov chain approximation methods (Kushner & Dupuis (2001)) to compute those approximate fundamental statistic characteristics.

6.1. A Convergence Theorem

Since $(Y, dY)$ is a complete separable metric space, $D_{\mathbb{R}^p+m \times Y}[0, \infty)$ embedded with the Skorohod topology is a complete separable metric space. There, we consider the family of Borel probability measures, topologized with the Prohorov metric. Under such a setup is the weak convergence

\(^8\)In the case of Kalman linear filter, $\pi_t$ is Gaussian, which can be determined by its conditional mean and variance. Hence, a finite dimensional filters can be derived. However, this is not true for the general nonlinear filtering problem such as the one discussed in this paper.

\(^9\)Hansen et al. (2010) apply the same idea of sufficient statistics to refine particle filtering methods.
theorem that we are going to establish. Chapter 3 of Ethier & Kurtz (1986) provides a formal description of all these concepts and the weak convergence employed here. Following the literature, we use the notation, $X_t \Rightarrow X$, to mean $X_t$ converges weakly to $X$ in the Skorohod topology as $\epsilon \to 0$. Such weak convergence is uniform in time in the sense that $X_t \Rightarrow X$ implies $(X_t(t_1), X_t(t_2), \ldots, X_t(t_k)) \Rightarrow (X(t_1), X(t_2), \ldots, X(t_k))$ for every finite set $\{t_1, t_2, \ldots, t_k\} \subset D(X)$, where $D(X) \equiv \{t \geq 0 : P(X(t) = X(t^-)) = 1\}$ (Theorem 3.7.8 of Ethier & Kurtz (1986)).

For Model $k$ ($\theta^{(k)}, X^{(k)}, \Phi^{(k)}, V^{(k)}$) where $k = 1, 2, \ldots, K$, let $(\theta^{(k)}_t, X^{(k)}_t)$ be an approximation of the signal, $(\theta^{(k)}_t, X^{(k)}_t)$. Since $(\theta^{(k)}_t, X^{(k)}_t)$ is a vector of stochastic process with sample paths in $D_{\mathbb{R}^{P+n}}[0, \infty)$, $(\theta^{(k)}_t, X^{(k)}_t)$ can be naturally taken as a sequence of pathwise approximating Markov chain index by $\epsilon$.

We define the observations for an approximate signal for Model $k$ as $\Phi^{(k)}_t = \{(T^{(k)}_t, Y^{(k)}_t)\}$. Then, an approximate model for Model $k$ is $(\theta^{(k)}_t, X^{(k)}_t, \Phi^{(k)}_t, V^{(k)}_t)$ with $P^{(k)}_t = \mathbb{P}^{(k),\epsilon}_{\theta,\epsilon;x} \times \mathbb{P}^{(k),\epsilon}_{\theta,\epsilon;y}$. The corresponding reference measure is $Q^{(k)}_t = \mathbb{P}^{(k),\epsilon}_{\theta,\epsilon;x} \times \mathbb{Q}_\Phi$ where $\mathbb{Q}_\Phi$ is common for all $k$ and all $\epsilon$. Then, the compensator of $\Phi^{(k)}_t = \{(T^{(k)}_t, Y^{(k)}_t)\}$ is $\mu(dy)dt$ under $\mathbb{Q}_\Phi$, and is $\lambda^{(k)}(t, dy)dt = \zeta^{(k)}(t, y)\mu(dy) = \lambda^{(k)}_\epsilon(t)\rho^{(k)}(dy|X^{(k)}_t(t), t)dt$ under $P^{(k)}_t$. Note that the filtration $\mathcal{F}^{\Phi,V}_t$ is the available information for all models of $k$ and $\epsilon$. Let $L^{(k)}_t(t) = \frac{d\rho^{(k)}_t}{dq^{(k)}_t}(t) = L\left((\theta^{(k)}_t(s), X^{(k)}_t(s), \Phi^{(k)}_t(s), s, B), V(s) : 0 \leq s \leq t, B \in \mathcal{Y}\right)$ as in Equation (5.1). Suppose that for $\epsilon > 0$, $(\theta^{(k)}_t, X^{(k)}_t, \Phi^{(k)}_t, V)$ lives on $(\Omega^{(k)}_t, \mathcal{F}_t^{(k)}, P^{(k)}_t)$ with Assumptions 4.1 - 4.5. Next, we define the approximations of $\rho^{(k)}_t(f_k(t), \pi^{(k)}_t(f_k(t), t), q^{(k)}_t(f_k(t), t)$.

**Definition 6.1** For $k = 1, 2, \ldots, K$ and $j \neq k$ in the same range, let $\rho^{(k)}_{t,1}$ be the unnormalized conditional measure of $(\theta^{(k)}_t(t), X^{(k)}_t(t))$ given $\mathcal{F}^{\Phi,V}_t$, and let

$$
\rho^{(k)}_{t,1}(f_k(t), \pi^{(k)}_t(f_k(t), t) = E^{Q^{(k)}_t}\left[f_k(\theta^{(k)}_t(t), X^{(k)}_t(t))I^{(k)}_t(t) | \mathcal{F}^{\Phi,V}_t\right].
$$

Let $\pi^{(k)}_{t,1}$ be the conditional distribution of $(\theta^{(k)}_t(t), X^{(k)}_t(t))$ given $\mathcal{F}^{\Phi,V}_t$, and let

$$
\pi^{(k)}_{t,1}(f_k(t), t) = E^{P^{(k)}_t}\left[f_k(\theta^{(k)}_t(t), X^{(k)}_t(t)) | \mathcal{F}^{\Phi,V}_t\right].
$$

Let $q^{(k)}_{t,1}(f_k(t), t) = \rho^{(k)}_{t,1}(f_k(t), t)/\rho^{(k)}_{t,1}(1, t)$. Let $P^{(k)}_t(M_k|\mathcal{F}^{\Phi,V}_t)$ the posterior probability of approximate Model $k$ given as below:

$$
P^{(k)}_t(M_k|\mathcal{F}^{\Phi,V}_t) = \frac{P(M_k|\mathcal{F}^{\Phi,V}_0)\rho^{(k)}_{t,1}(1, t)}{\sum_{l=1}^{K} P(M_l|\mathcal{F}^{\Phi,V}_0)\rho^{(k)}_{t,1}(1, t)} = \left[\sum_{l=1}^{K} P(M_l|\mathcal{F}^{\Phi,V}_0)q^{(k)}_{t,1}(1, t)\right]^{-1}.
$$

Now, we are ready to state the main convergence theorem.

**Theorem 6.1** Suppose that for $k = 1, 2, \ldots, K$, $(\theta^{(k)}_t, X^{(k)}_t, \Phi^{(k)}_t, V)$ satisfies Assumptions 4.1 - 4.5 with the reference measure $Q^{(k)}_t = \mathbb{P}^{(k),\epsilon}_{\theta,\epsilon;x} \times \mathbb{Q}_\Phi$, and the compensator of $\Phi^{(k)}_t$ is $\mu(dy)dt$ under $\mathbb{Q}_\Phi$. Suppose that Assumptions 4.1 - 4.5 hold for the approximate models $(\theta^{(k)}_t, X^{(k)}_t, \Phi^{(k)}_t, V)$ with the reference measure $Q^{(k)}_t = \mathbb{P}^{(k),\epsilon}_{\theta,\epsilon;x} \times \mathbb{Q}_\Phi$, and the compensator of $\Phi^{(k)}_t$ is $\mu(dy)dt$ under $\mathbb{Q}_\Phi$ for all $k$ and all $\epsilon$. 
If \((\theta^{(k)}_t, X^{(k)}_t) = (\theta_0, X_0)\) as \(\epsilon \to 0\), then, as \(\epsilon \to 0\), for all bounded continuous functions, \(f_k\) and \(f_j\), \(k, j = 1, 2, ..., K\) and \(j \neq k\) in the same range,

1. \(\hat{Y}_t^{(k)} \Rightarrow Y_t^{(k)}\) under physical measures;
2. \(\hat{\rho}_t^{(k)}(f_k, t) \Rightarrow \rho_t(f_k, t);\)
3. \(\hat{\pi}_t^{(k)}(f_k, t) \Rightarrow \pi_t(f_k, t);\)
4. \(q_{k,j}(f_k, t) \Rightarrow q_{k,j}(f_k, t)\) simultaneously for all pairs \((k, j)\) with \(k \neq j;\)
5. \(P_t(M_k, F_t^\Phi, V) \Rightarrow P(M_k| F_t^\Phi, V).\)

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 the observation of the true model, (2) the marginal likelihood of the approximate model to the true marginal likelihood, (3) the posterior of the approximate model to the true posterior, (4) the Bayes factors (or likelihood ratios) of the approximate models to the true Bayes factors (or likelihood ratios), and (5) the posterior model probabilities of the approximate models to the true posterior model probabilities.

6.2. The Blueprint for Recursive Algorithms

Theorem 6.1 provides a three-step blueprint for constructing a consistent and easily-parallelized recursive algorithm based on Markov chain approximation method to compute the continuous-time conditional measures and the needed statistical characteristics. For example, to compute the joint posterior and the marginal posterior mean \(^{10}\) for a model (then the superscript, “(k)” is excluded), we choose the simplified Equation (5.14) for illustrate the major steps in constructing a recursive algorithm, namely, \(H_t(\cdot, \cdot)\) as the approximation of \(H(\cdot, \cdot)\) in (5.21). After discretizing the state space, there are three main steps: Step 1 is to construct \((\theta_\epsilon, X_\epsilon)\), a Markov chain approximation to \((\theta, X)\), and obtain \(r_\epsilon(y) = r(y_\epsilon|\theta_\epsilon, x_\epsilon)\) as an approximation to \(r(y) = r(y|\theta, x)\), where \((\theta_\epsilon, x_\epsilon)\) and \(y_\epsilon\) are restricted to a discretized space of \((\theta_\epsilon, X_\epsilon)\) and \(Y_\epsilon\), respectively. Step 2 is to obtain the filtering equation for \(\pi_\epsilon(f, t)\) corresponding to \((\theta_\epsilon, X_\epsilon, Y_\epsilon, r_\epsilon)\) by applying Theorem 5.2. Recall that the filtering equation for the approximate model can also be dissected into the propagation equation:

\[
(6.1) \quad \pi_\epsilon(f, t_{i+1}^-) = \pi_\epsilon(f, t_i) + \int_{t_i}^{t_{i+1}^-} \pi_\epsilon(A_\epsilon f, s)ds,
\]

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

\[
(6.2) \quad \pi_\epsilon(f, t_{i+1}) = \frac{\pi_\epsilon(fr_\epsilon(y_{i+1}), t_{i+1}^-)}{\pi_\epsilon(r_\epsilon(y_{i+1}), t_{i+1}^-)}.
\]

If no new mark point arrives during \([t, t + dt]\), \(H_\epsilon\) is stipulated by the propagation equation (6.1). If a new mark point arrives, \(H_\epsilon\) is specified by the updating equation (6.2). Step 3 converts Equations (6.1) and (6.2) to the recursive algorithm for the discrete state space and in discrete times by two substeps: (a) represents \(\pi_\epsilon(\cdot, t)\) as a finite array with the components being \(\pi_\epsilon(f, t)\) for lattice-point indicators \(f\) and (b) approximates the time integral in (6.1) with an Euler scheme or other higher order numerical schemes.

\(^{10}\)That is the usual Bayes estimator for a parameter under squared error loss
7. MODELING AND INFERENCE VIA FILTERING: AN APPLICATION TO BOND TRANSACTIONS DATA

This section provides an example for the method of building a specific filtering model, and provides 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 hypotheses regarding market behavior are particularly amenable to the flexible form of our model.

7.1. 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. During this time period, GovPX was the leading interdealer broker for trades in the U.S. Treasury note market (Barclay et al. (2006)), acting as a clearinghouse for market activity transacted through five of the six largest Treasury interdealer brokers (Jordan & Kuipers (2005)). 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 Fleming & Remolona (1999), Huang et al. (2002), Brandt & Kavacejz (2004), and Green (2004), 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 (Fleming (1997)). 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.

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 inference tests regarding competing microstructure hypotheses in the Treasury market using the OTR note transactions data. 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. Numerous studies (e.g., Green (2004)) document that OTR Treasury notes fully respond to new public information upon its release within a few minutes time, at most.

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

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

---

11Research 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 Fleming (2003) and Brandt & Kavacejz (2004) for details.

12The 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 pre-processing is necessary to re-construct the original quotes and trades from this raw data; see Fleming (2003) in particular for details.
7.2.1. The Intrinsic Value Process

First, the state process becomes the intrinsic value process \( X_t \), of which we consider a general form of the GBM process for illustration. Alternative stochastic processes can be accommodated in our framework without loss of generality. We assume the intrinsic value process follows the SDE given by:

\[
\frac{dX_t}{X_t} = \mu dt + (\sigma_0 + \sigma_{en} V_{en}(t) + \sigma_{bs} V_{bs}(t)) dB_t
\]

where \( B_t \) is a standard Brownian Motion, and \((\mu, \sigma_0, \sigma_{bs}, \sigma_{bs})\) 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. Huang et al. (2002) and others find that volatility in the Treasury market increases significantly upon release of macroeconomic news, and Brandt & Kavacejz (2004) provide evidence that information-based trading is the predominant source of the observed change in volatility. Alternatively, the process \( V_{bs}(t) \) indicates whether a trade is buyer or seller initiated. Boni & Leach (2004) and Fleming & Rosenberg (2007) 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 differentiate between these competing theories using simple inferential tests.

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 \( \rho \) is used to model non-clustering noise to be defined later. The generator of (7.1) is:

\[
A_v f(x, \theta) = \mu x \frac{\partial}{\partial x} f(\theta, x) + \frac{1}{2}(\sigma_0 + \sigma_{en} V_{en} + \sigma_{bs} V_{bs})^2 x^2 \frac{\partial^2}{\partial x^2} f(\theta, x).
\]

7.2.2. Trading Durations and Noise

We assume the trade duration follows an EACD or WACD model, which is exogenous sampling. Hence, the simplified version of the normalized filtering equation for the posterior and that of the ratio filtering equation system for the Bayes factors can then 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., Harris (1991)): 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) is a discrete space \( Y = \{a \frac{s+1}{M}, a \frac{s+2}{M}, \ldots, a \frac{s+n}{M}\} \). 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:

\[
P\{U = u\} = \begin{cases} 
(1 - \rho) & \text{if } u = 0 \\
\frac{1}{2}(1 - \rho)\rho^{|u|} & \text{if } u = \pm \frac{1}{2M}, \pm \frac{3}{2M}, \ldots
\end{cases}
\]

**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 \( \{y'_i\} \). For consistency with market convention in the OTR 10-year note market, we construct a simple random biasing function for ticks with \( M = 64 \):

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

Details concerning \( b_i(\cdot) \) and the explicit \( p(y|x) \) for \( F \), both needed for Representation II, can be found in Appendix B. Simulations demonstrate that the \( F(x) \) is able to capture the sample characteristics of the tick transactions data.

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 \( B \subset Y \) with predictable stochastic intensity kernel for \( y \in Y \):

\[
(7.3) \quad \lambda(t, y) = \bar{\lambda}(t)p(y|X(t); \theta, t)
\]

where \( p(y|X(t); t) = p(y|x) \) is the transition probability.

### 7.2.3. Parameters and Recursive Algorithms

There are three groups of parameters. The first group is the parameters for trading duration. Following Engle & Russell (1998), we use cubic-spline interpolation to construct the diurnally-adjusted duration data to choose the best-fit model between EACD and WACD using the method of maximum likelihood.\(^{13}\) The second group is the clustering noise parameters, which can be estimated through the method of relative frequency. The focus of this example is the third group of parameters \( \theta = (\mu, \sigma_0, \sigma_{e\delta}, \sigma_{b\delta}, \rho) \), which are estimated from the Bayes filter using the recursive algorithm.

The choice of priors and the construction of the recursive algorithm for computing the joint posterior and the Bayes estimates is similar to that in Section 5.2.1 of Zeng (2003) but with more dimensions and with \( \sigma \) being replaced by \( \sigma_0 + \sigma_{e\delta}V_{e\delta} + \sigma_{b\delta}V_{b\delta} \). The construction of the recursive algorithm for calculating the related Bayes factors is similar to that illustrated in Section 4.2 of Kouritzin & Zeng (2005a). Interested readers are referred to those two papers for details.

With all Bayes factors available, the posterior model probabilities can be computed. For example, \( P(M_4|\mathcal{F}_t^\Phi) = \frac{1}{1 + B_{14} + B_{24} + B_{34}} \).

---

\(^{13}\)Because trade volume in the Treasury market is concentrated in the first and last hour of the trading day (Fleming (1997)), our stochastic intensity also contains a time-dependent deterministic component.
7.3. Simulation Tests

The recursive algorithms for calculating the Bayes estimates and Bayes factors are fast enough to generate real-time estimates. We develop a computer program for the recursive algorithms, and in this section, we test the program for convergence and model selection 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 (Fleming (1997)), 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 $\lambda(t)$ is not critical in the recursive algorithms, 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 Fleming & Rosenberg (2007): in the absence of inventory management, the arrival time for buyer- and seller-initiated trades should be completely random. In addition, between the 1000th and 1001st transaction, a news announcement impulse occurs. Given the results in Green (2004) 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.

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 V provides the results for the simulated dataset. The Bayes estimates are close to their true values, each within two SE bounds. This is in line with the consistency of the Bayes estimators for this model, which can be proved using the same method described in Appendix B of Zeng (2003). While the estimation error for $\mu$ is much larger than the other parameters, $\mu$ 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, $\mu$ is essentially a nuisance parameter for our purposes.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>Bayes Estimate</th>
<th>St. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>4.00E-08</td>
<td>4.37E-08</td>
<td>1.10E-08</td>
</tr>
<tr>
<td>$\sigma_0$</td>
<td>3.00E-05</td>
<td>3.03E-08</td>
<td>7.71E-07</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.06</td>
<td>0.07</td>
<td>0.01</td>
</tr>
<tr>
<td>$\sigma_{en}$</td>
<td>2.00E-05</td>
<td>2.27E-05</td>
<td>4.77E-06</td>
</tr>
<tr>
<td>$\sigma_{bs}$</td>
<td>1.00E-05</td>
<td>8.69E-06</td>
<td>1.25E-06</td>
</tr>
</tbody>
</table>

To assess model selection in recovering the true data generating process, Table VI reports the value for the Bayes factors ($B_{41}$, $B_{42}$, $B_{43}$, $B_{31}$, $B_{32}$, $B_{21}$) and posterior model probabilities at several iteration points (for all other pairs, use $B_{ij}B_{ji} = 1$). A uniform prior is used initially for each of the Models 1 - 4. At $N = 200$, the Bayes factors $B_{41}$, $B_{42}$, $B_{31}$ and $B_{32}$ exceed the 150 benchmark, favoring models that include the buyer-seller initiation dummy (Model 3 and Model 4). Models 3 and 4 perform equally well ($B_{43} = 1$), so they are assigned the same posterior model probability. The same logic applies to Models 1 and 2 ($B_{12} = 1$); however, the news announcement impulse has not yet been introduced.
Surrounding the simulated news release, at \( N = 1000 \) the Bayes factors \( (B_{41}, B_{42}, B_{31}, B_{32}) \) have increased to 115468, but the model posterior probabilities have not changed. Once the news occurs, the estimated Bayes factors and posterior model probabilities immediately capture the resulting volatility impact. Specifically, at \( N = 1001 \) the Bayes factor \( B_{21} \) jumps from 1 to 2.2, \( B_{43} \) increases from 1 to 2.4, and Model 4 now exhibits a higher posterior model probability (0.71) than does Model 3 (0.29). All of these show that the Bayes factors and the posterior model probabilities are very sensitive in detecting model change.

Once the announcement period ends \( (N = 1050) \), \( B_{21} \) has increased to 7.43E7 and \( B_{43} \) has increased to 180017. At this point, the recursive algorithm strongly prefers Model 4 to the alternatives, assigning 0.99999 for the posterior model probability. For the final 950 simulated trades, the Bayes factors continue to evolve (even sometimes change in large magnitude), but the posterior model probabilities remain stable. The recursive algorithm correctly identifies the true data generating process as Model 4. This confirms the consistency of Bayesian model selection approach (Berger & Pericchi (2001)). Figure 1 plots trade-by-trade posterior model probabilities of Models 1 to 4 with the shaded area indicating news period \( V_{cn} = 1 \) and further validates that the posterior model probabilities provides a sensitive but stable and powerful tool for selecting the right model.

Given these simulation results, there is reason to believe the recursive algorithm can provide model identification and selection 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 in the next subsection.

---

**Figure 1.**—Trade-by-Trade Posterior Model Probabilities (PMP) for Models 1-4 (Simulated data)
7.4. Empirical Results

We use the recursive algorithm to examine tick transaction data for the OTR 10-Year Treasury note during the period 8/15/2000 to 12/31/2000, where Figure 2 is the time series plot with the color-distinctive shaded areas marking news period ($V_{en} = 1$). Table VII 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 2](image)

**Figure 2.**—Times Series of Prices for 10-Year Note.

<table>
<thead>
<tr>
<th>TABLE VII</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary Statistics for the OTR 10-Year Note, 8/15/2000 - 12/31/2000</td>
</tr>
<tr>
<td># trades</td>
</tr>
<tr>
<td>10-Year Note</td>
</tr>
</tbody>
</table>

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 Green (2004) 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 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.
TABLE VIII  
ANNUALIZED BAYES ESTIMATES FOR THE OTR 10-YEAR NOTE, 8/15/2000 - 12/31/2000

<table>
<thead>
<tr>
<th>Model</th>
<th>µ</th>
<th>σ₀</th>
<th>ρ</th>
<th>σₑn</th>
<th>σₚₛ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>14.48%</td>
<td>5.01%</td>
<td>3.49%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(7.77%)</td>
<td>(0.046%)</td>
<td>(0.30%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td>13.95%</td>
<td>4.93%</td>
<td>3.42%</td>
<td>5.37%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(8.22%)</td>
<td>(0.073%)</td>
<td>(0.40%)</td>
<td>(0.62%)</td>
<td></td>
</tr>
<tr>
<td>Model 3</td>
<td>50.33%</td>
<td>5.30%</td>
<td>3.33%</td>
<td></td>
<td>-0.60%</td>
</tr>
<tr>
<td></td>
<td>(6.76%)</td>
<td>(0.06%)</td>
<td>(0.48%)</td>
<td></td>
<td>(0.07%)</td>
</tr>
<tr>
<td>Model 4</td>
<td>50.07%</td>
<td>5.28%</td>
<td>3.21%</td>
<td>4.99%</td>
<td>-0.80%</td>
</tr>
<tr>
<td></td>
<td>(7.39%)</td>
<td>(0.022%)</td>
<td>(0.60%)</td>
<td>(0.38%)</td>
<td>(0.13%)</td>
</tr>
</tbody>
</table>

period trades (Vₑn = 1) on Fed announcement days. In combination, there are 191 transactions in the data that occur with Vₑn = 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 VIII provides the results. The volatility parameter σ₀ is in the range of 4.9% - 5.5%, consistent with historical 10-year Treasury note volatility. The dummies Vₑn and Vₚₛ are both statistically significant. σₑn is positive and significant, consistent with the evidence that volatility in the Treasury market increases when new public information arrives (Fleming & Remolona (1999)). The magnitude of σₑn 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 Braudt & Kavacejz (2004).

For parameter σₚₛ, the estimates are negative and significant for both Models 3 and 4. Given that the dummy Vₚₛ is set to one for buyer-initiated trades, our finding confirms the supposition in Fleming & Rosenberg (2007) 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 σₚₛ is just 15% of σ₀.

Finally, we calculate the Bayes factors to determine which model best describes the OTR 10-year Treasury note transactions data. Table IX reports the daily average Bayes factors and the corresponding posterior model probabilities on selected days. By the third trading day (8/17/2000), the Bayes factors B₁₄, B₁₂, B₃₁ and B₃₂ exceed the benchmark 150 value. These factors increase to roughly 2.5 E6 by the end of the first week of data (8/21/2000), and favor models employing the buyer-seller initiation dummy (Models 3 and 4). The first news event in the sample occurs on 8/22/2000, when the Federal Reserve chose to keep the Fed-funds rate unchanged. The day before this announcement, the recursive algorithm assigns equal posterior probabilities for Models 3 and 4. After the announcement, the recursive algorithm begins to favor Model 4, with the posterior model probability slightly higher (62%) than for Model 3 (38%). Then, the posterior model probabilities keep stable until the second news announcement occurs on 9/1/2000, when the Labor Department released an unemployment report. The market exhibited increased volatility on this news release, and this is reflected in both the Bayes factors and the posterior probabilities. The recursive algorithm picks Model 4 at this point, assigning it a nearly-one (0.999999) posterior model probability. The posterior model probabilities remain stable thereafter for each of the models. Figure 3 plots the trade-by-trade posterior model probabilities for Models 1 to 4 with the same color-distinctive shaded
areas marking news period and further upholds the above description.

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 algorithms described in this paper can provide a new approach for tests of UHF financial market data.

\begin{table}[h]
\centering
\caption{Average Bayes Factors for selected dates, 10-year OTR Note, 8/15/2000 - 12/31/2000}
\begin{tabular}{|c|cccccc|}
\hline
Date & $B_{41}$ & $B_{42}$ & $B_{43}$ & $B_{44}$ & $B_{45}$ & $B_{46}$ & Posterior Model Probability \\
\hline
8/15/2000 & 30.3 & 30.3 & 1.0 & 30.3 & 30.3 & 1.0 & 0.015970 0.015970 0.484032 0.484028 \\
8/16/2000 & 665.7 & 665.7 & 1.0 & 665.7 & 665.7 & 1.0 & 0.000750 0.000750 0.499252 0.499248 \\
8/17/2000 & 3.4E+07 & 3.4E+07 & 1.0 & 3.4E+07 & 3.4E+07 & 1.0 & 0.000000 0.000000 0.589801 0.499999 \\
8/21/2000 & 3.7E+09 & 7.0E+09 & 1.7 & 2.2E+09 & 4.2E+09 & 0.52 & 0.000000 0.000000 0.376743 0.623257 \\
8/30/2000 & 3.5E+16 & 6.6E+16 & 1.7 & 2.1E+16 & 4.0E+16 & 0.52 & 0.000000 0.000000 0.375367 0.624633 \\
9/1/2000 & 1.3E+19 & 3.3E+11 & 1.2E+11 & 1.1E+08 & 2.9E+00 & 4.0E+07 & 0.000000 0.000000 0.089099 1.000901 \\
9/29/2000 & 1.53E+29 & 2.15E+12 & 2.64E+19 & 5.80E+09 & 8.16E+08 & 7.11E+16 & 0.000000 0.000000 0.089099 1.000901 \\
12/29/2000 & 3.41E+33 & 2.91E+07 & 1.84E+33 & 1.02E+08 & 1.45E+27 & 7.42E+14 & 0.000000 0.000000 0.089099 1.000901 \\
\hline
\end{tabular}
\end{table}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{chart.png}
\caption{Trade-by-Trade Posterior Model Probabilities (PMP) for Models 1-4 (Bond data)}
\end{figure}

8. CONCLUSIONS

In this paper, we develop a nonlinear filtering framework for UHF data that uses MPP observations with confounding observable factors. The proposed model unifies the general framework of
Engle (2000) and the filtering model of Zeng (2003), and connects to the literature for estimating realized volatility in the presence of microstructure noise, and to studies that estimate continuous-time Markov processes sampled at random times. Under our framework, complete information of UHF data is used for statistical inference.

Using stochastic filtering, we derive the unnormalized filtering equation to characterize the likelihood function, the normalized filtering equations to characterize the posterior distributions, and the ratio filtering equation system to characterize the likelihood ratio or Bayes factors.

To address the computational challenge, we construct consistent, easily-parallelizable, recursive algorithms via Markov chain approximation methods based on a powerful weak convergence theorem, thus developing the Bayesian inference (estimation and model selection) via filtering for the class of proposed models. The general statistical 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.

This paper provides several new areas for research. First, although the consistency of Bayes estimators for the exemplary model can be proven by an \textit{ad hoc} approach as in Zeng (2003), the consistency as well as the related asymptotic normality and efficiency of the Bayes estimators for the general model are not resolved. Based on simulation results, we conjecture that the consistency and asymptotic normality should be true under mild conditions, even without assuming stationarity of the latent Markov process. Second, as the exemplary model is extended to include stochastic volatility, (Lévy) jumps or regime-switching, more efficient algorithms need to be developed. Besides Markov chain approximation methods running on parallel supercomputing clusters, other numerical methods such as particle filtering, EM algorithms, Poisson chaos expansion methods, Markov chain Monte Carlo (MCMC) and their combination could be useful. Real-time model identification and selection may have a direct impact in the fast-growing high frequency trading in global financial markets, an issue of recent urgency in the United States. 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.

APPENDIX A: PROOFS FOR MAIN RESULTS

A.1. Proof of Proposition 1

If the \textit{stochastic intensity kernels} of the two MPPs are the same, then, they have the same compensator and therefore, have the same probability distribution. Related theorems supporting the above arguments can be found in the book of Last & Brandt (1995), especially, Theorems 8.3.3 and 8.2.2. Therefore, it suffices to show that the two representations have the same stochastic intensity kernel.

The stochastic intensity kernel of Representation II is \( \lambda(t, dy) = \lambda(t)p(dy|X(t); t) \).

By Assumption 4.2, the stochastic intensity kernel of Representation I can be obtained by a heuristic argument given in (2.2) as the following assuming a mark \((Y_i, T_i)\) happens during \([t,t+h]\)

\[
P(\Phi([t, t+h) \times dy) > 0 | \mathcal{F}_{t-}) \approx P\{T_i \in [t, t+h) | \mathcal{F}_{t-}\} P\{Y_i \in dy | T_i \in [t, t+h), \mathcal{F}_{t-}\}) \]

\[
\approx \lambda(\theta(t), X(t), \Phi^{t-}, V^{t-}, t) p(dy | X(t); \theta(t), \Phi^{t-}, V^{t-}, t) h.
\]

This confirms that both representations have the same stochastic intensity kernel. \(\square\)

A.2. Proof of Theorem 1

There are two basic approaches to derive filtering equation: filtering via innovation and filtering via reference measure. We adopt the latter here originally used in Zakai (1969).
First, we determine the SDE for $\phi(f,t)$ and we start with integration by parts (see Protter (2003), page 60),

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

and take $U(t) = f(\theta(t),X(t)), V(t) = L(t)$. Assumption 4.3 implies $f(\theta(t),X(t))$ and $L(t)$ have no simultaneous jumps w.p.1, which implies $[f(\theta,X),L]_t = 0$. Coupled with the fact that $\int A_s f(\theta(s),X(s))ds, L_t = 0$, we have that $[M_f, L]_t = 0$. Then, the Girsanov-Meyer Theorem (see Protter (2003), page 109) implies that $M_f(t)$ is also a martingale under $Q$. Observe that

$$f(\theta(t),X(t))L(t) = f(\theta(0),X(0))L(0) + \int_0^t L(s-)dM_f(s) + \int_0^t \int_s^t f(\theta(s-),X(s-))\left[\zeta(s-,y) - 1\right]L(s-)\Phi(d(s,y))$$

(A.2) $+ \int_0^t L(s)\left( A_s f(\theta(s),X(s)) - \int_s^t f(\theta(s),X(s))\left[\zeta(s,y) - 1\right]\mu(dy)\right)ds.$

We take conditional expectations with respect to the reference measure $Q$ given the observed history of $\mathcal{F}^{X,Y}_t$ on both sides of Equation (A.2). We state four lemmas that deal with the four terms on the right hand side of Equation (A.2). The proofs of these lemmas are available upon request.

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

Observe that $\mathcal{F}^{X}_0 < \mathcal{F}^{Y}_0 < \mathcal{F}^{X,Y}_0$, because of the independence between $\Phi$ and $(\theta, X)$ and the independent increments of $\Phi$. Since $V$ is a deterministic process, we obtain that $\mathcal{D} = \mathcal{F}^{X,Y}_0$ is independent of $\mathcal{G} = \mathcal{F}^X_0$ and $\mathcal{H} = \mathcal{F}_0^{X,Y}$. Then, Lemma A.1 implies $E^Q[f(\theta(0),X(0))L(0)|\mathcal{F}^{X,Y}_0] = E^Q[f(\theta(0),X(0))L(0)|\mathcal{F}^{X,Y}_0] = \rho(f(0),0)$.

**LEMMA A.2** Suppose two vector stochastic processes $X$ and $Y$ are independent. Let $M(t)$ be a martingale with respect to $\{\mathcal{F}^{X,Y}_t\}$ and let $U$ be $\mathcal{F}^{X,Y}_t$-predictable where $V$ is a deterministic process. If $E[\int_0^t U(s)dM(s)] < \infty$. Then $E[\int_0^t U(s)\Phi(dM(s))|\mathcal{F}^{X,Y}_t] = 0$.

Under the reference measure $Q$, $(\theta, X)$ and $\Phi$ are independent and $M_f(t)$ is still a $\mathcal{F}^{X,Y}_t$-martingale. Also, $U(t) = L(t)$ is $\mathcal{F}^{X,Y}_t$-predictable and $V$ is deterministic. Assumption 4.5 guarantees that $E^Q[\int_0^t L(s-)dM_f(s)] < \infty$. Hence, Lemma A.2 implies $E^Q[\int_0^t L(s-)dM_f(s)|\mathcal{F}^{X,Y}_t] = 0$.

**LEMMA A.3** Suppose a vector stochastic process $X$ and a Poisson random measure $Y$ with measure $m \times \mu$ are independent where $m$ is Lebesgue measure on $\mathbb{R}^+$ and $\mu$ is a finite measure on the mark space $Y$. Suppose that $U$ is $\mathcal{F}^{X,Y}_t$-predictable where $V$ is a deterministic process and $E[\int_0^t U(s)\mu(y)dy] < \infty$. Then, $E^Q[\int_0^t U(s-)dM_f(s)|\mathcal{F}^{X,Y}_t] = 0$.

Under $Q$, similarly, the conditions of Lemma A.3 are satisfied. Below, we give more details to show that the moment condition holds. Since $f$ is bounded and $\int_0^t \int_y E^Q[L(s-)]\mu(dy)ds = \mu(Y)< +\infty$ by Assumption 4.4, it suffices to show that $E^Q[\int_0^t \int_y \zeta(s,y)L(s-\mu(dy)ds] < +\infty$. Since $\zeta(t,y) = \lambda(t,dy)/\mu(dy) > 0$,$$

E^Q[\int_0^t \int_y \zeta(s,y)L(s-\mu(dy)ds] \leq E^Q[\int_0^t \int_y \lambda(s,y)L(s)\mu(dy)ds] = E^Q[\int_0^t \lambda(s)L(s)\int_y \mu(dy)X(s)\mu(dy)ds]$$

where the last inequality is by Assumption 4.5. Hence,

$$E^Q[\int_0^t \int_y f(\theta(s-),X(s-))\left[\zeta(s,y) - 1\right]L(s-\Phi(d(s,y))|\mathcal{F}^{X,Y}_t] = \int_0^t \int_y \rho((\zeta(y)-1)f(\theta,s)\Phi(d(s,y))).$$
LEMMA A.4 Suppose two vector stochastic processes $X$ and $Y$ are independent and $Y$ has independent increments. If $U$ is $\mathcal{F}_t^{X,Y,V}$-adapted satisfying $\int_0^t E[U(s)]ds < \infty$, then

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

Lemma A.4 implies

$$E^Q \left[ \int_0^t L(s) \left\{ A_v f(\theta(s),X(s)) - \int_0^t f(\theta(s),X(s)) [\zeta(s,y) - 1]\mu(dy) \right\} ds | \mathcal{F}_t^{X,Y,V} \right]$$

$$= \int_0^t \rho(A_v f(s),s)ds - \int_0^t \rho(\zeta(s,y) - 1)f(s)\mu(dy)ds$$

Summarizing the above, we have the SDE for $\rho(f,t)$, which is Equation (5.12).

A.3. Proof of Theorem 2

To determine the SDE for $\pi(f,t)$, we note that $\pi(f,t) = \frac{\rho(f,t)}{\rho(1,t)}$. Apply Itô’s formula for semimartingale (see Protter (2003), page 78) to $f(X,Y) = \frac{X}{Y}$ with $X = \rho(f,t)$ and $Y = \rho(1,t)$. After some simplifications, we have

$$\pi(f,t) = \pi(0,t) + \int_0^t \pi(A_v f(s),s)ds - \int_0^t \int_0^t \pi(\zeta(s,y),s) \pi(\zeta(y,s),\mu(dy)ds$$

(A.3)

$$+ \int_0^t \int_0^t \pi(s,y) - \pi(f,t)) \Phi(d(s,y)).$$

A last step remains to make the integrand of the last integral predictable. Assume a trade at price $y$ occurs. Then,

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

Hence, Equation (A.3) implies Equation (5.13).

Recall $\zeta(y) = \zeta(t,y) = \lambda(t)r(y|X(t),t)$. When $\lambda(t)$ is $\mathcal{F}_t^{X,Y,V}$-measurable, $\lambda(t)$ in $\lambda(t)$ can be pulled out of the conditional expectation of $\pi$. Two additional observations further simplify Equation (5.13) to Equation (5.14). First, $\int_0^t \pi(\zeta(y),t)\mu(dy) = \lambda(t)\pi(f,t)$, and $\int_0^t \pi(\zeta(y),t)\mu(dy) = \lambda(t)\pi(f,t)$. Then, the two terms in the integrand of $ds$ in Equation (5.13) cancel out. Second, $\frac{\pi(f,y,s)}{\pi(y,s)} = \frac{\pi(f,y,s)}{\pi(y,s)}$.

A.4. Proof of Theorem 3

We will show that $q_{jk}(f_k,t)$ satisfies Equation (5.15), and when $q_{jk}^{(k)}(\theta^{(k)}(t),\vec{X}^{(k)}(t),t) = a_j(t)$ Equation (5.15) reduces to Equation (5.17). Then, by symmetry, $q_{jk}(f_j,t)$ satisfies Equation (5.16) and, in the special case, (5.18). Recall that $\rho_j(f_j,t)$ satisfies Equation (5.12). Then applying Itô’s formula for semi-martingales (Protter (2003)) and simplifying gives us

$$\frac{\rho_k(f_k,t)}{\rho_j(1,t)} = \frac{\rho_k(f_k,0)}{\rho_j(1,0)} + \int_0^t \left[ \frac{\rho_k(A_v f(s),s)}{\rho_j(1,s)} - \frac{\rho_k(\zeta(s,y)f_k,s)}{\rho_j(1,s)} + \frac{\rho_k(f_k,s)}{\rho_j(1,s)} \right] ds$$

(A.4)

$$\int_0^t \int_0^t \left[ \frac{\rho_k(f_k,s)}{\rho_j(1,1)} - \frac{\rho_k(\zeta(s,y)f_k,s)}{\rho_j(1,s)} \right] \Phi(d(s,y)).$$

We make two observations: The first is that

$$\frac{\rho_k(f_k,s)^2}{\rho_j(1,s)} = \frac{\rho_k(f_k,s)}{\rho_j(1,s)} = \frac{\rho_k(\zeta(s,y)f_k,s)}{\rho_j(1,s)} = \frac{\rho_k(\zeta(s,y)f_k,s)}{\rho_j(1,s)} = \frac{q_{jk}(\zeta(s,y)f_k,s)}{q_{jk}(1,s)}.$$
A.5. Proof of the Uniqueness for Theorems 1, 2 and 3

We follow Kliemann et al. (1990). They formulate the problem as a filtered martingale problem (FMP) proposed by Kurtz & Ocone (1988) and then apply Kurtz and Ocone’s result about the unique solution for the FMP. For our case, Assumption 4.5 ensures the existence of a measure, under which the $\Phi$ is an MPP with compensator by Kurtz & Ocone (1988) and then apply Kurtz and Ocone’s result about the unique solution for the FMP. For (B.2)

\[ p(y|\theta, x, \xi) = \begin{cases} 2 \text{ if the fractional part of } y \text{ is odd } 1/64 \\ 1 \text{ if the fractional part of } y \text{ is odd } 1/32 \\ 0 \text{ if the fractional part of } y \text{ is odd } 1/16 \text{ or coarser} \end{cases} \]

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 $p(y|x) = \sum_{y'} p(y|y') p(y'|x) x \in [0, 1]$. Suppose $D = 64|y - R[x, \frac{1}{64}]|$. Then, $p(y|x)$ can be calculated as, for example, when $r(y) = 2$,

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


Conley, T., Hansen, L. P., Luttmer, E. & Scheinkman, J. (1997), ‘Short-term interest rates as subordinated diffu-


