SIAM/ASA J. Uncertainty Quantification
Vol. xx, pp. x  x–x

# Bayesian Inference via Filtering Equations for Ultra-High Frequency Data (II): Model Selection

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

**Abstract.** For the general partially-observed framework of Markov processes with marked point process observations proposed in [6], we develop the corresponding Bayesian model selection via filtering equations to quantify model uncertainty. To achieve this, we first derive the unnormalized filtering equation and the system of ratio filtering equations to, respectively, characterize the evolution of the marginal likelihood and the corresponding Bayes factors. Then, we prove a powerful weak convergence theorem. The theorem enables us to employ Markov chain approximation method to construct consistent, easily-parallelizable, recursive algorithms to calculate the related Bayes factors and posterior model probabilities of the candidate models in real time for streaming ultra-high frequency data. The general model selection theory is again illustrated by the four specific models built for U.S. Treasury Notes transactions data from GovPX via simulation and empirical studies.

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

**AMS subject classifications.** Primary: 60H35, 62F15, 62M02, 62P05, 93E11;
Secondary: 60F05, 60G55, 65C40, 65C60.

## 1. Introduction.

In our previous work [6], we propose a general nonlinear filtering framework with marked point process observations incorporating other observable economic variables for ultra-high frequency (UHF) data. The model well fits the stylized facts of UHF data in both macro- and micro-movements, unifies important existing models and provides extensions in new directions. We develop Bayesian inference via filtering equations to quantify uncertainty in the previous and the current papers. In the previous one, we develop the related Bayes estimation via filtering equation for the parameter uncertainty quantification. In the current paper, we extend [21, 9] to further develop the related Bayesian model selection via filtering equations for the model uncertainty quantification.

The previously proposed partially-observed model has two equivalent representations. Grounded on the representation of filtering with marked point process observations, we define the corresponding fundamental characteristics for model uncertainty quantification. Namely, we define the joint likelihood, the (integrated) marginal likelihood, the likelihood ratios, the Bayes factors for two models. Moreover, we define the posterior model probabilities for a collection of the proposed models.

The integrated likelihood and the Bayes factors may be characterized by conditional

*School of Economics and Finance, University of Hong Kong, Hong Kong (gracexhu@hku.hk).
†Department of Finance, Henry W. Bloch School of Management, University of Missouri at Kansas City, Kansas City, MO 64110 (kuipersd@umkc.edu).
‡Department of Mathematics and Statistics, University of Missouri at Kansas City, Kansas City, MO 64110 (zengy@umkc.edu). The author's work was supported in part by National Science Foundation under DMS–0604722, TG-DMS100019, and DMS-1228244, and by University Missouri Research Board.

measures that change over time, namely, by *measure-valued processes*. To characterize them, we use stochastic nonlinear filtering technique and derive the necessary filtering equations, which are *stochastic partial differential equations* (SPDEs). The system of ratio filtering equations stipulates the evolution of a pair of ratio conditional measures. Their totals are a pair of Bayes factors, which are at the center of quantifying the model uncertainty for Bayesian viewpoint.

Berger and Pericchi in [1] offers compelling reasons to use the Bayesian approach for model selection. Most of the reasons are suitable for the proposed models and cited below. First, Bayes factors and posterior model probabilities are easy to understand, and the approach is conceptually the same, regardless of the number of models under consideration. Second, Bayesian model selection is consistent. Namely, 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. 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. Fourth, the Bayesian procedures are automatic Ockham's razors, preferring simpler models over more sophisticated models when the data provide roughly equivalent fits. Fifth, the Bayesian approach does not require nested models. Finally, the Bayesian approach can account for model uncertainty via Bayesian averaging. All these desirable properties support our adoption of Bayesian approach for model uncertainty quantification.

Similar to the case of parameter uncertainty quantification, the evolving system of measure-valued processes is of infinite dimension and there is tremendous computational challenge in application. Fortunately, the system of ratio filters also has recursiveness, which remarkably benefits computation if the recursiveness is appropriately exploited. Then, we prove a weak convergence theorem, which shows that the weak convergence of the approximate signal guarantees the consistency of all of the useful approximate conditional measures. Based on the theorem, we employ the Markov chain approximation method to construct consistent, easily-parallelizable, efficient and recursive algorithms. The algorithms can quantify model uncertainty in real time for inflowing UHF data.

Finally, we exemplify the general Bayesian model selection theory again by the four specific models constructed for U.S. Treasury Notes transaction data from GovPX. We illustrate how to develop a recursive algorithm for efficiently propagating and updating the pair ratio conditional measures, whose total are a pair of Bayes factors. Simulation studies validate that the algorithm rightly selects the true model. Then, we employ the algorithm for model uncertainty quantification of the four models. The model selection empirical results further confirm that both information-based and inventory management based motives have significant impact on trade-to-trade price volatility.

The paper is organized as follows. The next section briefly review the previously proposed model. Section 3 develops Bayesian model selection via filtering equations. We define the related statistical fundamental characteristics, derive the filtering equations and prove a weak convergence theorem. Section 4 supplies simulation and model selection examples drawn from the microstructure of trading in U.S. Treasury notes. We provide concluding remarks in Section 5, with the mathematical proofs collected into Appendices.

**2. Brief Review of the Model.** For self-containing purpose, we concisely recap the model in two equivalent representations and interested readers can refer to [6] for the generality of the model and more detailed explanations of the assumptions.

**2.1. Representation I: Random Arrival Time State-Space Model.** This representation similarly has the state process, observation times, observations and noise.

**State Process:**

We make a a mild Markov assumption for the extended state process $(\theta, X)$, where $\theta$ is allowed to potentially change in continuous time. Let $V$ be a vector process, representing other observable economic or market factors.

Assumption 2.1. *$(\theta, X)$ is a $p+m$-dimension vector Markov process that is the solution of a martingale problem for a generator $\mathbf{A}_v$ such that $M_f(t) = f(\theta(t), X(t)) - \int_0^t \mathbf{A}_v f(\theta(s), X(s))ds$ is a $\mathcal{F}_t^{\theta,X,V}$-martingale, where $\mathcal{F}_t^{\theta,X,V}$ is the $\sigma$-algebra generated by $(\theta(s), X(s), V(s))_{0\leq s\leq t}$, and $f$ is in the domain of $\mathbf{A}_v$.*

**Observation times**

We assume that the observations, $T_1, T_2, \ldots, T_i, \ldots$, follow a general point process with a nonnegative $\mathcal{F}_t$-predictable stochastic intensity in the following form: $\bar{\lambda}(t) = \bar{\lambda}(\theta(t), X(t), V^{t-}, \Phi^{t-}, t)$ where $V^t = V(\cdot \wedge t)$ denotes the sample path of $V$ up to time $t$ and similarly $\Phi^{t-} = \{(T_i, Y_i) : T_i < t\}$.

**Observations and Noise**

For the mark space $\mathbb{Y}$, recall that $(\mathbb{Y}, \mathcal{Y}, \mu)$ is a measure space with a finite measure $\mu$ $(\mu(\mathbb{Y}) < \infty)$ and $(\mathbb{Y}, d_{\mathbb{Y}})$ is a complete, separable metric space.

The noisy observation at event time $T_i$, $Y(T_i)$, takes a value in mark space $\mathbb{Y}$ and is modeled by $Y(T_i) = F(X(T_i))$. Note that $F(\cdot)$ in $y = F(x)$ is a random transformation from $x$ to $y$, specified by a transition probability $p(y|x)$ with mild conditions given in Assumption 2.2.

This general representation has two significant differences from the usual *state-space models*: random arrival times and continuous-time state process.

**2.2. Representation II: Filtering with MPP Observations.** In Representation II, we have the extended Markov signal, $(\theta, X)$, which is partially observed through a MPP, $\Phi = \{T_i, Y_i\}_{i\geq 1}$. We incorporate an auxiliary predictable process $V$ in both the signal and the MPP, and formulate $(\theta, X, \Phi, V)$ an extended filtering problem with MPP observations. Below we recall the four more assumptions for the model.

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

$$\begin{aligned} &\lambda(t, dy; \theta(t), X(t), V^{t-}, Y^{t-}, t-) \\ =&\bar{\lambda}(\theta(t), X(t), V^{t-}, Y^{t-}, t-)p(dy|X(t); \theta(t), V^{t-}, Y^{t-}, t-). \end{aligned} \tag{2.1}$$

Assumption 2.3. *There exists a* reference measure $Q$ *such that (1) $P$ is absolutely continuous with respect to $Q$, namely, $P << Q$; and (2) under $Q$, $(\theta, X)$ and $V$ are independent of $\Phi = \{(T_n, Y_n)\}_{n\geq 1}$ and the compensator of MPP $\Phi$ is $\gamma_Q(d(t, y)) = \mu(dy)dt$.*

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

$$r(y) = r(y;\theta(t),X(t),V^{t-},\Phi^{t-},t-) = \frac{p(dy|X(t);\theta(t),V^{t-},\Phi^{t-},t-)}{\mu(dy)}. \tag{2.2}$$

and

$$\begin{aligned}\zeta(t,y) = &\frac{\gamma_P(d(t,y))}{\gamma_Q(d(t,y))} = \frac{\lambda(t,dy)}{\mu(dy)}\\ =&\bar{\lambda}(\theta(t),X(t),V^{t-},\Phi^{t-},t-)r(y;\theta(t),X(t),V^{t-},\Phi^{t-},t-).\end{aligned} \tag{2.3}$$

Let $L(t) = \frac{d\mathbb{P}}{dQ}(t)$ be the Radon-Nikodym derivative given in (3.1). Clearly, $\zeta(t,y)$ determines $L(t)$.

Assumption 2.4. *The $\zeta(t,y)$ defined in (2.3) satisfies the condition that $E^Q[L(T)] = 1$ for all $T > 0$.*

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

**3. Bayesian Model Selection via Filtering Equations.** The filtering formulation furnishes a natural setup for Bayesian inference. In this section, we present the unnormalized filtering equation, a system of ratio filtering equations, and a weak convergence theorem with the proofs given in Appendices.

We first define the related statistical foundations for Bayesian model selection.

**3.1. The Likelihoods, the Bayes Factors and the Posterior Model Probabilities.** For conducting hypotheses testing and model selection, we study the joint likelihood, the marginal (or integrated) likelihood processes of the proposed model, as well as the processes of the likelihood ratios, the Bayes factors, and the posterior model probabilities.

**3.1.1. The Joint Likelihood Process.** Let $D_U[0,\infty)$ be the space of right continuous with left limit functions with state space $U$. Then, the probability measure $P$ given $V$ on $\Omega = D_{\mathbb{R}^p\times\mathbb{R}^m\times\mathbb{Y}}[0,\infty)$ for $(\theta,X,\Phi)$ can be written as $P = P_{\theta,x,\Phi|v} = P_{\theta,x|v}\times P_{\Phi|\theta,x,v}$, where $P_{\theta,x|v}$ is the probability measure on $D_{\mathbb{R}^{p+m}}[0,\infty)$ for $(\theta,X)$ given $V$ such that $M_f(t)$ in Assumption 2.1 is a $\mathcal{F}_t^{\theta,X,V}$-martingale, and $P_{\Phi|\theta,x,v}$ is the conditional probability measure on $D_{\mathbb{Y}}[0,\infty)$ for $\Phi$ given $(\theta,X,V)$.

Under $P$, $\Phi$ relies on $(\theta,X,V)$. With Assumption 2.3, there exists a reference measure $Q$ such that under $Q$, $(\theta,X)$ and $\Phi$ become independent, $(\theta,X)$ remains the same probability law and $\Phi$ becomes a Poisson random measure with compensator $\mu(dy)dt$. Therefore, $Q$ can be decomposed as $Q = P_{\theta,x|v}\times Q_\Phi$, where $Q_\Phi$ is the probability measure for the Poisson random measure on $D_{\mathbb{Y}}[0,\infty)$ with compensator $\mu(dy)dt$.

One can obtain the Radon-Nikodym derivative guaranteed by Assumption 2.4, (see Proposition 14.4.I of [3], or Theorem 10.1.3. of [16]) and denoted by $L(t)$ as given below:

$$\begin{aligned}L(t) = &\frac{dP}{dQ}\Big|_{\mathcal{F}_t} = \frac{dP_{\theta,x|v}}{dP_{\theta,x|v}}\Big|_{\mathcal{F}_t} \times \frac{dP_{\Phi|\theta,x,v}}{dQ_\Phi}\Big|_{\mathcal{F}_t} = \frac{dP_{\Phi|\theta,x,v}}{dQ_\Phi}\Big|_{\mathcal{F}_t}\\ =&L(0)\exp\left\{\int_0^t\int_{\mathbb{Y}}\log\zeta(s-,y)\Phi(d(s,y)) - \int_0^t\int_{\mathbb{Y}}\Big[\zeta(s-,y)-1\Big]\mu(dy)ds\right\},\end{aligned} \tag{3.1}$$

where $\zeta(t,y)$ is given by (2.3) 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:

$$L(t) = L(0) + \int_0^t \int_{\mathbb{Y}} (\zeta(s,y) - 1)L(s-)(\Phi(d(s,y)) - \mu(dy)ds). \tag{3.2}$$

Because of Assumption 2.4, $L(t)$ is actually a martingale. The SDE form plays an important role in deriving the unnormalized filtering equation.

From a frequentist's view point, one takes $L(0) = 1$ and $L(t)$ is the joint likelihood of the model. In this paper, we take a Bayesian stand and regard $L(0)$ as a random variable with a prior distribution on $(\theta(0), X(0))$ satisfying $E^Q(L(0)) = 1$.

#### 3.1.2. The Unnormalized Conditional Measure-valued and Marginal Likelihood Processes.

In the case that $X$ is not observable and for the purpose of statistical analysis, we need the marginal likelihood of $\Phi$, which can be concisely expressed by conditional expectation. For the purpose of characterizing the evolution of the marginal likelihoods, we define the two terms below.

Definition 3.1. *Let $\rho_t$ be the unnormalized conditional measure of $(\theta(t), X(t))$ given $\mathcal{F}_t^{\Phi,V}$ defined as*

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

*where $A$ is a Borel set in $\mathbb{R}^{p+m}$.*

Following the nonlinear filtering literature ([2]), $\rho_t$ is called *the unnormalized conditional measure-value process*.

Definition 3.2. *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 $\Phi$ with $V$ is obtained by taking $f = 1$: $E^Q[L(t)\,|\,\mathcal{F}_t^{\Phi,V}] = \rho(1,t)$. A frequentist uses the marginal likelihood for statistical inference.

If a prior distribution is assumed on $(\theta(0), X(0))$ as in Bayesian paradigm, then the integrated marginal likelihood of $\Phi$ over the prior is also $\rho(1,t)$. If we have prior information, then Bayesian method offers a scientific way to integrate the prior information. If we have no prior information, then we can simply use uniform noninformative priors.

In order to compare $K$ models with the same observation $\Phi = \{T_i, Y_i\}_{i\geq 1}$ and the same observable factor $V$, we prepare the notations here. Denote Model $k$ by $(\theta^{(k)}, X^{(k)}, \Phi^{(k)}, V^{(k)})$ for $k = 1, 2, ..., K$ with $P^{(k)} = P^{(k)}_{\theta,x|v} \times \mathbb{P}^{(k)}_{\Phi|\theta,x,v}$. Denote $Q^{(k)} = P^{(k)}_{\theta,x|v} \times Q_\Phi$ where $Q_\Phi$ is the same for all $K$ models. Under $Q$, the compensator $\Phi$ is $\mu(dy)dt$. Denote $L^{(k)}(t)$ as the joint likelihood of $(\theta^{(k)}, X^{(k)}, \Phi^{(k)}, V^{(k)})$ with respect to $Q^{(k)}$. Since the available information, $\mathcal{F}_t^{\Phi,V}$, is the same for all models, we denote

$$\rho_k(f_k, t) = E^Q[f_k(\theta^{(k)}(t), X^{(k)}(t))L^{(k)}(t)|\mathcal{F}_t^{\Phi,V}]. \tag{3.3}$$

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

**3.1.3. The Ratio Conditional Measure-valued and Bayes Factor Processes.** For the general partially-observed model reviewed in Section 2, we define the corresponding Bayes factors, which are at the heart of Bayesian hypothesis testing and model selection between two given models.

Definition 3.3.*For $k = 1, 2, ..., K$ and $j \neq k$ in the same range, the Bayes factor of Model $k$ over Model $j$, $B_{kj}$, is defined as the ratio of integrated likelihoods of Model $k$ over Model $j$:*

$$B_{kj}(t) = \rho_k(1,t)/\rho_j(1,t). \tag{3.4}$$

The Bayes factor is usually interpreted 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 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 a table furnished by [7] as guideline. For example, if $B_{kj} > 20$ or 150, then there is strong or decisive evidence against Model $j$, and in favor of Model $k$.

Similar to defining $\rho_t$ and $\rho(f,t)$, in order to characterize the evolution of and to compute the Bayes factors, we define a conditional measure $q_t^{(kj)}$ and a conditional integral characteristic $q_{kj}(f_k, t)$ below.

First, we provide the intuition behind. Recall from [6] that $\pi_t$ is the conditional distribution of $(\theta(t), X(t))$ given $\mathcal{F}_t^{\Phi,V}$ and $\pi(f,t) = E^P[f(\theta(t), X(t))|\mathcal{F}_t^{\Phi,V}]$. Also recall the Bayes theorem: $\pi(f,t) = \rho(f,t)/\rho(1,t)$, implying that the conditional probability measure $\pi_t$ is obtained from normalizing the conditional measure $\rho_t$ by its (integrated) marginal likelihood in the case of Bayes estimation. Now, for the case of Bayesian model selection between two models, instead, we divide the conditional measure $\rho_t^{(k)}$ of Model $k$ by the integrated likelihood of Model $j$ so as to obtain another useful ratio conditional measure (but not a probability measure) and a related integral.

Definition 3.4. *For $k = 1, 2, ..., K$ and $j \neq k$ in the same range, let $q_t^{(kj)}$ be the ratio conditional measure of $(\theta^{(k)}(t), X^{(k)}(t))$ in Model $k$ given $\mathcal{F}_t^{\Phi,V}$ with respect to Model $j$, defined by: for a Borel set $A$ in $\mathbb{R}^{p^{(k)}+m^{(k)}}$,*

$$q_t^{(kj)}\left\{(\theta^{(k)}(t), X^{(k)}(t)) \in A\right\} = \frac{\rho_t^{(k)}\left\{(\theta^{(k)}(t), X^{(k)}(t)) \in A\right\}}{\rho_j(1,t)}$$
$$= \frac{E^Q\left[\mathbf{I}_{\{(\theta^{(k)}(t), X^{(k)}(t)) \in A\}}(\theta^{(k)}(t), X^{(k)}(t)) L^{(k)}(t) | \mathcal{F}_t^{\Phi,V}\right]}{\rho_j(1,t)}$$
$$= E^Q\left[\mathbf{I}_{\{(\theta^{(k)}(t), X^{(k)}(t)) \in A\}}(\theta^{(k)}(t), X^{(k)}(t)) \tilde{L}^{(k)}(t) | \mathcal{F}_t^{\Phi,V}\right]$$

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

In the last equality, we can move $\rho_j(1,t)$ inside the conditional expectation because $\rho_j(1,t)$ is $\mathcal{F}_t^{\Phi,V}$ measurable. Similarly, $q_t^{(kj)}$ is a measure-valued process.

Definition 3.5. *Let*

$$q_{kj}(f_k,t) = \int f_k(\theta^{(k)},x^{(k)}) q_t^{(kj)}(d(\theta^{(k)},x^{(k)})). \tag{3.5}$$

Similar to the integral form of $\pi(f,t)$, the integral forms of $\rho(f,t)$ 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.

By the definition of $q_t^{(kj)}$, we observe that

$$q_{kj}(f_k,t) = \int f_k(\theta^{(k)},x^{(k)}) \frac{\rho_t^{(k)}(d(\theta^{(k)},x^{(k)}))}{\rho_j(1,t)} = \frac{\rho_k(f_k,t)}{\rho_j(1,t)}.$$

Similarly, $q_{jk}(f_j,t) = \frac{\rho_j(f_j,t)}{\rho_k(1,t)}$. The pair, $(q_{kj}(f_k,t), q_{jk}(f_j,t))$, are called as the *filter ratio processes* for $f_k$ of Model $k$ and $f_j$ of Model $j$.

Clearly, the connection of $q_{kj}(f_k,t)$ to the Bayes factor is shown by taking $f_k = 1$ and $B_{kj}(t) = q_{kj}(1,t)$.

**3.1.4. The Posterior Model Probability Processes.** For the model selection with more than two models, the posterior model probabilities is a powerful tool to quantify model uncertainty. When prior model probabilities $P(M_k|\mathcal{F}_0^{\Phi,V})$, $k = 1,2,...,K$, are available, then we can calculate, by Bayes formula, the posterior probabilities of the models, $P(M_k|\mathcal{F}_t^{\Phi,V})$, which can also be expressed by Bayes factors as shown below:

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

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

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

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.

**3.2. Filtering Equations.** Stochastic partial differential equations (SPDEs) provide an powerful machinery to stipulate the evolution of the continuous-time conditional measure-valued processes, determining the likelihoods, the Bayes factors and the posterior model probabilities. The following two theorems summarize the useful SPDEs.

Before presenting the theorems, let us recall that we can alternatively present an MPP as a random counting measure, $\Phi(t,B)$, recording the cumulative number of marks in a set $B \in \mathcal{Y}$ up to time $t$ for all $B \in \mathcal{Y}$ and $t > 0$. Observe $\Phi(t,B) = \int_0^t \int_B \Phi(dt,dy)$ where $\Phi(dt,dy)$ is given by

$$\Phi(dt,dy) = \sum_{i\geq 1} \delta_{\{T_i,Y_i\}}(t,y)dtdy \tag{3.6}$$

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

Theorem 3.1. *Suppose that $(\theta,X,\Phi,V)$ satisfies Assumptions 2.1 - 2.5. Then, $\rho_t$ is the unique measure-valued solution of the SPDE under $Q$, the unnormalized filtering equation,*

$$\rho(f,t) = \rho(f,0) + \int_0^t \rho(\mathbf{A}_v f,s)ds + \int_0^t \int_{\mathbb{Y}} \rho((\zeta(y)-1)f,s-)(\Phi(d(s,y)) - \mu(dy)ds), \tag{3.7}$$

*for $t > 0$ and $f \in D(\mathbf{A}_v)$, the domain of generator $\mathbf{A}_v$, where $\zeta(y) = \zeta(s-,y)$ defined in (2.3).*

The unnormalized filtering equation characterizes the evolution of the unnormalized conditional measure, whose total measure is the (integrated) marginal likelihood. Under $Q$, the compensator of $\Phi$ is $\mu(dy)dt$ and the double integral is a martingale. This implies that the unnormalized filter has a semimartingale representation.

Theorem 3.2. *Suppose Model $k$ ($k = 1,2,...,K$) has generator $\mathbf{A}_v^{(k)}$ for $(\theta^{(k)},X^{(k)})$, the trading intensity $\bar\lambda_k(t) = \bar\lambda_k(\theta^{(k)}(t),X^{(k)}(t),\Phi^{t-},V^{t-})$, 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 2.1 - 2.5 with a common reference measure $Q_\Phi$, under which the compensator of $\Phi$ is $\mu(dy)dt$. Then, for $k = 1,2,...,K$ and $j \neq k$ in the same range, $\big(q_t^{(kj)}, q_t^{(jk)}\big)$ are the unique pair measure-valued solution of the following system of SPDEs, the ratio filtering equations,*

(3.8)

$$\begin{aligned} &q_{kj}(f_k,t) = q_{kj}(f_k,0) + \int_0^t q_{kj}(\mathbf{A}_v^{(k)} f_k,s)ds \\ &+ \int_0^t\int_{\mathbb{Y}} \Big[\frac{q_{kj}(f_k\zeta_k(y),s-)}{q_{jk}(\zeta_j(y),s-)} q_{jk}(1,s-) - q_{kj}(f_k,s-)\Big]\Big(\Phi(d(s,y)) - \frac{q_{jk}(\zeta_j(y),s)}{q_{jk}(1,s)}\mu(dy)ds\Big), \end{aligned}$$

(3.9)

$$\begin{aligned} &q_{jk}(f_j,t) = q_{jk}(f_j,0) + \int_0^t q_{jk}(\mathbf{A}_v^{(j)} f_j,s)ds \\ &+ \int_0^t\int_{\mathbb{Y}} \Big[\frac{q_{jk}(f_j\zeta_j(y),s-)}{q_{kj}(\zeta_k(y),s-)} q_{kj}(1,s-) - q_{jk}(f_j,s-)\Big]\Big(\Phi(d(s,y)) - \frac{q_{kj}(\zeta_k(y),s)}{q_{kj}(1,s)}\mu(dy)ds\Big), \end{aligned}$$

*for all $t > 0$, $f_k \in D(\mathbf{A}_v^{(k)})$ and $f_j \in D(\mathbf{A}_v^{(j)})$.*

*When $\bar\lambda_k(\theta^{(k)}(t),X^{(k)}(t),\Phi^{t-},V^{t-},t) = \bar\lambda_j(\theta^{(j)}(t),X^{(j)}(t),\Phi^{t-},V^{t-},t) = \bar\lambda(\Phi^{t-},V^{t-},t)$*

*and letting* $r_k(y) = p^{(k)}(dy|x)/\mu(dy)$*, we can simplify the above two equations as*

$$\begin{aligned} q_{kj}(f_k,t) &= q_{kj}(f_k,0) + \int_0^t q_{kj}(\mathbf{A}_v^{(k)} f_k, s)ds \\ &+ \int_0^t \int_{\mathbb{Y}} \Big[\frac{q_{kj}(f_k r_k(y), s-)}{q_{jk}(r_j(y), s-)} q_{jk}(1, s-) - q_{kj}(f_k, s-)\Big] \Phi(d(s,y)), \end{aligned} \tag{3.10}$$

$$\begin{aligned} q_{jk}(f_j,t) &= q_{jk}(f_j,0) + \int_0^t q_{jk}(\mathbf{A}_v^{(j)} f_j, s)ds \\ &+ \int_0^t \int_{\mathbb{Y}} \Big[\frac{q_{jk}(f_j r_j(y), s-)}{q_{kj}(r_k(y), s-)} q_{kj}(1, s-) - q_{jk}(f_j, s-)\Big] \Phi(d(s,y)). \end{aligned} \tag{3.11}$$

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 $\frac{q_{jk}(\zeta_j(y),s)}{q_{jk}(1,s)}\mu(dy)ds = \pi(\zeta_j(y),s)\mu(dy)ds$, which is the compensator of $\Phi$ under $P$ for Model $j$. This implies that the ratio filtering equations of the models also have semimartingale representations under suitable measures. Moreover, the filtering equations have the recursiveness, which make it possible for the real time model uncertainty quantification for inflowing UHF data.

Similar to the normalized filtering equation in the case of exogenous sampling, we observe that if event times $\{T_i\}$'s stochastic intensities $\bar\lambda(t)$ or $(\bar\lambda_k(t), \bar\lambda_j(t))$ are $\mathcal{F}_t^{\Phi,V}$-predictable including ACD models or Cox models, then Equations (3.8) and (3.9) are simplified to (3.10) and (3.11). The advantage is the great reduction of computation in calculating the Bayes factors and the posterior model probabilities. Moreover, observe that the stochastic intensities disappear in the simplified ratio filters. This implies the Bayesian model selection via filtering equations is *model-free* of the assumptions on durations as long as its stochastic intensities are $\mathcal{F}_t^{\Phi,V}$-predictable. This advantage is exploited in the example given in Section 4. The disadvantage is the exclusion of the relationship between durations and $(X(t), v(t))$ where $v(t)$ is the time-vary part of $\theta$ such as stochastic volatility and regime shifting.

To prepare for the construction of recursive algorithms, we point out that the ratio filtering equations, such as (3.10), can be separated into the propagation equation for the evolution without trade,

$$q_{kj}(f_k, t_{i+1}-) = q_{kj}(f_k, t_i) + \int_{t_i}^{t_{i+1}-} q_{kj}(\mathbf{A}_v^{(k)} f_k, s)ds, \tag{3.12}$$

and the updating equation,

$$q_{kj}(f_k, t_{i+1}) = \frac{q_{kj}(f_k r_k(y_{i+1}), t_{i+1}-)}{q_{jk}(r_j(y_{i+1}), t_{i+1}-)} q_{jk}(1, t_{i+1}-), \tag{3.13}$$

when a trade happens at time $t_{i+1}$ with price $y_{i+1}$.

**3.3. A Weak Convergence Theorem.** Theorems 3.1 and 3.2 deliver the filtering equations of the model-selection-related continuous-time conditional measure-valued processes, which are defined on $D_{\mathbb{R}^{p+m}\times\mathbb{Y}}[0,\infty)$ and thus are all infinite dimensional. To compute them, the infinite dimensional problem has to be reduced to a finite dimensional problem so as to construct algorithms. One fundamental requirement for the recursive algorithms is consistency: the approximate conditional measure-valued processes computed by the algorithms converge to the true ones. The theorem below furnishes the foundation for the consistency.

**3.3.1. The Setup.** Since $(\mathbb{Y}, d_{\mathbb{Y}})$ is a complete separable metric space, so is $D_{\mathbb{R}^{p+m}\times\mathbb{Y}}[0,\infty)$ embedded with the Skorohod topology (see Theorem 3.5.6 of [4]), where we consider the family of Borel probability measures, topologized with the Prohorov metric. Under such a setup is the weak convergence theorem that we are going to establish. Chapter 3 of [4] provides a formal description of all these concepts and the weak convergence employed here.

Following the literature, we use the notation, $X_\epsilon \Rightarrow X$, to mean $X_\epsilon$ converges weakly to $X$ in the Skorohod topology as $\epsilon \to 0$. Such weak convergence is uniform in time in the sense that $X_\epsilon \Rightarrow X$ implies $(X_\epsilon(t_1), X_\epsilon(t_2), ..., X_\epsilon(t_k)) \Rightarrow (X(t_1), X(t_2), ..., X(t_k))$ for every finite set $\{t_1, t_2, ..., t_k\} \subset D(X)$, where $D(X) \equiv \{t \geq 0 : P\{X(t) = X(t-)\} = 1\}$ is the continuity set of $X$ (Theorem 3.7.8 of [4]).

For Model $k$ $(\theta^{(k)}, X^{(k)}, \Phi^{(k)}, V^{(k)})$ where $k = 1, 2, ..., K$, let $(\theta_\epsilon^{(k)}, X_\epsilon^{(k)})$ be an approximation of the signal, $(\theta^{(k)}, X^{(k)})$. Since $(\theta^{(k)}, X^{(k)})$ is a vector of stochastic process with sample paths in $D_{\mathbb{R}^{p+m}}[0,\infty)$, $(\theta_\epsilon^{(k)}, X_\epsilon^{(k)})$ 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 $\Phi_\epsilon^{(k)} = \{(T_{i,\epsilon}^{(k)}, Y_{i,\epsilon}^{(k)})\}$. Then, an approximate model for Model $k$ is $(\theta_\epsilon^{(k)}, X_\epsilon^{(k)}, \Phi_\epsilon^{(k)}, V_\epsilon^{(k)})$ with $P_\epsilon^{(k)} = P_{\theta,x|v}^{(k),\epsilon} \times \mathbb{P}_{\Phi|\theta,x,v}^{(k),\epsilon}$. The corresponding reference measure is $Q_\epsilon^{(k)} = P_{\theta,x|v}^{(k),\epsilon} \times Q_\Phi$ where $Q_\Phi$ is common for all $k$ and all $\epsilon$. Then, the compensator of $\Phi_\epsilon^{(k)} = \{(T_{i,\epsilon}^{(k)}, Y_{i,\epsilon}^{(k)})\}$ is $\mu(dy)dt$ under $Q_\Phi$, and is $\lambda_\epsilon^{(k)}(t, dy)dt = \zeta_\epsilon^{(k)}(t, y)\mu(dy) = \bar{\lambda}_\epsilon^{(k)}(t)\, p_\epsilon^{(k)}(dy|X_\epsilon^{(k)}(t), t)dt$ under $P_\epsilon^{(k)}$. Note that the filtration $\mathcal{F}_t^{\Phi,V}$ is the same available information for all models of $k$ and $\epsilon$. Let $L_\epsilon^{(k)}(t) = \frac{dP_\epsilon^{(k)}}{dQ_\epsilon^{(k)}}(t) = L\Big(\big(\theta_\epsilon^{(k)}(s), X_\epsilon^{(k)}(s), \Phi_\epsilon^{(k)}(s, B), V_\epsilon^{(k)}(s)\big) : 0 \leq s \leq t, B \in \mathcal{Y}\Big)$ as in Equation (3.1). Suppose that for $\epsilon > 0$, $(\theta_\epsilon^{(k)}, X_\epsilon^{(k)}, \Phi_\epsilon^{(k)}, V_\epsilon^{(k)})$ lives on $(\Omega_\epsilon^{(k)}, \mathcal{F}_\epsilon^{(k)}, P_\epsilon^{(k)})$ with Assumptions 2.1 - 2.5.

**3.3.2. The Convergence Theorem.** We first define the approximations of $\rho_k(f_k, t)$, $\pi_k(f_k, t)$, and $q_{kj}(f_k, t)$. Recall for Model $k$,

$$\pi_k(f_k, t) \equiv E^{P^{(k)}}\Big[f_k\big(\theta^{(k)}(t), X^{(k)}(t)\big)|\mathcal{F}_t^{\Phi,V}\Big].$$

Definition 3.6. *For $k = 1, 2, ..., K$, $j \neq k$ in the same range and an approximate model of Model $k$, let $\rho_{\epsilon,t}^{(k)}$ be the unnormalized conditional measure of $(\theta_\epsilon^{(k)}(t), X_\epsilon^{(k)}(t))$ given $\mathcal{F}_t^{\Phi,V}$, and let*

$$\rho_{\epsilon,k}(f_k, t) = E^{Q_\epsilon^{(k)}}\Big[f_k\big(\theta_\epsilon^{(k)}(t), X_\epsilon^{(k)}(t)\big)L_\epsilon^{(k)}(t)|\mathcal{F}_t^{\Phi,V}\Big].$$

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

$$\pi_{\epsilon,k}(f_k,t) = E^{P^{(k)}_\epsilon}\Big[f_k\big(\theta^{(k)}_\epsilon(t), X^{(k)}_\epsilon(t)\big)|\mathcal{F}^{\Phi,V}_t\Big].$$

*Let $q^{(kj)}_{\epsilon,t}$ be the ratio conditional measure of $(\theta^{(k)}_\epsilon(t), X^{(k)}_\epsilon(t))$ given $\mathcal{F}^{\Phi,V}_t$ with respect to Model $j$, and let $q_{\epsilon,kj}(f_k,t) = \rho_{\epsilon,k}(f_k,t)/\rho_{\epsilon,j}(1,t)$.*

*Let $P_\epsilon(M_k|\mathcal{F}^{\Phi,V}_t)$ be the posterior model probability of an approximate Model $k$ given as below:*

$$P_\epsilon(M_k|\mathcal{F}^{\Phi,V}_t) = \frac{P(M_k|\mathcal{F}^{\Phi,V}_0)\rho_{\epsilon,k}(1,t)}{\sum_{l=1}^K P(M_l|\mathcal{F}^{\Phi,V}_0)\rho_{\epsilon,l}(1,t)} = \Big[\sum_{l=1}^K \frac{P(M_l|\mathcal{F}^{\Phi,V}_0)}{P(M_k|\mathcal{F}^{\Phi,V}_0)} q_{\epsilon,lk}(1,t)\Big]^{-1}.$$

Now, we state the main convergence theorem.

Theorem 3.3. *Suppose that for $k = 1,2,...,K$, $(\theta^{(k)}, X^{(k)}, \Phi^{(k)}, V^{(k)})$ satisfies Assumptions 2.1 - 2.5 with the reference measure $Q^{(k)} = P^{(k)}_{\theta,x|v} \times Q_\Phi$, and the compensator of $\Phi^{(k)}$ is $\mu(dy)dt$ under $Q_\Phi$. Suppose that Assumptions 2.1 - 2.5 hold for the approximate models $(\theta^{(k)}_\epsilon, X^{(k)}_\epsilon, \Phi^{(k)}_\epsilon, V^{(k)}_\epsilon)$ with the reference measure $Q^{(k)}_\epsilon = P^{(k),\epsilon}_{\theta,x|v} \times Q_\Phi$, and the compensator of $\Phi^{(k)}_\epsilon$ is also $\mu(dy)dt$ under $Q_\Phi$ for all $k$ and all $\epsilon$.*

*If $(\theta^{(k)}_\epsilon, X^{(k)}_\epsilon) \Rightarrow (\theta^{(k)}, X^{(k)})$ as $\epsilon \to 0$, then, as $\epsilon \to 0$, for all bounded continuous functions, $f_k$ and $f_j$, $k = 1,2,...,K$ and $j \neq k$ in the same range,*
*(1) $\Phi^{(k)}_\epsilon \Rightarrow \Phi^{(k)}$ under physical measures;*
*(2) $\rho_{\epsilon,k}(f_k,t) \Rightarrow \rho_k(f_k,t)$;*
*(3) $\pi_{\epsilon,k}(f_k,t) \Rightarrow \pi_k(f_k,t)$;*
*(4) $q_{\epsilon,kj}(f_k,t) \Rightarrow q_{kj}(f_k,t)$ simultaneously for each pair $(k,j)$ with $k \neq j$;*
*(5) $P_\epsilon(M_k|\mathcal{F}^{\Phi,V}_t) \Rightarrow P(M_k|\mathcal{F}^{\Phi,V}_t)$ simultaneously for $k = 1,2,\cdots,K$.*

This theorem states that as long as the approximate signal weakly converges to the true signal, we have the weak convergence of (1) the observation of an approximate model to that of the true model under physical measures, (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.

Built on this theorem, we can adopt the blueprint described in Section 3.3 of [6] for constructing consistent and easily-parallelized recursive algorithms through Kushner's Markov chain approximation methods ([15]) to compute the approximate Bayes factors and posterior model probabilities for model uncertainty quantification.

**4. Model Selection via Filtering: An Application to Bond Transactions Data.** This section continues to employ the examples given in [6]. We illustrate the method of constructing a recursive algorithm to compute the Bayes factors and the posterior model probabilities. We supply simulation and model selection examples drawn from the microstructure of trading in U.S. Treasury notes, and exemplify how to empirically test market microstructure hypotheses.

First, we concisely review the data and the four partially-observed models for bond transactions data used in [6].

**4.1. A Brief Review of the Data and the Four Models.** The data we study involves quotes and trades for the "on-the-run" (hereafter, "OTR") 10-year U.S. Treasury note. Specifically, the data includes a tick-by-tick record for the OTR 10-year note issued on August 15, 2000 through the end of 2000, during which time the note stayed OTR in the Treasury market.

We only present the model by Representation I. Recall that the state process turns into the intrinsic value process $X_t$. We assume $X_t$ is a geometric Brownian motion (GBM) with the SDE given by:

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

where $B_t$ is a standard Brownian Motion, and $(\mu, \sigma_0, \sigma_{en}, \sigma_{bs})$ are model parameters. The observable factors $V(t) = \{V_{en}(t), V_{bs}(t)\}$ consist of two indicator processes. $V_{en}(t)$ indicates whether trades observed at time $t$ are coincident with significant macroeconomic news announcements and $V_{bs}(t)$ indicates whether a trade is buyer or seller initiated. A benefit of the employed framework is that we can readily discriminate between different competing theories presented in [6] using simple inferential tests. Especially, we investigate the four formats of the process $V(t)$ presented above with different 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.

Similarly, we consider an exogenous sampling case where the trade duration follows an Exponential ACD or Weibull ACD model. Hence, the simplified version of the ratio filtering equation system for the Bayes factors can be employed.

We combine trade noise onto the intrinsic values at observed trading times to generate the model price process. 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 and has a doubly geometric distribution with parameter $\rho$. For consistency with market convention in the OTR 10-year note market, we construct a simple random biasing function $b_i(\cdot)$ for ticks with $M = 64$. Hence, we develop a random transformation $y = F(x)$ by adding non-clustering noise $U_i$, rounding to $1/M$ and biasing to clustering ticks as below:

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

The corresponding transition probability $p(y|x)$ can be calculated accordingly. Details concerning $b_i(\cdot)$ and the explicit $p(y|x)$ for $F$ can be found in [6].

There are three groups of parameters in the models. One group is the clustering noise parameters, which can be estimated through the method of relative frequency. Another group is the trading duration parameters, which can be estimated independently by another approach such maximum likelihood. In this paper, we focus on model selection, involving the third group of parameters $\theta = (\mu, \sigma_0, \sigma_{en}, \sigma_{bs}, \rho)$, which are estimated from the Bayes filter using the recursive algorithm in [6]. Since we consider only the exogenous sampling case for trading duration, the recursive algorithm for Bayes factors constructed below works regardless of the model assumption on trading durations.

**4.2. The Recursive Algorithms for Bayes Factors.** We illustrate how to follow the three-step blueprint of Markov chain approximation method in Section 3.3 of [6] to construct a consistent and easily-parallelized recursive algorithm for real time computing of Bayes factors. For our four model candidates, we need to calculate six pairs of Bayes Factors: $\{B_{21}, B_{12}\}$,$\{B_{31}, B_{13}\}$,$\{B_{41}, B_{14}\}$,$\{B_{32}, B_{23}\}$,$\{B_{42}, B_{24}\}$ and $\{B_{43}, B_{34}\}$. In this section, we show how to construct the recursive algorithm for pair $\{B_{41}, B_{14}\}$. The algorithm can be easily modified to apply on other pairs. Let $k = 1$ or 4 in the construction below. Observe that $\mathbb{Y}$ is a discrete space for the models, and $r_k(y)$ in (3.10) and (3.11) turns into $p^{(k)} \equiv p^{(k)}(y|X(t); \theta)$ for $y \in \mathbb{Y}$. We focus on the things related to Model 4 or Models 4 to 1. The things related to Model 1 or Models 1 to 4 can be obtained similarly.

*Step 1: Construct approximating Markov chains.* As in Bayes estimation, we use $(\theta_\epsilon^{(k)}(t), X_\epsilon^{(k)}(t))$ as approximations to $(\theta^{(k)}(t), X^{(k)}(t))$. Here, the parameters in Model 4 are denoted by $\theta^{(4)}(t) = (\mu^{(4)}, \sigma_0^{(4)}, \sigma_{en}^{(4)}, \sigma_{bs}^{(4)}, \rho^{(4)})$. The generator of (4.1) using the above parameter notations is given by:

$$\begin{aligned}\mathbf{A}_v^{(4)} f_4(x^{(4)}, \theta^{(4)}) &= \mu^{(4)} x^{(4)} \frac{\partial}{\partial x^{(4)}} f_4(\theta^{(4)}, x^{(4)}) \\ &+ \frac{1}{2}\Big(\sigma_0^{(4)} + \sigma_{en}^{(4)} V_{en} + \sigma_{bs}^{(4)} V_{bs}\Big)^2 (x^{(4)})^2 \frac{\partial^2}{\partial (x^{(4)})^2} f_4(\theta^{(4)}, x^{(4)}).\end{aligned} \tag{4.2}$$

For Model 4, we follow Section 4.3 of [6] and discretize the six-dimension state space to form a state grid. Let

$$\theta_{\vec{v}}^{(4)} = (\mu_i^{(4)}, \sigma_{0,j}^{(4)}, \sigma_{en,k}^{(4)}, \sigma_{bs,l}^{(4)}, \rho_m^{(4)})$$

where $\vec{v}$ is $(i, j, k, l, m)$. Let $\theta_\epsilon = (\mu_\epsilon^{(4)}, \sigma_{\epsilon,0}^{(4)}, \sigma_{\epsilon,en}^{(4)}, \sigma_{\epsilon,bs}^{(4)}, \rho_\epsilon^{(4)})$ to denote an approximate discretized parameter signal, which is random in the Bayesian framework. Then, we construct the Markov chain approximate generator $\mathbf{A}_{\epsilon,v}^{(4)}$, which is given by

$$\begin{aligned}&\mathbf{A}_{\epsilon,v}^{(4)} f_4(\theta_{\vec{v}}^{(4)}, x_w^{(4)}) \\ =&\beta_v^{(4)}\Big(\theta_{\vec{v}}^{(4)}, x_w^{(4)})(f_4(\theta_{\vec{v}}^{(4)}, x_w^{(4)} + \epsilon_x^{(4)}) - f_4(\theta_{\vec{v}}^{(4)}, x_w^{(4)})) \\ &+ \delta_v^{(4)}(\theta_{\vec{v}}^{(4)}, x_w^{(4)})(f_4(\theta_{\vec{v}}^{(4)}, x_w^{(4)} - \epsilon_x^{(4)}) - f_4(\theta_{\vec{v}}^{(4)}, x_w^{(4)})\Big),\end{aligned} \tag{4.3}$$

where

$$\beta_v^{(4)}(\theta_{\vec{v}}^{(4)}, x_w^{(4)}) = \frac{1}{2}\left(\frac{\big(\sigma_{0,j}^{(4)} + \sigma_{en,k}^{(4)} V_{en} + \sigma_{bs,l}^{(4)} V_{bs}\big)^2 (x_w^{(4)})^2}{(\epsilon_x^{(4)})^2} + \frac{\mu_i^{(4)} x_w^{(4)}}{\epsilon_x^{(4)}}\right),$$

and

$$\delta_v^{(4)}(\theta_{\vec{v}}^{(4)}, x_w^{(4)}) = \frac{1}{2}\left(\frac{(\sigma_{0,j}^{(4)} + \sigma_{en,k}^{(4)} V_{en} + \sigma_{bs,l}^{(4)} V_{bs})^2 (x_w^{(4)})^2}{(\epsilon_x^{(4)})^2} - \frac{\mu_i^{(4)} x_w^{(4)}}{\epsilon_x^{(4)}}\right).$$

*Step 2: Obtain the ratio filtering equations of Bayes factors for the approximate models.* When $(\theta^{(k)}, X^{(k)})$ is replaced by $(\theta_\epsilon^{(k)}, X_\epsilon^{(k)})$, $\mathbf{A}_v^{(k)}$ by $\mathbf{A}_{\epsilon,v}^{(k)}$, $\Phi^{(k)}$ by $\Phi_\epsilon^{(k)}$, and $p^{(k)}$ by $p_\epsilon^{(k)}$, there also exists a probability measure $P_\varepsilon^{(k)}$ in the place of $P^{(k)}$. Then Assumptions 2.1 - 2.5 also hold for $(\theta_\epsilon^{(k)}, X_\epsilon^{(k)}, \Phi_\epsilon^{(k)}, V)$.

Recall from Definition (3.6) that $q_{\epsilon,t}^{(41)}$ is the ratio conditional measure of approximate Models 4 to 1, and that the approximate filter ratio process can be expressed as

$$q_{\epsilon,41}(f_4, t) = \sum_{\vec{v}', w'} f_4(\theta_{\vec{v}'}^{(4)}, x_{w'}^{(4)}) q_{\epsilon,t}^{(41)}(\theta_{\vec{v}'}^{(4)}, x_{w'}^{(4)}),$$

where the sum goes over the whole state grid. Similarly, $q_{\epsilon,t}^{(14)}$ and $q_{\epsilon,14}(f_1, t)$ can be properly defined. Hence, Theorem 3.2 with $k = 4$ and $j = 1$ gives the following approximate ratio filtering equation, which can be separated as in (3.12) and (3.13):

$$q_{\epsilon,41}(f_4, t_{i+1}-) = q_{\epsilon,41}(f_4, t_i) + \int_{t_i}^{t_{i+1}-} q_{\epsilon,41}(\mathbf{A}_v^{(4)} f_4, s) ds, \tag{4.4}$$

when there is no trade and when a trade happens at time $t_{i+1}$ with price $y_{i+1}$,

$$q_{\epsilon,41}(f_4, t_{i+1}) = \frac{q_{\epsilon,41}(f_4 r_4(y_{i+1}), t_{i+1}-)}{q_{\epsilon,14}(r_1(y_{i+1}), t_{i+1}-)} q_{\epsilon,14}(1, t_{i+1}-). \tag{4.5}$$

*Step 3 Convert Equations (4.4) and (4.5) and the related ones of $q_{\epsilon,14}(f_1, t)$ to the recursive algorithm*, which computes a pair of approximate ratio conditional measures, $(q_{\epsilon,t}^{(41)}, q_{\epsilon,t}^{(14)})$, whose total is the pair Bayes factors. Namely, for example, $B_{\epsilon,41}(t) = q_{\epsilon,41}(1, t) = \sum_{\vec{v}', w'} q_{\epsilon,41}(\theta_{\vec{v}'}^{(4)}, x_{w'}^{(4)}; t)$, where the sum goes over all the state grid of Model 4, and $q_{\epsilon,41}(\theta_{\vec{v}}^{(4)}, x_w^{(4)}; t)$ is to be defined below.

Definition 4.1. *The ratio conditional measure mass of approximate Models 4 to 1 at $(\theta_{\vec{v}}, x_w)$ and at time $t$ is denoted by*

$$q_{\epsilon,41}(\theta_{\vec{v}}^{(4)}, x_w^{(4)}; t) = q_{\epsilon,t}^{(41)}\left\{\theta_\epsilon^{(4)} = \theta_{\vec{v}}^{(4)}, X_\epsilon^{(4)}(t) = x_w^{(4)}\right\}.$$

The collection of $q_{\epsilon,41}(\theta_{\vec{v}}^{(4)}, x_w^{(4)}; t)$ for the whole state grid is an discretized approximation of the ratio conditional measure $q_t^{(41)}$, and evolves continuously in time according to Equations (4.4) and (4.5).

There are two more substeps in making the recursive algorithm. The first one is to take $f_4$ as the lattice-point indicator given below:

$$\mathbf{I}_{\{\theta_{\vec{v}}^{(4)}, x_w^{(4)}\}}(\theta_\epsilon^{(4)}, X_\epsilon^{(4)}(t)), \tag{4.6}$$

which takes value one when $\theta_\epsilon^{(4)} = \theta_{\vec{v}}^{(4)}$ and $X_\epsilon^{(4)}(t) = x_w^{(4)}$, and value zero otherwise. Then,

$$q_{\epsilon,41}\Big(\delta_v^{(4)}(\theta_\epsilon^{(4)}, X_\epsilon^{(4)}(t))\mathbf{I}_{\{\theta_{\vec{v}}^{(4)}, x_w^{(4)}\}}(\theta_\epsilon^{(4)}, X_\epsilon^{(4)}(t) - \epsilon_x^{(4)}), t\Big) = q_{\epsilon,41}(\theta_{\vec{v}}^{(4)}, x_{w+1}^{(4)}; t).$$

The second substep is to approximate the time integral in Equation (4.4) by an Euler scheme, which is similar to the approximation in Section 4.3 of [6]. The final propagation part of the recursive algorithm has two cases. Case 1, if $t_{i+1} - t_i \le LL$, the length controller for Euler scheme, then

$$\begin{aligned} & q_{\epsilon,41}(\theta_{\vec{v}}^{(4)}, x_w^{(4)}; t_{i+1}-) \\ \approx & q_{\epsilon,41}(\theta_{\vec{v}}^{(4)}, x_w^{(4)}; t_i) + \Big[\beta_v^{(4)}(\theta_{\vec{v}}^{(4)}, x_{w-1}^{(4)}) q_{\epsilon,41}(\theta_{\vec{v}}^{(4)}, x_{w-1}^{(4)}; t_i) \\ & - \big(\beta_4^{(4)}(\theta_{\vec{v}}^{(4)}, x_w^{(4)}) + \delta_4^{(4)}(\theta_{\vec{v}}^{(4)}, x_w^{(4)})\big) q_{\epsilon,41}(\theta_{\vec{v}}^{(4)}, x_w^{(4)}; t_i) \\ & + \delta_4^{(4)}(\theta_{\vec{v}}^{(4)}, x_{w+1}^{(4)}) q_{\epsilon,41}(\theta_{\vec{v}}^{(4)}, x_{w+1}^{(4)}; t_i)\Big](t_{i+1} - t_i). \end{aligned} \tag{4.7}$$

Case 2, if $t_{i+1} - t_i > LL$, then we can choose a thinner partition $\{t_{i,0} = t_i, t_{i,1}, \ldots, t_{i,n} = t_{i+1}\}$ of $[t_i, t_{i+1}]$ with $\max_j |t_{i,j+1} - t_{i,j}| < LL$ and then apply repeatedly the recursive algorithms given by the above equations from $t_{i,0}$ to $t_{i,1}$, then $t_{i,2}$,. . ., until $t_{i,n} = t_{i+1}$.

Suppose a trade at price $y_{i+1}$ occurs at time $t_{i+1}$. Then, the updating Equation in (4.5) becomes,

$$\begin{aligned} & q_{\epsilon,41}(\theta_{\vec{v}}^{(4)}, x_w^{(4)}; t_{i+1}) \\ = & \frac{q_{\epsilon,41}(\theta_{\vec{v}}^{(4)}, x_w^{(4)}; t_{i+1}-) p_\epsilon^{(4)}(y_{i+1}|x_w^{(4)}, \rho_m)}{\sum_{\vec{v}', w'} q_{\epsilon,14}(\theta_{\vec{v}'}^{(1)}, x_{w'}^{(1)}; t_{i+1}-) p(y_{i+1}|x_{w'}^{(1)}, \rho_{m'})} \times \Big(\sum_{\vec{v}', w'} q_{\epsilon,14}(\theta_{\vec{v}'}^{(1)}, x_{w'}^{(1)}; t_{i+1}-)\Big) \end{aligned} \tag{4.8}$$

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

Equations (4.7), (4.8) and the two corresponding equations for $q_{\epsilon,14}(\cdot, \cdot; t)$ compose the recursive algorithm we employ to calculate the approximate conditional measures at time $t_{i+1}$ from those at time $t_i$ to time $t_{i+1}-$, and then to time $t_{i+1}$. At time $t_{i+1}$, the Bayes factor $B_{\epsilon,41}(t_{i+1}) = \sum_{\vec{v}', w'} q_\varepsilon^{(4)}(\theta_{\vec{v}'}^{(4)}, x_{w'}^{(4)}; t_{i+1})$, where the sum goes over all the lattices in the discretized state space.

As for the choice of the prior, we can follow [6] to assign uniform priors on parameters (if prior information) and count on the final marginal posterior obtained to identify an appropriate range and step size for the prior of each parameter. These priors are used to compute the Bayes factors. The consistency of the recursive algorithm comes from the consistency of the Euler scheme in approximating the time integral as well as result (4) of Theorem 3.3.

Similarly, we can calculate the other approximate Bayes factors $B_{\epsilon,kj}$ for $k, j = 1, 2, 3, 4$ and $k \neq j$. With all the Bayes factors available, the approximate posterior model probabilities can be computed. For example, $P_\epsilon(M_4|\mathcal{F}_t^{\Phi,V}) = \frac{1}{1+B_{\epsilon,14}(t)+B_{\epsilon,24}(t)+B_{\epsilon,34}(t)}$.

**4.3. Simulation Tests.** We code a Fortran program for the recursive algorithms, which are able to output real-time Bayes factors. In this section, we examine the program for model selection using the simulated data described in Section 4.4 of [6]. The simulated data set is generated based on Model 4 and contains 2,000 observations. Between the $1000^{th}$ and $1001^{st}$ transaction, a news announcement impulse occurs and the 50 subsequent trades with the news dummy equal to one. After that, things are back to normal, namely, news dummy equal to zero.

To assess model selection in recovering the true data generating process, Table 4.1 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 time 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 and 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 ([1]). Figure 4.1 plots trade-by-trade posterior model probabilities of Models 1 to 4 with the shaded area indicating news period ($V_{en} = 1$) and further validates that the posterior model probabilities provides a sensitive but stable and powerful tool for selecting the right model.

With these simulation results, we have grounds to accept that the recursive algorithm can deliver model selection when employed to real market data. We explore such tests in the next subsection.

**4.4. Empirical Tests.** We use the Bayes factors recursive algorithm to examine the tick transaction data for the OTR 10-Year Treasury note during the period 8/15/2000 to 12/31/2000. Related data description and Bayes estimation results are given in Section 4.5 of [6]. Here, we further provide the results for Bayesian model selection.

We calculate the Bayes factors to determine which model best describes the OTR 10-year Treasury note transactions data. Table 4.2 reports the daily average Bayes factors and the corresponding posterior model probabilities on selected days. By the third trading

**Table 4.1**
*Bayes factors for the simulated dataset*

| | Bayes Factors | | | | | | Posterior Model Probabilities | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| N | $B_{41}$ | $B_{42}$ | $B_{43}$ | $B_{31}$ | $B_{32}$ | $B_{21}$ | P(Model 1) | P(Model 2) | P(Model 3) | P(Model 4) |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.25 | 0.25 | 0.25 | 0.25 |
| 200 | 253.6 | 253.6 | 1.0 | 253.6 | 253.6 | 1.0 | 0.0020 | 0.0020 | 0.49804 | 0.49803 |
| 1000 | 115485.7 | 115485.7 | 1.0 | 115486.1 | 115486.1 | 1.0 | 0.00000 | 0.00000 | 0.50000 | 0.50000 |
| 1001 | 250876.6 | 114265.0 | 2.4 | 104208.0 | 47462.8 | 2.2 | 0.00000 | 0.00001 | 0.29347 | 0.70652 |
| 1050 | 5.89E+12 | 180017.5 | 214912.578 | 27396860 | 0.80 | 7.43E+07 | 0.00000 | 0.00001 | 0.00000 | 0.99999 |
| 2000 | 2.87E+21 | 9.51E+13 | 262700.594 | 1.09E+16 | 3.62E+08 | 6.77E+07 | 0.00000 | 0.00000 | 0.00000 | 1.00000 |

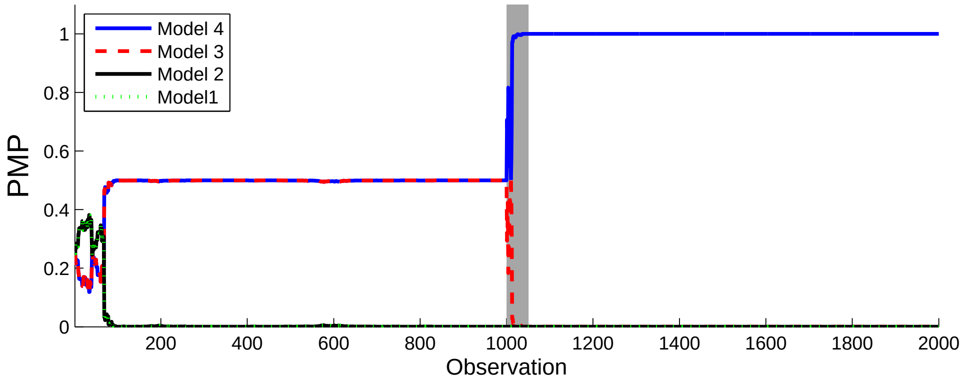


**Figure 4.1.** *Trade-by-Trade Posterior Model Probabilities (PMP) for Models 1-4 (Simulated data)*

day (8/17/2000), the Bayes factors $B_{41}$, $B_{42}$, $B_{31}$ and $B_{32}$ exceed the benchmark 150 value. These factors increase to roughly $2.5E6$ 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 4.2 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.

The above empirical mode selection results further confirm that both information-based and inventory-based effects can explain the observed volatility microstructure in the OTR 10-year Treasury note market. Moreover, the example given here further demonstrates that the Bayesian model selection approach with computational algorithms can provide a new method for tests of UHF financial market data.

**Table 4.2**

*Average Bayes Factors for selected dates, 10-year OTR Note, 8/15/2000 - 12/31/2000*

| | Bayes Factors | | | | | | Posterior Model Probability | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Date | $B_{41}$ | $B_{42}$ | $B_{43}$ | $B_{31}$ | $B_{32}$ | $B_{21}$ | P(Model 1) | P(Model 2) | P(Model 3) | P(Model 4) |
| 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 | 94.8 | 94.8 | 1.0 | 94.8 | 94.8 | 1.0 | 0.005219 | 0.005219 | 0.494782 | 0.494780 |
| 8/17/2000 | 665.7 | 665.7 | 1.0 | 665.7 | 665.7 | 1.0 | 0.000750 | 0.000750 | 0.499252 | 0.499248 |
| 8/21/2000 | 3.4E+07 | 3.4E+07 | 1.0 | 3.4E+07 | 3.4E+07 | 1.0 | 0.000000 | 0.000000 | 0.500001 | 0.499999 |
| 8/22/2000 | 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 | 0.000000 | 1.000000 |
| 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 | 0.000000 | 1.000000 |
| 12/29/2000 | 3.41E+33 | 2.91E+07 | 1.84E+34 | 1.12E+08 | 1.45E-27 | 7.42E+34 | 0.000000 | 0.000000 | 0.000000 | 1.000000 |

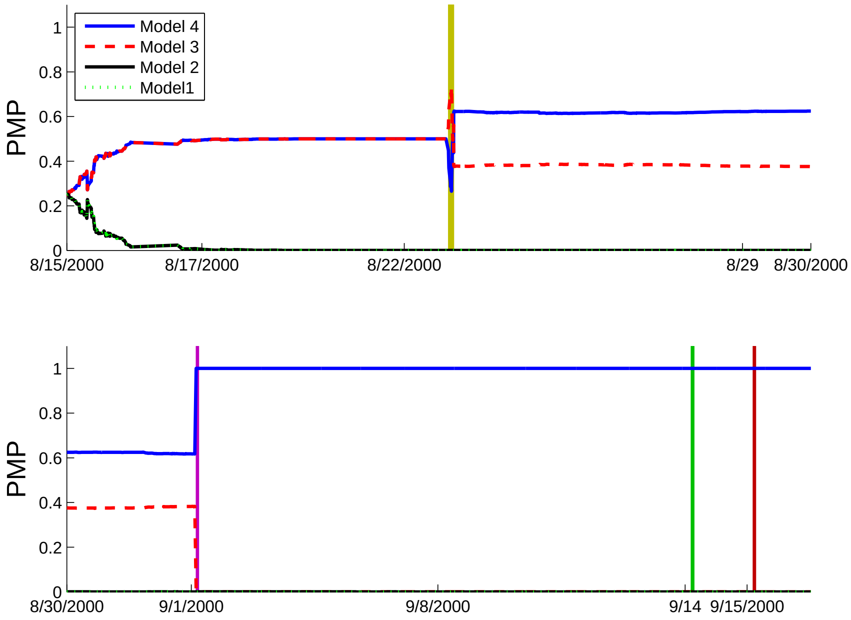


**Figure 4.2.** *Trade-by-Trade Posterior Model Probabilities (PMP) for Models 1-4 (Bond data)*

**5. Conclusions.** To quantify model uncertainty for a class of partially-observed Markov models with MPP observations for UHF data, we adopt Bayesian approach and develop Bayesian model selection via filtering equations. We derive the related filtering equations and prove a weak convergence theorem. We illustrate how to develop efficient recursive algorithms to implement online model selection, and apply the approach to the four concrete models constructed for U.S. Treasury Notes trade-by-trade data from GovPX. We show that the computed Bayes factors and the posterior model probabilities are sensitive in discovering model change by simulation. We apply the computational algorithms to the Treasury Notes transaction data and further support the empirical findings in [6].

Besides the research topics discussed in [6], we would like to add one more possible topic particularly related to model selection. There are two classes of important benchmark models in empirical and mathematical finance: stochastic volatility models and exponential Lévy models. With the speed gained by more efficient algorithms and the GPU high performance computing power, real-time model uncertainty quantification for the two model classes is an exciting research area.

**Acknowledgements:** The previous version of this paper and [6] was a working paper "Econometric Analysis via Filtering for Ultra-High Frequency Data". Part of the works were done when Yong Zeng visited and taught a special topic graduate course on *Statistical Analysis of Ultra-high Frequency Financial Data - An Overview and A New Filtering Approach* in Department of Operations Research and Financial Engineering (ORFE) in Princeton University in Spring 2007. He especially thanks Savas Dayanik for the invitation and the ORFE Department for the hospitality. We thank Yacine Aït-Saharia, Rene Carmona, Jianqing Fan, Tom Kurtz, Per Mykland, Ronnie Sircar, Wei Sun, Yazhen Wang, Jie Xiong, and Thaleia Zariphopoulou for helpful discussions. The works were presented at several seminars and conferences and we thanks participants for comments and suggestions.

## Appendix A. Proofs for Main Results.

First, we prove four lemmas, which are used in proving Theorem 3.1.

### A.1. Four Lemmas and their Proofs.

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, $E[X|\mathcal{G}\vee\mathcal{D}] = E[X|\mathcal{G}]$.*

*Proof.* $E[X|\mathcal{G}]$ is $\mathcal{G}$-measurable, and therefore, $\mathcal{G}\vee\mathcal{D}$-measurable. *Dynkin Class Theorem* (see, for example, Page 497 of [4]) implies that it suffices to show that for $A = G\bigcap D$, where $G \in \mathcal{G}$ and $D \in \mathcal{D}$, $\int_A XdP = \int_A E[X|\mathcal{G}]dP$. Now, $\int_A XdP = \int_{G\bigcap D} XdP = E[XI_G I_D] = E[XI_G]E[I_D] = E[E[XI_G|\mathcal{G}]]E[I_D] = E[I_G E[X|\mathcal{G}]]E[I_D] = E[I_G E[X|\mathcal{G}]I_D] = \int_{G\bigcap D} E[X|\mathcal{G}]dP = \int_A E[X|\mathcal{G}]dP$. The third and sixth equalities are because of the independence between $\mathcal{D}$ and $\mathcal{H}\vee\mathcal{G}$. ■

Lemma A.2. *Suppose that a vector stochastic processes $X$ is independent of a marked point process $\Phi$ with compensator $\mu(dy)dt$ and that $V$ is an observable deterministic process. Let $M(t)$ be a martingale with respect to $\{\mathcal{F}_t^{X,V}\}$ with $E(M(0)) = 0$ and let $U$ be $\mathcal{F}_t^{X,V,\Phi}$-predictable. If $E[|\int_0^t U(s)dM(s)|] < \infty$. Then $E[\int_0^t U(s)dM(s)|\mathcal{F}_t^{\Phi,V}] = 0$.*

*Proof.* Independence of $X$ and $\Phi$ implies that $M(t)$ is still a martingale with respect to $\mathcal{F}_t^{X,V,\Phi}$ and makes $\int U(s)dM(s)$ well defined. It suffices to show that for every bounded, $\mathcal{F}_t^{\Phi}$-measurable random variable H, $E^Q[H\int_0^t U(s-)dM(s)] = 0$. Since $\Phi$ is a unit Poisson random measure under $Q$, martingale representation theorem (page 343 in [16]) gives that there exists a predictable $g(t, y)$ such that

$$H = H_0 + \int_0^t \int_{\mathbb{Y}} g(s, y)(\Phi(d(s, y)) - \mu(dy)ds),$$

where $H_0$ is a constant and $\mathbb{Y}$ is the marked space of $\Phi$. Observe that the independence implies $[M, \Phi] = 0$. The formula of integration by part gives

$$\Big(\int_0^t U(s)dM(s)\Big)\Big(\int_0^t\int_{\mathbb{Y}} g(s,y)(\Phi(d(s,y)) - \mu(dy)ds)\Big)$$

$$= \int_0^t \Big(\int_0^s\int_{\mathbb{Y}} g(u,y)\Big(\Phi(d(u,y)) - \mu(dy)du\Big)\Big) U(s)dM(s)$$

$$+ \int_0^t\int_{\mathbb{Y}} \Big(\int_0^s U(u)dM(u)\Big) g(s,y)\Big(\Phi(d(s,y)) - \mu(dy)ds\Big)$$

Then, $E^Q[H \int_0^t U(s-)dM(s)] = 0$ under the given moment condition, because all the integrators are martingales. ■

Lemma A.3. *Suppose a vector stochastic process $X$ and a Poisson random measure $\Phi$ with compensator $\mu(dy)dt$ are independent. Suppose that $U$ is $\mathcal{F}_t^{X,\Phi,V}$-predictable where $V$ is an observable deterministic process and $E[\int_0^t\int_{\mathbb{Y}} |U(s,y)|\mu(dy)ds] < \infty$. Then,*

$$E\Big[\int_0^t\int_{\mathbb{Y}} U(s-,y)\Phi(d(s,y))\Big|\mathcal{F}_t^{\Phi,V}\Big] = \int_0^t\int_{\mathbb{Y}} E[U(s-,y)|\mathcal{F}_s^{\Phi,V}]\Phi(d(s,y)).$$

*Proof.* Again, independence of $X$ and $\Phi$ makes $\int_0^t\int_{\mathbb{Y}} U(s-,y)\Phi(d(s,y))$ well defined with the filtration $\mathcal{F}_t^{X,V,\Phi}$ where $\mathbb{Y}$ is the marked space of $\Phi$. Since $U$ is cadlag, there is a cadlag version of $E[U(t,y)|\mathcal{F}_t^{\Phi,\vec{V}}]$ by [19].

We first start with $U$ as a simple process. Assume that $\{A_j\}_{j=1}^m$ is a partition of $\mathbb{Y}$ and that $t_0 = 0$ and $t_{n+1} = t$. Let $U(t,y) = \sum_{j=1}^m \xi_{0j} I_{\{0\}\times A_j}(t,y) + \sum_{j=1}^m\sum_{i=0}^n \xi_{ij} I_{(t_i,t_{i+1}]\times A_j}(t,y)$ where $\xi_{ij}$ is $\mathcal{F}_{t_i}^{X,\Phi}$-measurable. Then, with $\Phi(A_j, 0) = 0$,

$$
\begin{aligned}
& E[\int_0^t\int_{\mathbb{Y}} U(s-,y)\Phi(d(y,s))|\mathcal{F}_t^{\Phi,V}] \\
=& \sum_{j=1}^m\sum_{i=0}^n E[\xi_{ij}|\mathcal{F}_t^{\Phi,V}]\big(\Phi(A_j,t_{i+1}) - \Phi(A_j,t_i)\big) \\
=& \sum_{j=1}^m\sum_{i=0}^n E[\xi_{ij}|\mathcal{F}_{t_i}^{\Phi,V}]\big(\Phi(A_j,t_{i+1}) - \Phi(A_j,t_i)\big) \\
=& \int_0^t\int_{\mathbb{Y}} E[U(s-,y)|\mathcal{F}_s^{\Phi,V}]\Phi(d(y,s))
\end{aligned}
$$

Since

$$E\Big|\int_0^t\int_{\mathbb{Y}} U(s-,y)\Phi(d(y,s))\Big| \le E[\int_0^t\int_{\mathbb{Y}} |U(s-,y)|\mu(dy)ds] < \infty,$$

and $\Phi$ is a random count measure with finite variation, we can extend $U$ to predictable integrand as we define the stochastic integral for predictable integrand. (See, for example, Section IV.2 of [17]). ■

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*

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

The proof is similar to that of Lemma A.3.

**A.2. Proof of Theorem 3.1.** We employ the reference measure approach originally used in [20].

Let $M_f(t) = f(\theta(t), X(t)) - f(\theta(0), X(0)) - \int_0^t \mathbf{A}_v f(\theta(s), X(s))ds$. Under $Q$, $M_f(t)$ remains a mean-zero martingale. Since $[\int \mathbf{A}_v f(\theta(s), X(s))ds, L]_t = 0$, we have that $[M_f, L]_t = 0$. Then, the formula of integration by parts (see [17], page 60) gives

$$\begin{aligned} f(\theta(t), X(t))L(t) &= f(\theta(0), X(0))L(0) + \int_0^t L(s-)dM_f(s) \\ &+ \int_0^t \int_{\mathbb{Y}} f(\theta(s-), X(s-))\Big[\zeta(s-, y) - 1\Big]L(s-)\Phi(d(s, y)) \\ &+ \int_0^t L(s)\Big\{\mathbf{A}_v f(\theta(s), X(s)) - \int_{\mathbb{Y}} f(\theta(s), X(s))\Big[\zeta(s, y) - 1\Big]\mu(dy)\Big\}ds. \end{aligned} \tag{A.1}$$

We take conditional expectations with respect to the reference measure $Q$ given the observed history of $\mathcal{F}_t^{\Phi,V}$ on both sides of Equation (A.1). Then, we apply the four lemmas given in the previous section.

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

Under the reference measure $Q$, $(\theta, X)$ and $\Phi$ are independent and $M_f(t)$ is still a $\mathcal{F}_t^{\theta,X}$-martingale. We can write $Q = P_{\theta,x|v} \times Q_\Phi$. Also, $U(t) = L(t)$ is $\mathcal{F}_t^{\theta,X,\Phi,V}$-predictable and $V$ is deterministic. Assumption 2.5 and (3.2) imply that $L(t)$ is a local martingale. Since $L(t)$ is nonnegative, it is a supermartingale. Then, $E^{Q_\Phi}[L(t)] \le 1$ $P_{\theta,x|v}$−a.s., and we obtain that $E^Q[|\int_0^t L(s-)dM_f(s)|] \le E^Q[|M_f(t)|] < \infty$. Hence, Lemma A.2 implies $E^Q[\int_0^t L(s-)dM_f(s)|\mathcal{F}_t^{\Phi,V}] = 0$.

To apply Lemma A.3 under $Q$, we show that the moment condition holds. We can take $f$ to be bounded. Since $\int_0^t \int_{\mathbb{Y}} E^Q[L(s-)]\mu(dy)ds = \mu(\mathbb{Y})t < +\infty$ by Assumption 2.4, it suffices to show that $E^Q[\int_0^t \int_{\mathbb{Y}} |\zeta(s-, y)|L(s-)\mu(dy)ds] < +\infty$. Since $\zeta(t, y) =$

$\lambda(t,dy)/\mu(dy) > 0$,

$$\begin{aligned} E^Q\left[\int_0^t \int_{\mathbb{Y}} |\zeta(s,y)|L(s)\mu(dy)ds\right] &\leq E^Q\left[\int_0^t \int_{\mathbb{Y}} \lambda(s,y)L(s)ds\right] \\ &= E^Q\left[\int_0^t \bar{\lambda}(s)L(s)\Big[\int_{\mathbb{Y}} p(dy|X(s))\Big]ds\right] \\ &\leq \int_0^t E^Q[\bar{\lambda}(s)L(s)]ds = \int_0^t E^P[\bar{\lambda}(s)]ds < \infty, \end{aligned}$$

where the last inequality is by Assumption 2.5. Hence, Lemma A.3 gives

$$\begin{aligned} &E^Q\Big[\int_0^t \int_{\mathbb{Y}} f(\theta(s-),X(s-))\big[\zeta(s-,y)-1\big]L(s-)\Phi(d(s,y))|\mathcal{F}_t^{Y,V}\Big] \\ &= \int_0^t \int_{\mathbb{Y}} \rho((\zeta(y)-1)f,s-)\Phi(d(s,y)). \end{aligned}$$

Similarly, the conditions of Lemma A.4 are met and it implies

$$\begin{aligned} &E^Q\Big[\int_0^t L(s)\left\{\mathbf{A}_v f(\theta(s),X(s)) - \int_0^t \int_{\mathbb{Y}} f(\theta(s),X(s))\big[\zeta(s,y)-1\big]\mu(dy)\right\}ds|\mathcal{F}_t^{\Phi,V}\Big] \\ &= \int_0^t \rho(\mathbf{A}_v f,s)ds - \int_0^t \int_{\mathbb{Y}} \rho(\zeta(s,y)-1)f,s)\mu(dy)ds \end{aligned}$$

Summarizing the above, we have the SDE for $\rho(f,t)$, which is Equation (3.7). □

**A.3. Proof of Theorem 3.2.** We will show that $q_{kj}(f_k,t)$ satisfies Equation (3.8), and when $a_j^{(k)}(\theta^{(k)}(t),\vec{X}^{(k)}(t),t) = a_j(t)$ Equation (3.8) reduces to Equation (3.10). Then, by symmetry, $q_{jk}(f_j,t)$ satisfies Equation (3.9) and, in the special case, (3.11). Recall that $\rho_j(f_s,t)$ satisfies Equation (3.7). Then applying Ito's formula for semi-martingales ([17]) and simplifying gives us

$$\begin{aligned} \frac{\rho_k(f_k,t)}{\rho_j(1,t)} &= \frac{\rho_k(f_k,0)}{\rho_j(1,0)} + \int_0^t \int_{\mathbb{Y}} \Big[\frac{\rho_k(f_k,s)}{\rho_j(1,s)} - \frac{\rho_k(f_k,s-)}{\rho_j(1,s-)}\Big]\Phi(d(s,y)) \\ &\quad + \int_0^t \int_{\mathbb{Y}} \Big[\frac{\rho_k(\mathbf{A}_v^{(1)} f_k,s)}{\rho_j(1,s)} - \frac{\rho_k(\zeta_k(y)f_k,s)}{\rho_j(1,s)} + \frac{\rho_k(f_k,s)\rho_j(\zeta_j(y),s)}{\rho_j^2(1,s)}\Big]ds. \end{aligned} \tag{A.2}$$

First, observe that

$$\frac{\rho_k(f_k,s)\rho_j(\zeta_j(y),s)}{\rho_j^2(1,s)} = \frac{\frac{\rho_k(f_k,s)}{\rho_j(1,s)}\frac{\rho_j(\zeta_j(y),s)}{\rho_k(1,s)}}{\frac{\rho_j(1,s)}{\rho_k(1,s)}} = \frac{q_{kj}(f_k,s)q_{jk}(\zeta_j(y),s)}{q_{jk}(1,s)}.$$

Second, we make the integrand of the last integral in Equation (A.2)predictable. If a trade at $y$ occurs at time $s$, then Equation (3.7) implies that

$$\frac{\rho_k(f_k,s)}{\rho_j(1,s)} = \frac{\rho_k(f_k,s-) + \rho_k((\zeta_k(y)-1)f_k,s-)}{\rho_j(1,s-) + \rho_j(\zeta_j(y)-1,s-)}$$

$$= \frac{\rho_k(\zeta_k(y)f_k, s-)}{\rho_j(\zeta_j(y), s-)} = \frac{q_{kj}(\zeta_k(y)f_k, s-)}{q_{jk}(\zeta_j(y), s-)} q_{jk}(1, s-).$$

With the above two observations, Equation(A.2) implies Equation (3.8).

When both models have the common $\mathcal{F}_t^{\Phi,V}$-measurable trading intensity, Equation (3.8) clearly is simplified to Equation (3.10) by the similar two observations that simplifies $\pi(f,t)$ in the proof of Theorem 3.1 of [6]. □

**A.4. Proof of the Uniqueness for Theorems 3.1, 3.2 and Theorem 3.1 in [8].** Assumption 2.5 ensures the existence of a measure, under which the $\Phi$ is an MPP with compensator $\pi(\zeta(y),t)\mu(dy)dt$. As in [8], this allows us to formulated the problems as *filtered martingale problems* (FMP) proposed in [12] and further developed in [11]. Then, the uniqueness of FMP gives the uniqueness of the unnormalized and normalized filtering equations. The uniqueness in Theorem 3.2 can proven in a similar manner as in [9] and [18]. For technical details, interested readers are referred to the aforementioned references. □.

**A.5. Proof of Theorem 3.3.** Because $(\theta_\epsilon, X_\epsilon) \Rightarrow (\theta, X)$ in the Skorohod topology, (1) is obtained by *Continuous Mapping Theorem* (Corollary 3.1.9 of [4]).

For proving (2) or (3), the idea comes from Theorem 2.1 of [5] and Theorem 1 and Corollary 1 of [10] on *weak convergence of conditional expectation*, respectively. As in our setup excluding the superscript $(k)$, $(\theta, X)$ and $\Phi$ are independent under a reference measure $Q$ with R-N derivative $L$, and $(\theta_\epsilon, X_\epsilon)$ and $\Phi_\epsilon$ are independent under a reference measure $Q_\epsilon$ with R-N derivative $L_\epsilon$. According to the aforementioned theorems, it suffices to prove $((\theta_\epsilon, X_\epsilon), \Phi_\epsilon, L_\epsilon) \Rightarrow ((\theta, X), \Phi, L)$. Recall the R-N derivative $dP_\epsilon/dQ_\epsilon$ is

$$L_\epsilon(t) = \exp\left\{ \int_0^t \int_{\mathbb{Y}} \log \zeta_\epsilon(s-, y) \Phi(d(s,y)) - \int_0^t \int_{\mathbb{Y}} \Big[ \zeta_\epsilon(s-, y) - 1 \Big] \mu(dy) ds \right\}. \tag{A.3}$$

The tool to prove the above joint weak convergence is the *weak convergence of stochastic integrals* with respect to *semimartingale random measure* developed in [14] (Theorem 4.2). This result is a generalization of Theorem 2.2 of [13] on the weak convergence of stochastic integrals with respect to semimartingale. Observe that $(\theta_\epsilon, X_\epsilon) \Rightarrow (\theta, X)$ implies $(\theta_\epsilon, X_\epsilon, \zeta_\epsilon) \Rightarrow (\theta, X, \zeta)$ again by Continuous Mapping Theorem. Since $\Phi_\epsilon$ and $\Phi$ have the same distribution under reference measures, we actually have $(\theta_\epsilon, X_\epsilon, \zeta_\epsilon, \Phi_\epsilon) \Rightarrow (\theta, X, \zeta, \Phi)$. Note that $\Phi_\epsilon$ and $\Phi$ are $H^\#$-semimartingale (Example 3.18 of [14]). Hence, Continuous Mapping Theorem and Theorem 4.2 of [14] imply that

$$(\theta_\epsilon, X_\epsilon, \Phi_\epsilon, \int \log(\zeta_\epsilon) d\Phi_\epsilon) \Rightarrow (\theta, X, \Phi, \int \log(\zeta) d\Phi),$$

and

$$(\theta_\epsilon, X_\epsilon, \Phi_\epsilon, \int (\zeta_\epsilon - 1) ds) \Rightarrow (\theta, X, \Phi, \int (\zeta - 1) ds),$$

and then $((\theta_\epsilon, X_\epsilon), \Phi_\epsilon, L_\epsilon) \Rightarrow ((\theta, X), \Phi, L)$.

The weak convergence of (4) and (5) comes from the weak convergence of (2) and again Continuous Mapping Theorem. □

## REFERENCES

[1] J. O. Berger and L. Pericchi. Objective Bayesian methods for model selection: introduction and comparison. In P. Lahiri, editor, *Model Selection*, volume 38, pages 135–207. IMS Lecture Notes – Monograph Series, Beachwood Ohio, 2001.

[2] D. Crisan and B. Rozovsky, editors. *Oxford Handbook of Nonlinear Filtering*. Oxford University Press, Oxford, UK, 2011.

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

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

[5] E.M. Goggin. Convergence in distribution of conditional expectations. *Annals of Probability*, 22:1097–1114, 1994.

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

[7] R. E. Kass and A. E. Raftery. Bayes factors and model uncertainty. *Journal of the American Statistical Association*, 90:773–795, 1995.

[8] W.H. Kliemann, G. Koch, and F. Marchetti. On the unnormalized solution of the filtering problem with counting process observations. *IEEE Transactions on Information Theory*, 36:1415–1425, 1990.

[9] M. Kouritzin and Y. Zeng. Bayesian model selection via filtering for a class of micro-movement models of asset price. *International Journal of Theoretical and Applied Finance*, 8:97–121, 2005.

[10] M. Kouritzin and Y. Zeng. Weak convergence for a type of conditional expectation: Application to the inference for a class of asset price models. *Nonlinear Analysis: Theory, Methods and Applications*, 60:231–239, 2005b.

[11] T.G. Kurtz and G. Nappa. The filtered martingale problem. In D. Crisan and B. Rozovsky, editors, *Oxford Handbook of Nonlinear Filtering*, pages 129 – 165. Oxford University Press, 2011.

[12] T.G. Kurtz and D.L. Ocone. Unique characterization of conditional distributions in nonlinear filtering. *Annals of Probability*, 16:80–107, 1988.

[13] T.G. Kurtz and P.E. Protter. Weak limit theorems for stochastic integrals and stochastic differential equations. *Annals of Probability*, 19:1035–1070, 1991.

[14] T.G. Kurtz and P.E. Protter. Weak convergence of stochastic integrals and differential equations ii: Infinite dimensional case. In D. Talay and L. Tubaro, editors, *Probabilistic Models for Nonlinear Partial Differential Equations*, volume 38, pages 197–285. Springer, Beachwood Ohio, 1996.

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

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

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

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

[19] M. Yor. *Sur les théories du filtrage et de la prédiction*, volume 581, pages 257–297. Springer, Berlin, 1977.

[20] M. Zakai. On the optimal filtering of diffusion processes. *Z. Wahrsch. verw. Gebiete*, 11:230 – 243, 1969.

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