Kernel Weighted Approximate MLE of Parameter Volatility Model Based on log-Likelihood Expansion

Jing Wang
College of Mathematics and Information Science
Henan Normal University, Henan, 453007, China

Ji-Xia Wang
College of Mathematics and Information Science
Henan Normal University, Henan, 453007, China

Abstract

This paper studies the Kernel weighted approximate maximum likelihood estimation in diffusion process model with time-dependent parameter, which is observed at discrete sample paths. Based on the approximate maximum likelihood function, the Kernel weighted approximate maximum likelihood estimation of time-dependent parameter is proposed. The consistency of the proposed estimation is established under two regimes.

Mathematics Subject Classification: Primary 62M05; Secondary 62F12

Keywords: time-dependent parameter; consistency; asymptotic normality; diffusion model; maximum likelihood estimation

1This work is supported by the National Natural Science Foundation (NNSF) of China (No. U1504701); Doctoral research startup project of Henan Normal University (No. qd15184).
1 Introduction

The time-dependent parameter models are important for the dynamic models in many scientific fields. Various efforts include the time-dependent models of the references [1-6].

In this paper, we consider more general diffusion model with time-dependent parameter than the above model, which is defined by

\[ dX(t) = \mu(X_t, \theta_t)dt + \sigma(X_t, \theta_t)dW_t. \]

Here, we use the approximate forms to the transitional densities. The method is discussed in Aït-Sahalia(2002) and Chang and Chen(2011). Aït-Sahalia(2002) has proved that the approximate likelihood function converges to the true likelihood function.

In our model, we suppose that the parameters is a function of time. By using the local constant approximation and maximizing the local approximate kernel weighted log-likelihood function, the kernel weighted local approximate maximum likelihood estimation(AMLE for brief) is proposed. We prove the consistency and gain the rate of convergence of the local AMLE.

The rest of this paper is organized as follows. In Section 2, we give some assumptions and state the transitional density approximations of Aït-Sahalia(2002) and Chang and Chen(2011). The kernel weighted AMLE of time-dependent parameter and some lammas are given in Section 3. Section 4 establishes the consistency and rates of convergence of local AMLE.

2 Transitional density approximation and assumptions

On \((\Omega, F, (F_t)_{t \geq 0}, P)\), we consider a diffusion process \(\{X_t\}_{t \geq 0}\) as the following

\[ dX(t) = \mu(X_t, \theta_t)dt + \sigma(X_t, \theta_t)dW_t \quad (1) \]

where \(W_t\) is the standard Brownian motion, \(\mu(\cdot, \cdot)\) and \(\sigma(\cdot, \cdot)\) are respectively the drift and diffusion functions, which are known functions. \(\theta_t = (\theta_1(t), \theta_2(t), \ldots, \theta_d(t))^T\) is an unknown parameter vector taking values in a compact set \(\Theta \in \mathbb{R}^d\) for a arbitrary given time point \(t_0\).

Let \(f_{X}(x| x_0, \delta; \theta_t)\) be the density of \(X_{t+\delta}\) given \(X_t = x_0\) for \((x, x_0) \in \chi \times \chi\), where \(\delta\) is the sampling internal and \(\chi\) is the domain of \(X_t\).

The first step is transforming \(X_t\) to \(Y_t\) defined as

\[ Y_t = \gamma(X_t, \theta_t) \equiv \int_{X_t}^{\delta} \frac{du}{\sigma(u; \theta_t)} \quad (2) \]
By applying Itô formula, we have
\[ dY_t = \mu_Y(Y_t; \theta_t)dt + dW_t \]
where \( \mu_Y(Y_t; \theta_t) = \frac{\mu(y^{-1}(y; \theta_t)\gamma_1 - \frac{1}{2} \frac{\partial \sigma}{\partial x}(y^{-1}(y; \theta_t))}{\sigma(y; \theta_t)} \).

Let \( f_Y(y|y_0, \delta; \theta_t) \) be the density of \( Y_{t+\delta} \) given \( Y_t = y_0 \). Then,
\[ f_X(x_t|x_{t-1}, \delta; \theta_t) = \frac{f_Y(\gamma(x_t; \theta_t)\gamma(x_{t-1}; \theta_t), \delta; \theta_t)}{\sigma(x_t; \theta_t)} \]
(3)

The second step is standardizing \( Y_{t+\delta} \) by \( Z_t = \delta^{-\frac{1}{2}}(Y_{t+\delta} - y_0) \). Let \( f_Z(z|y_0, \delta; \theta_t) \) be the density of \( Z_{t+\delta} \) given \( Z_t = 0 \). Then
\[ f_Z(z|y_0, \delta; \theta_t) = \delta^{\frac{1}{2}} f_Y(\delta^{\frac{1}{2}}z + y_0|y_0, \delta; \theta_t). \]

Now, the Hermite polynomial is defined by
\[ H_j(z) = \phi^{-1}(z)(-1)^j \frac{d^j \phi(z)}{dz^j}, j = 0, 1, 2, \ldots \]
where \( \phi \) is the standard normal density function and \( \{H_j(z)\}_{j=0}^{\infty} \) are orthogonal with respect to \( \phi \), namely \( \int H_j(x)H_k(x)\phi(x)dx = 0 \) for \( j \neq k \). So, the Hermite expansion of the density function \( f_Z(z|y_0, \delta; \theta_t) \) is
\[ f_Z^H(z|y_0, \delta; \theta_t) = \phi(z) \sum_{j=0}^{\infty} \eta_j(y_0, \delta; \theta_t)H_j(z) \]
(4)
where
\[ \eta_j(y_0, \delta; \theta_t) = (j!)^{-1} \int H_j(z)f_Z(z|y_0, \delta; \theta_t)dz = (j!)^{-1} \int H_j(\delta^{-\frac{1}{2}}(y-y_0))f_Y(y|y_0, \delta; \theta_t)dy \]
\[ = (j!)^{-1}E[H_j(\delta^{-\frac{1}{2}}(Y_{t+\delta} - y_0))|Y_t = y_0]. \]

Define the infinitesimal generator of \( Y_t \) as the following
\[ A_{\theta_t}g(y) = \mu_Y(y; \theta_t) \frac{\partial g}{\partial y} + \frac{1}{2} \frac{\partial^2 g}{\partial y^2} \]
(5)

Then,
\[ E[H_j(\delta^{-\frac{1}{2}}(Y_{t+\delta} - y_0))|Y_t = y_0] \]
\[ = \sum_{k=0}^{\infty} A_{\theta_t}^k H_j(\delta^{-\frac{1}{2}}(y-y_0))|y = y_0 \frac{\delta^k}{k!} + E[A_{\theta_t}^{k+1}H_j(\delta^{-\frac{1}{2}}(Y_{t+\delta} - y_0))|Y_t = y_0] \frac{\delta^{k+1}}{(k + 1)!}. \]
(6)
Jing Wang and Ji-Xia Wang

So, a J-term approximation to the density \( f_Y(y|y_0, \delta; \theta_t) \) is

\[
f_Y^{(J)}(y|y_0, \delta; \theta_t) = \delta^{-\frac{1}{2}} \phi\left(\frac{y-y_0}{\delta}\right) \exp\left(\int_{y_0}^{y} \mu_Y(u; \theta_t) du\right) \sum_{j=0}^{J} c_j(y|y_0; \theta_t) \delta^j / j!
\]

where \( c_0(y|y_0; \theta_t) \equiv 1 \) and for \( j \geq 1, \)

\[
c_j(y|y_0; \theta_t) = j(y-y_0)^{-j} \int_{y_0}^{y} (\omega-y_0)^{j-1} \cdot \left\{ \lambda_Y(\omega; \theta_t) c_{j-1}(\omega|y_0; \theta_t) + \frac{1}{2} \frac{\partial^2 c_{j-1}(\omega|y_0; \theta_t)}{\partial \omega^2} \right\} d\omega
\]

here \( \lambda_Y(y; \theta_t) = -\{\mu_Y^2(y; \theta_t) + \partial \mu_Y(y; \theta_t)/\partial y\} / 2. \)

Now, we have

\[
f_X^{(J)}(x|x_0, \delta; \theta_t) = \sigma^{-1}(x; \theta_t) \delta^{-\frac{1}{2}} \phi(A) \cdot \exp\left(\int_{x_0}^{x} \frac{\mu_Y(y; \theta_t)}{\sigma(y; \theta_t)} du\right)
\]

\[
\cdot \sum_{j=0}^{J} c_j(\gamma(x; \theta_t)|\gamma(x_0; \theta_t); \theta_t) \delta^j / j!, \tag{7}
\]

where \( A = \frac{\gamma(x; \theta_t)-\gamma(x_0; \theta_t)}{\delta^2} \). So, we have the following

\[
f_X^{(J)}(x|x_0, \delta; \theta_{t_0}) \to f_X(x|x_0, \delta; \theta_{t_0}) \tag{8}
\]

uniformly with respect to \( \theta_{t_0} \in \Theta \) and \( x_0 \) over compact of \( \chi. \)

Now we give the following assumptions. \( C \) denotes a positive constant.

(A1) The parameter vector \( \theta_t \) is twice continuously differentiable in \( t. \)

(A2) The kernel function \( K(\cdot) \) is symmetric and \( |K(x)| \leq C \) for all \( x \in \mathbb{R}. \)

(A3) For a given time point \( t_0, \) the parameter \( \theta_{t_0} \) is an interior point of \( \Theta; \)

(A4) (1) For a given time point \( t_0 \) and each \( \delta > 0, \)

\[
E\left[ \frac{\partial \log f(X_t|X_{t-1}; \delta; \theta_{t_0})}{\partial \theta} \right] = 0
\]

and \( \theta_{t_0} \) is the root of \( E\left[ \frac{\partial \log f(X_t|X_{t-1}; \delta; \theta)}{\partial \theta} \right] = 0. \)

(11) For a given time point \( t_0, \) the MLE \( \hat{\theta}_{t_0,n} \) is consistent to \( \theta_{t_0} \) and asymptotically normal.

3 Kernel weighted approximate maximum likelihood estimation

Considering that the form of the parameter vector \( \theta_t \) is not specified and is smooth according to Assumption (A1), the parameter vector can be locally
approximated by a constant. That is, at a given time point $t_0$, we use the approximation $\hat{\theta}_t = \theta_{t_0} = (\theta_1(t_0), \theta_2(t_0), \ldots, \theta_d(t_0))^T$, for $t$ in a small neighborhood of $t_0$. Let $h$ denote the size of the neighborhood and $K(\cdot)$ be a nonnegative weight function. These are called a bandwidth parameter and a kernel function, respectively.

Let $\{X_t, i = 1, 2, \cdots, n\}$ be discretely observed sample of the diffusion process $\{X_t\}_{t \geq 0}$, where $t_i = i\delta, i = 1, 2, \cdots, n$. We use $f$ and $f^{(J)}$ to denote $f_X$ and $f_X^{(J)}$ respectively. Based on the (7), the $ith$ approximate log-density of $X_{t_i}$ given $X_{t_{i-1}}$ is

$$\log f^{(J)}(x_{t_i} | x_{t_{i-1}}, \delta; \theta_{t_0}) = - \log \sqrt{2\pi \delta} + A_1(x_{t_i} | x_{t_{i-1}}, \delta; \theta_{t_0}) + A_2(x_{t_i} | x_{t_{i-1}}, \delta; \theta_{t_0})$$

where

$$A_1(x_{t_i} | x_{t_{i-1}}, \delta; \theta_{t_0}) = - \log [\sigma(x_{t_i}, \theta_{t_0})] - \frac{1}{2 \delta} [\gamma(x_{t_i}, \theta_{t_0}) - \gamma(x_{t_{i-1}}, \theta_{t_0})]^2$$

$$A_2(x_{t_i} | x_{t_{i-1}}, \delta; \theta_{t_0}) = \int_{x_{t_{i-1}}}^{x_{t_i}} \frac{\mu(Y(\gamma(u, \theta_{t_0}); \theta_{t_0}))}{\sigma(u, \theta_{t_0})} du$$

and

$$A_3(x_{t_i} | x_{t_{i-1}}, \delta; \theta_{t_0}) = \log \left\{ \sum_{j=0}^{J} \left[ c_j(\gamma(x_{t_i}; \theta_{t_0})|\gamma(x_{t_{i-1}}; \theta_{t_0}); \theta_{t_0})^j/j! \right] \right\}$$

By using the local constant approximation and introducing the kernel weight, we obtain, at an arbitrary given time point $t_0$, the local approximate kernel weighted log-likelihood function

$$\ell_n^{(J)}(\theta_{t_0}, t_0) = \sum_{i=0}^{n} K_h(t_i - t_0) \log f^{(J)}(x_{t_i} | x_{t_{i-1}}, \delta; \theta_{t_0})$$

$$= - \log \sqrt{2\pi \delta} \sum_{i=0}^{n} K_h(t_i - t_0) + \sum_{i=0}^{n} K_h(t_i - t_0) A_1(x_{t_i} | x_{t_{i-1}}, \delta; \theta_{t_0})$$

$$+ \sum_{i=0}^{n} K_h(t_i - t_0) A_2(x_{t_i} | x_{t_{i-1}}, \delta; \theta_{t_0}) + \sum_{i=0}^{n} K_h(t_i - t_0) A_3(x_{t_i} | x_{t_{i-1}}, \delta; \theta_{t_0})$$

where $K_h(\cdot) = K(\cdot/h)/h$. Now, we obtain the local AMLE of $\theta_{t_0}$, that is

$$\hat{\theta}_{t_0, n}^{(J)} = \arg \max_{\theta \in \Theta} \ell_n^{(J)}(\theta, t_0).$$

The whole parameter vector $\theta_t$ can be estimated by repeatedly maximizing (9) over a grid of time points. This method is related to the generalized method of moments of Hansen(1982) and is used in a local neighborhood[see Fan(1998), Florens-Zmirou(1993) and Genon-Catalot and Jacod(1993)].
Now, we select the bandwidth of the local AMLE $\hat{\theta}_{t_0,n}^{(j)}$. For an arbitrary given time point $t_0$, the bandwidth can be chosen to minimize the approximate log-likelihood function:

$$\sum_{i=0}^{n} \log f^{(j)}(x_{t_i}|x_{t_{i-1}},\delta;\theta_{t_0,n}^{(j)}) = -n \log \sqrt{2\pi \delta}$$

$$+ \sum_{i=0}^{n} A_1(x_i|x_{t_{i-1}},\delta;\theta_{t_0,n}^{(j)}) + \sum_{i=0}^{n} A_2(x_i|x_{t_{i-1}},\delta;\theta_{t_0,n}^{(j)}) + \sum_{i=0}^{n} A_3(x_i|x_{t_{i-1}},\delta;\theta_{t_0,n}^{(j)})$$

This idea has also been used by Fan et al. (2003).

**Lemma 3.1** Under Assumptions (A3) and (A6), for any $\delta \in (0, \Delta]$, we have

$$\sum_{l=0}^{\infty} c_l(\gamma(x_{t_i};\theta_0)|\gamma(x_{t_{i-1}};\theta_0);\theta_0)\delta^l/l!$$

converges with probability 1, and for $k = 1, 2, \text{ and } 3$, and $i, j, k \in \{1, 2, \cdots, d\}$,

$$\frac{\partial^k}{\partial \theta_1 \partial \theta_2 \cdots \partial \theta_k} \sum_{l=0}^{\infty} c_l(\gamma(x_{t_i};\theta_0)|\gamma(x_{t_{i-1}};\theta_0);\theta_0)\delta^l/l!$$

$$= \sum_{l=0}^{\infty} \frac{\partial^k}{\partial \theta_1 \partial \theta_2 \cdots \partial \theta_k} c_l(\gamma(x_{t_i};\theta_0)|\gamma(x_{t_{i-1}};\theta_0);\theta_0)\delta^l/l!.$$

**Lemma 3.2** Under Assumptions (A7) and (A8), for any $\beta > 1$, there exists $m(\beta) < \infty$ and $\Delta_1(\beta)$ such that for any $\delta \in (0, \Delta_1(\beta)]$ and $J$, then

$$E\{ \sup_{\theta_0 \in \Theta} | \sum_{l=0}^{J} c_l(\gamma(x_{t_i};\theta_0)|\gamma(x_{t_{i-1}};\theta_0);\theta_0)\delta^l/l!|^{-\beta} \} \leq m(\beta).$$

The proofs of lemma 3.1 and lemma 3.2 are given in Chang and Chen (2011).

**Lemma 3.3** Under Assumptions (A3), (A5)-(A8), there exist $M < \infty$ and $\Delta_2 > 0$ such that, for any $J, \delta \in (0, \Delta_2]$ and $s, j, k \in \{1, 2, \cdots, d\},$

$$E\{ \sup_{\theta_0 \in \Theta} | \frac{\partial^3 K_h(t_i - t_0) \log f^{(j)}(X_{t_i}|X_{t_{i-1}},\delta;\theta_0)}{\partial \theta_s \partial \theta_j \partial \theta_k} | \} \leq M.$$

**Proof** Note that

$$| \frac{\partial^3 K_h(t_i - t_0) A_l(x_i|x_{t_{i-1}},\delta;\theta_0)}{\partial \theta_s \partial \theta_j \partial \theta_k} | \leq |K_h(t_i - t_0)|| \frac{\partial^3 A_l(x_i|x_{t_{i-1}},\delta;\theta_0)}{\partial \theta_s \partial \theta_j \partial \theta_k} |$$

where $l = 1, 2, 3$. By using Assumptions (A2), (A5) and lemma 3 in Chang and Chen (2011), the result of lemma 3.3 is obtained. □
4 Consistency

For a given time point \( t_0 \), the maximum likelihood estimation is defined as

\[
\hat{\theta}_{t_0, n} = \arg\max_{\theta \in \Theta} \sum_{i=0}^{n} K_h(t_i - t_0) \log f(x_{t_i}|x_{t_{i-1}}, \delta; \theta_t).
\]

We have

\[
\sum_{i=0}^{n} K_h(t_i - t_0) \frac{\partial}{\partial \theta} \log f(x_{t_i}|x_{t_{i-1}}, \delta; \hat{\theta}_{t_0, n})
\]

\[
= \sum_{i=0}^{n} K_h(t_i - t_0) \frac{\partial}{\partial \theta} \log f(x_{t_i}|x_{t_{i-1}}, \delta; \hat{\theta}_{t_0, n}) = 0 \quad \text{(10)}
\]

Subtracting \( \sum_{i=0}^{n} K_h(t_i - t_0) \frac{\partial}{\partial \theta} \log f(x_{t_i}|x_{t_{i-1}}, \delta; \theta_0) \) from both sides of (10), we obtain

\[
\sum_{i=0}^{n} K_h(t_i - t_0) \frac{\partial}{\partial \theta} \log f(x_{t_i}|x_{t_{i-1}}, \delta; \hat{\theta}_{t_0, n}) - \sum_{i=0}^{n} K_h(t_i - t_0) \frac{\partial}{\partial \theta} \log f(x_{t_i}|x_{t_{i-1}}, \delta; \theta_0) \]

\[
= \sum_{i=0}^{n} K_h(t_i - t_0) \frac{\partial}{\partial \theta} \log f(x_{t_i}|x_{t_{i-1}}, \delta; \hat{\theta}_{t_0, n}) - \sum_{i=0}^{n} K_h(t_i - t_0) \frac{\partial}{\partial \theta} \log f(x_{t_i}|x_{t_{i-1}}, \delta; \theta_0)
\]

\[
= \sum_{i=0}^{n} K_h(t_i - t_0) \frac{\partial}{\partial \theta} [\tilde{A}_3(x_{t_i}|x_{t_{i-1}}, \delta; \theta_t) - A_3(x_{t_i}|x_{t_{i-1}}, \delta; \theta_0)]
\]

\[
+ \sum_{i=0}^{n} K_h(t_i - t_0) \frac{\partial}{\partial \theta} \log f(x_{t_i}|x_{t_{i-1}}, \delta; \hat{\theta}_{t_0, n}) - \sum_{i=0}^{n} K_h(t_i - t_0) \frac{\partial}{\partial \theta} \log f(x_{t_i}|x_{t_{i-1}}, \delta; \theta_0)
\]

Now, we have

\[
\frac{1}{n} \sum_{i=0}^{n} K_h(t_i - t_0) \frac{\partial^2}{\partial \theta \partial \theta^T} \log f(x_{t_i}|x_{t_{i-1}}, \delta; \theta_0)(\hat{\theta}_{t_0, n} - \theta_0) +
\]

\[
\frac{1}{2} [E_d \otimes (\hat{\theta}_{t_0, n} - \theta_0)^T] \frac{1}{n} \sum_{i=0}^{n} K_h(t_i - t_0) \frac{\partial^3}{\partial \theta \partial \theta^T \partial \theta} \log f(x_{t_i}|x_{t_{i-1}}, \delta; \theta^*)(\hat{\theta}_{t_0, n} - \theta_0)
\]

\[
= \frac{1}{n} \sum_{i=0}^{n} K_h(t_i - t_0) \frac{\partial}{\partial \theta} [\tilde{A}_3(x_{t_i}|x_{t_{i-1}}, \delta; \theta_t) - A_3(x_{t_i}|x_{t_{i-1}}, \delta; \theta_0)]
\]

\[
+ \frac{1}{n} \sum_{i=0}^{n} K_h(t_i - t_0) \frac{\partial^2}{\partial \theta \partial \theta^T} \log f(x_{t_i}|x_{t_{i-1}}, \delta; \theta_0)(\hat{\theta}_{t_0, n} - \theta_0)
\]
\[ + \frac{1}{2} [E_d \otimes (\hat{\theta}_{t_0,n} - \hat{\theta}_t)^T ] \frac{1}{n} \sum_{i=0}^{n} K_h(t_i - t_0) \frac{\partial^3}{\partial \theta \partial \theta^T \partial \theta} \log f(x_t|x_{t_{i-1}}, \delta; \theta^{**})(\hat{\theta}_{t_0,n} - \hat{\theta}_t) \]

(12)

where \( \theta^* \) is a parameter which is on the line between \( \hat{\theta}^{(j)}_{t_{0,n}} \) and \( \theta_t \) and \( \theta^{**} \) is is a parameter which is on the line between \( \hat{\theta}_{t_{0,n}} \) and \( \theta_t \). \( E_d \) is the \( d \times d \) identity matrix.

Let \( \Delta_{n1}(\hat{\theta}^{(j)}_{t_{0,n}}, \theta_t) \) and \( \Delta_{n2}(\hat{\theta}_{t_{0,n}}, \theta_t) \) denote the last terms on the left and the right hand sides of (12), respectively. Furthermore, let

\[
F_n(\theta_t, J, \delta) = \frac{1}{n} \sum_{i=0}^{n} K_h(t_i - t_0) \frac{\partial^2}{\partial \theta \partial \theta^T} [\hat{A}_3(x_t|x_{t_{i-1}}, \delta; \theta_t) - A_3(x_t|x_{t_{i-1}}, \delta; \theta_t)]
\]

\[
U_n(\theta_t, J, \delta) = \frac{1}{n} \sum_{i=0}^{n} K_h(t_i - t_0) \frac{\partial}{\partial \theta} [\hat{A}_3(x_t|x_{t_{i-1}}, \delta; \theta_t) - A_3(x_t|x_{t_{i-1}}, \delta; \theta_t)]
\]

and

\[
N_n(\theta_t, J, \delta) = \frac{1}{n} \sum_{i=0}^{n} K_h(t_i - t_0) \frac{\partial^2}{\partial \theta \partial \theta^T} \log f^{(j)}(x_t|x_{t_{i-1}}, \delta; \theta_t).
\]

Then, (12) can be written as

\[
N_n(\theta_t, J, \delta)(\hat{\theta}_{t_{0,n}} - \hat{\theta}_t) + \Delta_{n1}(\hat{\theta}^{(j)}_{t_{0,n}}, \theta_t) = U_n(\theta_t, J, \delta)
\]

\[
+ [N_n(\theta_t, J, \delta) + F_n(\theta_t, J, \delta)](\hat{\theta}_{t_{0,n}} - \hat{\theta}_t) + \Delta_{n2}(\hat{\theta}_{t_{0,n}}, \theta_t)
\]

(13)

Now, we study the asymptotic propositions as the following. Let \( \|A\|_2 = [\rho(A^T A)]^{\frac{1}{2}} \) be the norm of a matrix \( A \), where \( \rho(A^T A) \) denotes the largest eigenvalue of \( A^T A \).

From the lemma 3.3, we have

\[ \Delta_{n1}(\hat{\theta}^{(j)}_{t_{0,n}}, \theta_t) = O_p(\|\hat{\theta}^{(j)}_{t_{0,n}} - \theta_t\|_2^2) \] and \( \Delta_{n2}(\hat{\theta}_{t_{0,n}}, \theta_t) = O_p(\|\hat{\theta}_{t_{0,n}} - \theta_t\|_2^2) \).

\[ E[F_n(\theta_t, J, \delta)], \ E[U_n(\theta_t, J, \delta)] \] and \( E[N_n(\theta_t, J, \delta)] \) exist. Let \( N(\theta_t, J, \delta) = E[N_n(\theta_t, J, \delta)] \). We think that

\[
N(\theta_t, J, \delta) \rightarrow - \frac{1}{n} \sum_{i=1}^{n} K_h(t_i - t_0) I(\delta)
\]

as \( J \rightarrow \infty \) for fixed \( \delta \) or \( J \) being fixed but \( \delta \rightarrow 0 \). The following proposition give the speed of the convergence.

**Proposition 4.1** Under the Assumptions (A2)-(A3),(A6)-(A8), there exist a constant \( C \), which is not dependent on \( J \) and \( \delta \), and a \( \Delta > 0 \) such that

\[
\|N(\theta_t, J, \delta) + \frac{1}{n} \sum_{i=1}^{n} K_h(t_i - t_0) I_i(\delta)\|_2 \leq C \delta^{J+1}.
\]
where \( I_i(\delta) \equiv -E \left[ \frac{\partial^2 \log f(X_{t_i} \mid X_{t_{i-1}}, \delta; \theta_{t_0})}{\partial \theta_{t_0} \partial \theta_{t_0}} \right] \)

**Proposition 4.2** If Assumptions (A2)-(A3),(A5)-(A9) hold, we have

\[
\| N^{-1}(\theta_{t_0}, J, \delta) \frac{1}{n} \sum_{i=1}^{n} K_h(t_i - t_0) I_i(\delta) + E_d \|_2 = O(\delta^J)
\]

and

\[
\| N^{-1}(\theta_{t_0}, J, \delta) U(\theta_{t_0}, J, \delta) \|_2 = O(\delta^J)
\]

under either (1) for any fixed \( \delta \in (0, \bar{\Delta}] \), where \( \bar{\Delta} \) is the quantity in Proposition 4.1 and \( J \to \infty \) or (2) for any fixed \( J \), and \( \delta \to 0 \).

The proof of the above propositions are given in Appendix.

**Proposition 4.3** Under the Assumptions (A2)-(A3),(A6)-(A8), there exist constants \( \tilde{\Delta} \) and \( C \), which is not dependent on \( J \) and \( \delta \), such that

\[
E \left\{ \sup_{\theta_{t_0} \in \Theta} \left\| \frac{1}{n} \sum_{i=1}^{n} K_h(t_i - t_0) \frac{\partial}{\partial \theta} \log f(X_{t_i} \mid X_{t_{i-1}}, \delta; \theta_{t_0}) \right. \right. - \left. \left. \frac{1}{n} \sum_{i=1}^{n} K_h(t_i - t_0) \frac{\partial}{\partial \theta} \log f^{(J)}(X_{t_i} \mid X_{t_{i-1}}, \delta; \theta_{t_0}) \right\|_2 \right\} \leq C\delta^{J+1}.
\]

**Proof** We can use the same method in the proof of Proposition 4.1.

**Theorem 4.1** Under the Assumptions (A1)-(A8), for a arbitrary given time point \( t_0 \),

\[ \hat{\theta}^{(j)}_{t_{0,n}} \to \theta_{t_0} \]

in probability under (1) \( \delta \in (0, \bar{\Delta} \wedge \bar{\Delta}] \) being fixed, \( J \to \infty \) and \( n \to \infty \) or (2) \( J \) being fixed, \( n \to \infty \), \( \delta \to 0 \), \( n\delta \to \infty \).

**Proof**

From Proposition 4.3, we have

\[
\sup_{\theta_{t_0} \in \Theta} \left\| \frac{1}{n} \sum_{i=1}^{n} K_h(t_i - t_0) \frac{\partial}{\partial \theta} \log f(X_{t_i} \mid X_{t_{i-1}}, \delta; \theta_{t_0}) \right. \right. - \left. \left. \frac{1}{n} \sum_{i=1}^{n} K_h(t_i - t_0) \frac{\partial}{\partial \theta} \log f^{(J)}(X_{t_i} \mid X_{t_{i-1}}, \delta; \theta_{t_0}) \right\|_2 \to 0
\]

in probability for (1) \( \delta \in (0, \bar{\Delta}] \) being fixed, \( n \to \infty \) or (2) \( n \to \infty \), \( \delta \to 0 \) but \( n\delta \to \infty \).

We obtain

\[
\left\| \frac{1}{n} \sum_{i=1}^{n} K_h(t_i - t_0) E \left[ \frac{\partial}{\partial \theta} \log f(X_{t_i} \mid X_{t_{i-1}}, \delta; \hat{\theta}^{(j)}_{t_{0,n}}) \right] \right\|_2
\]
\[
\leq \sup_{\theta_0 \in \Theta} \left\| \frac{1}{n} \sum_{i=1}^{n} K_h(t_i - t_0) \frac{\partial}{\partial \theta} \log f(X_{t_i} | X_{t_{i-1}}, \delta; \theta_0) - \frac{1}{n} \sum_{i=1}^{n} K_h(t_i - t_0) E\left[ \frac{\partial}{\partial \theta} \log f(X_{t_i} | X_{t_{i-1}}, \delta; \theta_0) \right] \right\|_2
\]

\[
+ \sup_{\theta_0 \in \Theta} \left\| \frac{1}{n} \sum_{i=1}^{n} K_h(t_i - t_0) \frac{\partial}{\partial \theta} \log f^{(j)}(X_{t_i} | X_{t_{i-1}}, \delta; \theta_0) \right\|_2
\]

\[
\to 0
\]

in probability under (1) $\delta \in (0, \Delta \wedge \tilde{\Delta}]$ being fixed, $J \to \infty$ and $n \to \infty$ or (2) $J$ being fixed, $n \to \infty$, $\delta \to 0$ but $n\delta \to \infty$. Hence, From Assumption (A4), we can obtain the consistency of the $\hat{\theta}_{t_0,n}$.

This completes the proof. $\square$

References


Received: February 8, 2018; Published: March 6, 2018