Available online at www.sciencedirect.com
SCIENCE DIRECT®

Nonlinear Analysis 60 (2005) 231 – 239

Nonlinear Analysis

www.elsevier.com/locate/na

Weak convergence for a type of conditional expectation: application to the inference for a class of asset price models☆

Michael A. Kouritzin[a], Yong Zeng[b,*]

[a] Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Al, Canada T6G 2G1
[b] Department of Mathematics and Statistics, University of Missouri at Kansas City, Kansas City, MO 64110, USA

Received 12 March 2004; accepted 1 July 2004

Abstract

We prove weak convergence of a type of conditional expectation, which provides a straightforward proof of Goggin's Theorem and further proves the consistency of (integrated) likelihood, posterior, and Bayes factor for a class of transactional asset price models recently developed.

Keywords: Conditional expectation; Filtering; Counting process

1. Introduction

The weak convergence of (X^N, Y^N) on a probability space $(\Omega^N, \mathscr{F}^N, P^N)$ to (X, Y) on $(\Omega, \mathscr{F}, P)$ does not necessarily imply that $E^{P^N}[F(X^N) \mid Y^N] \Rightarrow E[F(X) \mid Y]$, even for bounded, continuous functions, F. However, Goggin [2] showed that the limit of the conditional expectations can be determined if it is possible to make an absolutely continuous change of probability measure, from P^N to Q^N with $L^N(X^N, Y^N) = \mathrm{d}P^N/\mathrm{d}Q^N$ being the Radon–Nikodym derivative, so that X^N and Y^N are independent under Q^N. Goggin's main condition guaranteeing the weak convergence of the conditional expectation is

☆ We thank Thomas G. Kurtz for a productive conversation.

* Corresponding author. Tel.: +1-816-235-5850.
E-mail addresses: mkouritz@math.ualberta.ca (M.A. Kouritzin), zeng@mendota.umkc.edu (Y. Zeng).

0362-546X/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.na.2004.07.052

that $(X^N, Y^N, L^N(X^N, Y^N))$ under Q^N converges in distribution to $(X, Y, L(X, Y))$ under Q. Under the same conditions of Goggin [2], we provide a short proof of the weak convergence of this type of conditional expectation: $E^{Q^N}[F(X^N)L^N(X^N, Y^N) \mid Y^N] \Rightarrow E^Q[F(X)L(X, Y) \mid Y]$ as $N \to \infty$, establishing Goggin's theorem. We note that Crimaldi and Pratelli in [7] at about the same time independently proved a similar result.

We note that $E^Q[F(X)L(X, Y) \mid Y]$ for suitably chosen F solves the Duncan–Mortensen–Zakai (unnormalized filtering) equation in the classical filtering problem. Hence, weak convergence of this type of conditional expectation can be used in proving the consistency of some numerical algorithms for computing Duncan–Mortensen–Zakai equation. However, this paper focuses on another important application to the statistical inference for a class of partially observed models of asset price for transaction data recently proposed in [6]. The main appeal of the class of models is that they are framed as a filtering problem with counting process observations. Then, $E^Q[F(X)L(X, Y) \mid Y]$ relates to the continuous-time likelihood (or integrated likelihood in Bayesian statistics) of the model, which is specified by the unnormalized filtering equation. The Markov chain approximation method can be applied to develop recursive algorithms based on the unnormalized filtering equation to compute the continuous-time likelihood. Therefore, weak convergence of this type of conditional expectation is useful in proving the consistency of the recursive algorithms for likelihoods. Furthermore, together with the continuous mapping theorem, weak convergence of this type of conditional expectation can also prove the consistency of the recursive algorithms for computing posterior and Bayes factor, which is the ratio of the integrated likelihoods of the two models considered in model selection.

Section 2 presents and proves the main theorem with two corollaries. The first corollary is Goggin's theorem and is related to the consistency of posterior. The second relates to the consistency of Bayes factor. It should be noted that the results in Section 2 are fairly general, because they are for the random variables in the complete, separable metric spaces. To present an application, Section 3 first reviews the class of models proposed in [6], then proves the consistencies of the likelihoods, posterior and Bayes factor.

2. The main result

Theorem 1. *Suppose that S_1 and S_2 are complete, separable metric spaces and that (X^N, Y^N), $N = 1, 2, \ldots,$ and (X, Y) are $S_1 \times S_2$-valued random variables defined on the probability spaces $(\Omega^N, \mathscr{F}^N, P^N)$ and $(\Omega, \mathscr{F}, P)$, respectively. Suppose that $\{(X^N, Y^N)\}$ converges in distribution to (X, Y), that $P^N \ll Q^N$ on $\sigma(X^N, Y^N)$ with $\mathrm{d}P^N/\mathrm{d}Q^N = L^N(X^N, Y^N)$, and that $Q^N(Q)$ is a probability measure on $\mathscr{F}^N(\mathscr{F})$ such that X^N, Y^N (X, Y) are independent under $Q^N(Q)$.*

Suppose that the Q^N-distribution of $(X^N, Y^N, L^N(X^N, Y^N))$ converges weakly to the Q-distribution of $(X, Y, L(X, Y))$, where $E^Q[L(X, Y)] = 1$. Then, the following hold:

(i) *$P \ll Q$ on $\sigma(X, Y)$ and $\mathrm{d}P/\mathrm{d}Q = L(X, Y)$;*

(ii) *for every bounded continuous function $F : S_1 \to R$, $E^{Q^N}[F(X^N)L^N(X^N, Y^N) \mid Y^N]$ converges weakly to $E^Q[F(X)L(X, Y) \mid Y]$ as $N \to \infty$.*

Proof. Part (i) is from part (i) of Theorem 2.1 in [2]. The following is the proof for part (ii).

Pick $\varepsilon > 0$. Then, as in [2], there exists a bounded, continuous function L^{ε}, w.l.o.g. strictly positive, such that $E^{Q}[|L(X, Y) - L^{\varepsilon}(X, Y)|] < \varepsilon$.

Let $Z^N = L^N(X^N, Y^N)$ and $Z = L(X, Y)$. Using Skorohod's representation, we can find a probability space $(\hat{\Omega}, \hat{\mathscr{F}}, \hat{Q})$ where $(\hat{X}^N, \hat{Y}^N, \hat{Z}^N)$ and $(\hat{X}, \hat{Y}, \hat{Z})$ are defined with the following three properties: for each N, $(\hat{X}^N, \hat{Y}^N, \hat{Z}^N)$ has the same distribution as $(X^N, Y^N, L^N(X^N, Y^N))$ under Q^N; $(\hat{X}, \hat{Y}, \hat{Z})$ has the same distribution as $(X, Y, L(X, Y))$ under Q; and $(\hat{X}^N, \hat{Y}^N, \hat{Z}^N) \to (\hat{X}, \hat{Y}, \hat{Z})\,\hat{Q}$-a.s. Suppose $|F| \leqslant C$. We observe that

$$\begin{aligned} &E^{\hat{Q}}\left[|E^{\hat{Q}}\left[F(\hat{X}^N)\hat{Z}^N|\hat{Y}^N\right] - E^{\hat{Q}}\left[F(\hat{X})\hat{Z}|\hat{Y}\right]|\right] \\ &\quad \leqslant C E^{\hat{Q}}\left[|\hat{Z} - L^{\varepsilon}(\hat{X}, \hat{Y})|\right] + C E^{\hat{Q}}\left[|\hat{Z}^N - L^{\varepsilon}(\hat{X}^N, \hat{Y}^N)|\right] \\ &\qquad + E^{\hat{Q}}\left[|E^{\hat{Q}}\left[F(\hat{X}^N)L^{\varepsilon}(\hat{X}^N, \hat{Y}^N)|\hat{Y}^N\right] - E^{\hat{Q}}\left[F(\hat{X})L^{\varepsilon}(\hat{X}, \hat{Y})|\hat{Y}\right]|\right]. \end{aligned} \tag{1}$$

Now, it suffices to prove the right-hand side of the above inequality can be arbitrary small for large enough N. The *first term* is less than $C\varepsilon$. For the second term, we observe

$$\begin{aligned} E^{\hat{Q}}\left[|L^{\varepsilon}(\hat{X}^N, \hat{Y}^N) - \hat{Z}^N|\right] &\leqslant E^{\hat{Q}}\left[|L^{\varepsilon}(\hat{X}^N, \hat{Y}^N) - L^{\varepsilon}(\hat{X}, \hat{Y})|\right] \\ &\quad + E^{\hat{Q}}\left[|L^{\varepsilon}(\hat{X}, \hat{Y}) - \hat{Z}|\right] + E^{\hat{Q}}\left[|\hat{Z} - \hat{Z}^N|\right]. \end{aligned}$$

Since $(\hat{X}^N, \hat{Y}^N) \to (\hat{X}, \hat{Y})$ ($\hat{Q}$-a.s., and L^{ε} is bounded and continuous, the first expectation is less than ε for large N. The second expectation is less than ε by the assumption on L^{ε}. Since $\hat{Z}^N \to \hat{Z}$ $\hat{Q}$-a.s., and $\int \hat{Z}^N \,\mathrm{d}\hat{Q} \to \int \hat{Z}\,\mathrm{d}\hat{Q}$ (since all the integrals are identically 1), the last expectation is less than ε for large N. Therefore, the *second term* of (1) is less than $3C\varepsilon$.

Noting that both F and L^{ε} are bounded and continuous real variables, one finds that the mapping: $(y, \mu) \to \int F(x)L^{\varepsilon}(x, y)\mu(\mathrm{d}x)$ is continuous. Therefore, noting that $(\hat{Y}^N, \mu^N) \to (\hat{Y}, \mu)$, where $\mu^N = \hat{Q}^{-1}(\hat{X}^N)$ and $\mu = \hat{Q}^{-1}(\hat{X})$, we obtain that the *third term* of (1) is less than ε for large N. □

Theorem 1 relates to the consistency of likelihood in the application of Section 3. The following corollary is Goggin Theorem and relates to the consistency of posterior in the application of Section 3. We use the notation, $X^N \Rightarrow X$ ($X_{\varepsilon} \Rightarrow X$), to mean X^N (X_{ε}) converges weakly to X in the Skorohod topology as $N \to \infty$ ($\varepsilon \to 0$).

Corollary 1. *Under the same conditions of Theorem* 1 *and the assumption that* $Q\{E^Q[L(X, Y)\,|\,Y] = 0\} = 0$, *for every bounded continuous function* $F : S_1 \to R$, *we have*: $E^{P^N}[F(X^N)\,|\,Y^N]$ *converges weakly to* $E^P[F(X)\,|\,Y]$ *as* $N \to \infty$.

Proof. Bayes theorem, Theorem 1 and continuous mapping theorem imply that

$$\begin{aligned} E^{P^N}[F(X^N)\,|\,Y^N] &= \frac{E^{Q^N}[F(X^N)L^N(X^N, Y^N)\,|\,Y^N]}{E^{Q^N}[L^N(X^N, Y^N)\,|\,Y^N]} \;\Rightarrow\; \frac{E^Q[F(X)L(X, Y)\,|\,Y]}{E^Q[L(X, Y)\,|\,Y]} \\ &= E^P[F(X)\,|\,Y]. \qquad \square \end{aligned}$$

Corollary 2 relates to the consistency of Bayes factor in the application of Section 3.

Corollary 2. *Suppose that for* $k=1,2$, $S_1^{(k)}$ *and* $S_2^{(k)}$ *are complete, separable metric spaces, and that* (X_k^N, Y_k^N) $((X_k, Y_k))$ *on* $(\Omega_k^N, \mathscr{F}_k^N, P_k^N)$ $((\Omega_k, \mathscr{F}_k, P_k))$ *satisfies the same conditions of Theorem* 1. *Suppose that* $Q_k\{E^{Q_k}[L_k(X_k, Y_k) \mid Y_k] = 0\} = 0$ *for* $k = 1, 2$. *Define*

$$q_1^N(F_1) = \frac{E^{Q_1^N}[F_1(X_1^N)L_1^N(X_1^N, Y_1^N) \mid Y_1^N]}{E^{Q_2^N}[L_2^N(X_2^N, Y_2^N) \mid Y_2^N]}$$

and

$$q_2^N(F_2) = \frac{E^{Q_2^N}[F_2(X_2^N)L_2^N(X_2^N, Y_2^N) \mid Y_2^N]}{E^{Q_1^N}[L_1^N(X_1^N, Y_1^N) \mid Y_1^N]}.$$

Define $q_1(F_1)$ *and* $q_2(F_2)$ *similarly. Then, for every bounded continuous function* $F_k : S_1^{(k)} \to R$, $(q_1^N(F_1), q_2^N(F_2))$ *converges weakly to* $(q_1(F_1), q_2(F_2))$.

Proof. By the proof of Theorem 1, $\left(E^{Q_1^N}\left[F_1(X_1^N)L_1^N(X_1^N, Y_1^N) \mid Y_1^N\right], E^{Q_2^N}\left[F_2(X_2^N)L_2^N(X_2^N, Y_2^N) \mid Y_2^N\right]\right) \Rightarrow \left(E^{Q_1}\left[F_1(X_1)L_1(X_1, Y_1) \mid Y_1\right], E^{Q_2}\left[F_2(X_2)L_2(X_2, Y_2) \mid Y_2\right]\right)$, and the continuous mapping theorem implies that

$$\begin{bmatrix} q_1^N(F_1) \\ q_2^N(F_2) \end{bmatrix} = \begin{bmatrix} \frac{E^{Q_1^N}[F_1(X_1^N)L_1^N(X_1^N,Y_1^N) \mid Y_1^N]}{E^{Q_2^N}[L_2^N(X_2^N,Y_2^N) \mid Y_2^N]} \\ \frac{E^{Q_2^N}[F_2(X_2^N)L_2^N(X_2^N,Y_2^N) \mid Y_2^N]}{E^{Q_1^N}[L_1^N(X_1^N,Y_1^N) \mid Y_1^N]} \end{bmatrix} \Rightarrow \begin{bmatrix} \frac{E^{Q_1}[F_1(X_1)L_1(X_1,Y_1) \mid Y_1]}{E^{Q_2}[L_2(X_2,Y_2) \mid Y_2]} \\ \frac{E^{Q_2}[F_2(X_2)L_2(X_2,Y_2) \mid Y_2]}{E^{Q_1}[L_1(X_1,Y_1) \mid Y_1]} \end{bmatrix}$$

$$= \begin{bmatrix} q_1(F_1) \\ q_2(F_2) \end{bmatrix}. \qquad \square$$

3. Application to a class of asset price models

In this section, we first review the class of partially-observed models for transactional asset price in [6] and their continuous-time likelihoods, posterior and Bayes factors. Then, we apply the theorem and the two corollaries in Section 2 to prove the consistency of the likelihoods, posterior and Bayes factors.

3.1. The class of models

In trade-by-trade level, the asset price does not move continuously as the common asset price models such as geometric Brownian motion or jump-diffusion processes suggest, but move level-by-level due to price discreteness. When we view the prices according to price level, we model the prices of an asset as a collection of counting processes in the following form:

$$\vec{Y}(t) = \begin{pmatrix} N_1(\int_0^t \lambda_1(\theta(s), X(s), s)\,\mathrm{d}s) \\ N_2(\int_0^t \lambda_2(\theta(s), X(s), s)\,\mathrm{d}s) \\ \vdots \\ N_n(\int_0^t \lambda_n(\theta(s), X(s), s)\,\mathrm{d}s) \end{pmatrix}, \tag{2}$$

where $Y_j(t) = N_j(\int_0^t \lambda_j(\theta(s), X(s), s)\,\mathrm{d}s)$ is the counting process recording the cumulative number of trades that have occurred at the jth price level (denoted by y_j) up to time t. Moreover, $\theta(t)$ is a vector of parameters in the model and $X(t)$ is the intrinsic value process, which cannot be observed directly, but can be partially observed through $\vec{Y}$. Suppose that P is the probability measure of $(\theta, X, \vec{Y})$. We invoke five mild assumptions on the model.

Assumption 1. $\{N_j\}_{j=1}^n$ are unit Poisson processes under measure P.

Assumption 2. $(\theta, X), N_1, N_2, \ldots, N_n$ are independent under measure P.

Assumption 3. The total intensity process, $a(\theta, x, t)$, is uniformly bounded above, namely, there exists a positive constant, C, such that $0 < a(\theta, x, t) \leqslant C$ for all $t > 0$ and all (θ, x).

Remark 1. Assumption 3 is for technical reasons. These three assumptions imply that there exists a reference measure Q and that after a suitable change of measure to Q, $(\theta, X), Y_1, \ldots, Y_n$ become independent, and $Y_1, Y_2, \ldots, Y_n$ become unit Poisson processes. Also, the independence in Assumption 2 implies the probability to have any simultaneous jumps is zero.

Assumption 4. The intensity, $\lambda_j(\theta, x, t) = a(\theta, x, t)p(y_j \mid x)$, where $a(\theta, x, t)$ is the total trading intensity at time t and $p(y_j \mid x)$ is the transition probability from x to y_j, the jth price level.

Remark 2. This assumption imposes a desirable structure for the intensities of the model. It means that the total trading intensity $a(\theta(t), X(t), t)$ determines the overall rate of trade occurrence at time t and $p(y_j \mid x)$ determines the proportional intensity of trade at the price level, y_j, when the value is x. Note that $p(y_j \mid x)$ models how the trading noise enters the price process.

Assumption 5. (θ, X) is the unique cadlag solution of a martingale problem for a generator $\mathbf{A}$ such that

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

is a $\mathscr{F}_t^{\theta,X}$-martingale for $f \in \mathscr{D}(\mathbf{A})$, where $\mathscr{F}_t^{\theta,X}$ is the σ-algebra generated by $(\theta(s), X(s))_{0 \leqslant s \leqslant t}$.

Remark 3. The martingale problem and the generator approach (see [1]) provide a powerful tool for the characterization of Markov processes. Assumption 5 includes all relevant stochastic processes such as diffusion and jump-diffusion processes.

Remark 4. Under this representation, $(\theta(t), X(t))$ becomes the signal process, which cannot be observed directly, but can be partially observed through the counting processes, $\vec{Y}(t)$, corrupted by trading noise, which is modeled by $p(y_j \mid x)$ (see [6] for more about trading noise). Hence, $(\theta, X, \vec{Y})$ is framed as a *filtering problem with counting process observations*.

3.2. Foundations of statistical inference

3.2.1. The continuous-time joint likelihood

The probability measure P of $(\theta, X, \vec{Y})$ can be written as $P = P_{\theta,x} \times P_{y \mid \theta,x}$, where $P_{\theta,x}$ is the probability measure for (θ, X) such that $M_f(t)$ in Assumption 5 is a $\mathscr{F}_t^{\theta,X}$-martingale, and $P_{y \mid \theta,x}$ is the conditional probability measure for $\vec{Y}$ given (θ, X). Under P, $\vec{Y}$ depends on (θ, X). Recall from Remark 1 that there exists a reference measure Q such that under Q, (θ, X) and $\vec{Y}$ become independent, and $Y_1, Y_2, \ldots, Y_n$ become unit Poisson processes. Therefore, Q can be decomposed as $Q = P_{\theta,x} \times Q_y$, where Q_y is the product probability measure for n independent unit Poisson processes. Then, the Radon–Nikodym derivative of the model, $L(t)$, that is the *continuous-time joint likelihood* of $(\theta, X, \vec{Y})$, is,

$$
\begin{aligned}
L(t) &= \frac{\mathrm{d}P}{\mathrm{d}Q}(t) = \frac{\mathrm{d}P_{\theta,x}}{\mathrm{d}P_{\theta,x}}(t) \times \frac{\mathrm{d}P_{y \mid \theta,x}}{\mathrm{d}Q_y}(t) = \frac{\mathrm{d}P_{y \mid \theta,x}}{\mathrm{d}Q_y}(t) \\
&= \prod_{k=1}^{n} \exp\left\{ \int_0^t \log \lambda_k(\theta(s-), X(s-), s-)\, \mathrm{d}Y_k(s) \right. \\
&\qquad \left. - \int_0^t [\lambda_k(\theta(s), X(s), s) - 1]\, \mathrm{d}s \right\}. \qquad (3)
\end{aligned}
$$

3.2.2. The continuous-time likelihoods of $\vec{Y}$

However, X cannot be observed. To obtain the integrated likelihood of the model, which is the marginal likelihood of $\vec{Y}$, we may integrate $L(t)$ on (θ, X), or equivalently, write the integrated likelihood in terms of conditional expectation. Let $\mathscr{F}_t^{\vec{Y}} = \sigma\{(\vec{Y}(s)) \mid 0 \leqslant s \leqslant t\}$ be the available information up to time t.

Definition 1. Let $\phi(f, t) = E^Q[f(\theta(t), X(t))L(t) \mid \mathscr{F}_t^{\vec{Y}}]$.

If $(\theta(0), X(0))$ is fixed, then the likelihood of Y is $E^Q[L(t) \mid \mathscr{F}^{\vec{Y}}] = \phi(1, t)$. If a prior is assumed on $(\theta(0), X(0))$, then the *integrated* (*or marginal*) *likelihood* of Y is also $\phi(1, t)$.

3.2.3. The continuous-time posterior

Definition 2. Let π_t be the conditional distribution of $(\theta(t), X(t))$ given $\mathscr{F}_t^{\vec{Y}}$.

Definition 3. $\pi(f,t) = E^P[f(\theta(t), X(t)) \mid \mathscr{F}_t^{\vec{Y}}] = \int f(\theta, x)\pi_t(\mathrm{d}\theta, \mathrm{d}x)$.

If a prior is assumed on $(\theta(0), X(0))$, then π_t becomes the continuous-time posterior, which is determined by $\pi(f,t)$ for all continuous and bounded *f*.

3.2.4. Continuous-time Bayes factors

For the class of micromovement models, we suppose that Model *k* is denoted by $(\theta^{(k)}, X^{(k)}, \vec{Y}^{(k)})$ for $k = 1, 2$. Denote the joint likelihood of $(\theta^{(k)}, X^{(k)}, \vec{Y}^{(k)})$ by $L^{(k)}(t)$, which is given by Eq. (3). Denote $\phi_k(f_k, t) = E^{Q^{(k)}}[f_k(\theta^{(k)}(t), X^{(k)}(t))L^{(k)}(t) \mid \mathscr{F}_t^{\vec{Y}^{(k)}}]$. Then, the integrated likelihood of $\vec{Y}$ is $\phi_k(1,t)$, for Model *k*.

In general, the Bayes factor of Model 2 over Model 1, B_{21}, is defined as the ratio of integrated likelihoods of Model 2 over Model 1, i.e. $B_{21}(t) = \phi_2(1,t)/\phi_1(1,t)$. Suppose B_{21} has been calculated. Then, we can interpret it using the table furnished by Kass and Raftery [3] (a survey paper of Bayes factor) as guideline.

Similarly, we can define B_{12}. Obviously, $B_{12} \times B_{21} = 1$.

Definition 4. Define the filter ratio processes:

$$q_1(f_1, t) = \frac{\phi_1(f_1, t)}{\phi_2(1, t)}, \quad \text{and} \quad q_2(f_2, t) = \frac{\phi_2(f_2, t)}{\phi_1(1, t)}.$$

Remark 5. Observe that the Bayes factors, $B_{12}(t) = q_1(1,t)$ and $B_{21}(t) = q_2(1,t)$.

In summary, $\phi(f,t)$, which determines the likelihood or integrated likelihood, is characterized by the unnormalized filtering equation. $\pi(f,t)$, which determines the posterior, is characterized by the normalized filtering equation, and is the optimum filter in the sense of least mean square error. $q_k(f_k,t)$, $k = 1, 2$, which determines the Bayes factors, is characterized by the system of evolution equations. The filtering equations for $\phi(f,t)$ and $\pi(f,t)$ are derived in [6] and the system of evolution equations for $q_k(f_k,t)$, $k = 1, 2$, is derived in [4].

3.3. Consistency of the likelihoods, posterior and Bayes factors

To compute them for statistical inference, one constructs recursive algorithms to approximate them. Zeng [6,4] applied Markov chain approximation methods to construct recursive algorithms for computing the posterior and Bayes estimates (the Bayes factors). One basic requirement for the recursive algorithms is consistency: The approximate versions (i.e. the likelihoods, posterior and Bayes factors), computed by the recursive algorithms, must converge to the true ones. The following theorem proved by Theorem 1 and Corollaries 1 and 2 provides the theoretical foundation for consistency.

Let $(\theta_\varepsilon^{(k)}, X_\varepsilon^{(k)})$ be an approximation of $(\theta^{(k)}, X^{(k)})$. Then, we define

$$\vec{Y}_\varepsilon^{(k)}(t) = \begin{pmatrix} N_1(\int_0^t \lambda_1(\theta_\varepsilon^{(k)}(s), X_\varepsilon^{(k)}(s), s)\,\mathrm{d}s) \\ N_2(\int_0^t \lambda_2(\theta_\varepsilon^{(k)}(s), X_\varepsilon^{(k)}(s), s)\,\mathrm{d}s) \\ \vdots \\ N_n(\int_0^t \lambda_n(\theta_\varepsilon^{(k)}(s), X_\varepsilon^{(k)}(s), s)\,\mathrm{d}s) \end{pmatrix}, \tag{4}$$

set $\mathscr{F}_t^{\vec{Y}_\varepsilon^{(k)}} = \sigma(\vec{Y}_\varepsilon^{(k)}(s), 0 \leqslant s \leqslant t)$, and take $L_\varepsilon^{(k)}(t) = L\left(\left(\theta_\varepsilon^{(k)}(s), X_\varepsilon^{(k)}(s), Y_\varepsilon^{(k)}(s)\right)_{0\leqslant s\leqslant t}\right)$. Suppose that $(\theta_\varepsilon^{(k)}, X_\varepsilon^{(k)}, \vec{Y}_\varepsilon^{(k)})$ lives on $(\Omega_\varepsilon^{(k)}, \mathscr{F}_\varepsilon^{(k)}, P_\varepsilon^{(k)})$, with Assumptions 1–5. Then, there also exists a reference measure $Q_\varepsilon^{(k)}$ with similar properties. Next, we define the approximate versions that the recursive algorithms compute.

Definition 5. For $k = 1, 2$, let $\phi_{\varepsilon,k}(f_k, t) = E^{Q_\varepsilon^{(k)}}\left[f_k\left(\theta_\varepsilon^{(k)}(t), X_{\varepsilon_x}^{(k)}(t)\right) L_\varepsilon^{(k)}(t) \mid \mathscr{F}_t^{\vec{Y}_\varepsilon^{(k)}}\right]$, $\pi_{\varepsilon,k}(f_k, t) = E^{P_\varepsilon^{(k)}}\left[f_k\left(\theta_\varepsilon^{(k)}(t), X_{\varepsilon_x}^{(k)}(t)\right) \mid \mathscr{F}_t^{\vec{Y}_\varepsilon^{(k)}}\right]$, $q_{\varepsilon,1}(f_1, t) = \phi_{\varepsilon,1}(f_1, t)/\phi_{\varepsilon,2}(1, t)$ and $q_{\varepsilon,2}(f_2, t) = \phi_{\varepsilon,2}(f_2, t)/\phi_{\varepsilon,1}(1, t)$.

Theorem 2. *Suppose that Assumptions* 1–5 *hold for the models* $(\theta^{(k)}, X^{(k)}, \vec{Y}^{(k)})_{k=1,2}$ *and that Assumptions* 1–5 *hold for the approximate models* $(\theta_\varepsilon^{(k)}, X_\varepsilon^{(k)}, \vec{Y}_\varepsilon^{(k)})_{k=1,2}$. *Suppose* $(\theta_\varepsilon^{(k)}, X_\varepsilon^{(k)}) \Rightarrow (\theta^{(k)}, X^{(k)})$ *as* $\varepsilon \to 0$. *Then, as* $\varepsilon \to 0$, for k = 1,2,

(i) $\vec{Y}_\varepsilon^{(k)} \Rightarrow \vec{Y}^{(k)}$;
(ii) $\phi_{\varepsilon,k}(f, t) \Rightarrow \phi_k(f, t)$;
(iii) $\pi_{\varepsilon,k}(f, t) \Rightarrow \pi_k(f, t)$;
(iv) $(q_{\varepsilon,1}(f_1, t), q_{\varepsilon,2}(f_2, t)) \Rightarrow (q_1(f_1, t), q_2(f_2, t))$ *for all bounded continuous functions,* f, f_1, *and* f_2, *and* $t > 0$.

Proof. To simplify notation, we exclude the superscript, (k). Since $(\theta_\varepsilon, X_\varepsilon) \Rightarrow (\theta, X)$, continuous mapping theorem implies that the stochastic intensities of $\vec{Y}_\varepsilon$ converges weakly to that of $\vec{Y}$. This implies $\vec{Y}_\varepsilon \Rightarrow \vec{Y}$. Note that Eq. (3) implies that $Q_k\{E^{Q_k}[L_k(X_k, Y_k) \mid Y_k] = 0\} = 0$ for $k = 1, 2$. To show parts (ii)–(iv), since (by Remark 1) under the reference measure Q_ε, $(\theta_\varepsilon, X_\varepsilon)$ and $\vec{Y}_\varepsilon$ are independent, it suffices to show $((\theta_\varepsilon, X_\varepsilon), \vec{Y}_\varepsilon, L_\varepsilon) \Rightarrow ((\theta, X), \vec{Y}, L)$ under the reference measures. The Radon–Nikodym derivative $\mathrm{d}P_\varepsilon/\mathrm{d}Q_\varepsilon$ is

$$L_\varepsilon(t) = \prod_{j=1}^n \exp\left\{\int_0^t \log\,\lambda_j(\theta_\varepsilon(s-), X_\varepsilon(s-), s-)\,\mathrm{d}Y_{\varepsilon,j}(s) - \int_0^t [\lambda_j(\theta_\varepsilon(s), X_\varepsilon(s), s) - 1]\,\mathrm{d}s\right\}.$$

Noting $(\theta_\varepsilon, X_\varepsilon, \vec{Y}_\varepsilon) \Rightarrow (\theta, X, \vec{Y})$ under the reference measures, one finds that continuous mapping theorem and *Kurtz and Protter's Theorem* [5] imply that

$$\int_0^t \log \lambda_j(\theta_\varepsilon(s-), X_\varepsilon s-), s-)\, \mathrm{d}Y_{\varepsilon,j}(s) \Rightarrow \int_0^t \log \lambda_j(\theta(s-), X(s-), s-)\, \mathrm{d}Y_k(s),$$

$$\int_0^t [\lambda_j(\theta_\varepsilon(s), X_\varepsilon(s), s) - 1]\, \mathrm{d}s \Rightarrow \int_0^t [\lambda_j(\theta(s), X(s), s) - 1]\, \mathrm{d}s,$$

and $((\theta_\varepsilon, X_\varepsilon), \vec{Y}_\varepsilon, L_\varepsilon) \Rightarrow ((\theta, X), \vec{Y}, L)$ under the reference measures (condition C2.2(i) of Kurtz and Protter [5] holds in our case, see Example 3.3 there,). Hence, Theorem 1 and Corollaries 1 and 2 imply (ii)–(iv), respectively. □

References

[1] S. Ethier, T. Kurtz, Markov Processes: Characterization and Convergence, Wiley, New York, 1986.

[2] E. Goggin, Convergence in distribution of conditional expectations, Ann. Probab. 22 (1994) 1097–1114.

[3] R.E. Kass, A.E. Raftery, Bayes factors and model uncertainty, J. Am. Stat. Assoc. 90 (1995) 773–795.

[4] M. Kouritzin, Y. Zeng, Bayesian model selection via filtering for a class of micro-movement models of asset price. International Journal of Theoretical and Applied Finance, submitted.

[5] T. Kurtz, P. Protter, Weak limit theorems for stochastic integrals and stochastic differential equations, Ann. Probab. 19 (1991) 1035–1070.

[6] Y. Zeng, A partially-observed model for micro-movement of asset prices with bayes estimation via filtering, Math. Financ. 13 (2003) 411–444.

[7] I. Crimaldi, L. Pratelli, Convergence results for conditional expectations, to appear in Bernoulli.