Parametric estimation for volatility of ergodic diffusion process with unspecified drift

Yoichi Nishiyama

The Institute of Statistical Mathematics

10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan

nisiyama@ism.ac.jp

http://www.ism.ac.jp/~nisiyama/

October 8, 2009

Running title: Estimation for volatility

Abstract

We consider a statistical model of one-dimensional ergodic diffusion processes where the diffusion coefficient contains a finite-dimensional parameter of interest and the drift coefficient is not specified. We constract an estimator for the diffusion coefficient not using any information on the drift coefficient, and prove its asymptotic normality. The limit distribution is the same as that of the LAN theory in the case where the drift coefficient is known, and thus our estimator is asymptotically optimal.

Keywords. Estimating function, asymptotic efficiency, ergodic diffusion, discrete observation.

MSC Classification. Primary, 62F12; secondary, 62M05.

1 Introduction

Let us begin with stating our motivation. First consider the one-dimensional ergodic diffusion process X which is a solution to the stochastic differential equation (SDE)

$$X_t = X_0 + \int_0^t S(X_s;\theta)ds + \int_0^t \sigma(X_s;\theta)dW_s,$$

where $s \rightsquigarrow W_s$ is a standard Brownian motion. Here θ is a parameter from a finite-dimensional space $\Theta \subset \mathbb{R}^d$. The problem of estimating the true value $\theta_0 \in \Theta$ based on high frequency data $\{X_{t_i^n}, i = 0, 1, ..., n\}$ where $0 = t_0^n < t_1^n < \cdots < t_n^n$ are the time points such that $t_n^n \to \infty$ and $\Delta_n = \max_{1\le i\le n} |t_i^n - t_{i-1}^n| \to 0$ with an appropriate rate has been considered by many authors; see Florens-Zmirou (1989), Yoshida (1992) and Kessler (1997) among others. However, any parametric models for the drift coefficient $\{S(\cdot;\theta); \theta \in \Theta\}$ and the diffusion coefficient $\{\sigma(\cdot;\theta); \theta \in \Theta\}$ are exposed at a risk of missspecification. So if we are interested only in estimating the drift coefficient, it is preferable not to specify any model for the diffusion coefficient. Recently, Nishiyama (2009a, b) considered this problem by a semiparametric approach. He treated the diffusion coefficient as an infinite-dimensional nuisance parameter, and showed that having a consistent estimator for the nuisance parameter is sufficient for obtaining an asymptotically optimal estimator for the finite-dimensional parameter of the drift coefficient.

This paper considers the opposite problem. Suppose that we are interested only in estimating the diffusion coefficient, and we do not specify any parametric structure of the drift coefficient. Namely, we consider the SDE

$$X_t = X_0 + \int_0^t S(X_s)ds + \int_0^t \sigma(X_s;\theta)dW_s,$$

and our goal is to estimate the parameter θ of the diffusion coefficient under the assumptions that

$$t_n^n \to \infty, \quad n\Delta_n^2 \to 0 \tag{1}$$

and

$$\sum_{i=1}^n \left| \frac{|t_i^n - t_{i-1}^n|}{t_n^n} - \frac{1}{n} \right| \to 0. \tag{2}$$

The assumption (2) is satisfied for the equidistant sampling scheme $t_i^n = i\Delta_n$. Although our result does not cover those of Yoshida (1992) and Kessler (1997) at least because our assumption $n\Delta_n^2 \to 0$ is stronger than theirs, the point of our work is that we do not use any information on the drift coefficient when we construct our estimator.

An approximate likelihood is

$$L_n(S,\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma(X_{t_{i-1}^n};\theta)^2|t_i^n - t_{i-1}^n|}} \exp\left(-\frac{|X_{t_i^n} - X_{t_{i-1}^n} - S(X_{t_{i-1}^n})|t_i^n - t_{i-1}^n||^2}{2\sigma(X_{t_{i-1}^n};\theta)^2|t_i^n - t_{i-1}^n|} \right).$$

We will see that the influence of the drift is negligible if we assume $n\Delta_n^2 \to 0$. Based on this fact, by putting $S = 0$, we propose to use the estimating function

$$\begin{aligned}\Psi_n(\theta) &= \frac{1}{n}\left(\frac{\partial}{\partial\theta_1}\log L_n(0,\theta), \cdots, \frac{\partial}{\partial\theta_d}\log L_n(0,\theta) \right)^T \\ &= \frac{1}{n}\sum_{i=1}^n \frac{\dot\sigma(X_{t_{i-1}^n};\theta)}{\sigma(X_{t_{i-1}^n};\theta)^3}\left\{ \frac{|X_{t_i^n} - X_{t_{i-1}^n}|^2}{|t_i^n - t_{i-1}^n|} - \sigma(X_{t_{i-1}^n};\theta)^2 \right\},\end{aligned}$$

where $\dot{\sigma}(x;\theta) = (\frac{\partial}{\partial\theta_1}\sigma(x;\theta), \cdots, \frac{\partial}{\partial\theta_d}\sigma(x;\theta))^T$. Our estimator $\widehat{\theta}_n$ is an (approximate) solution to the estimating equation $\Psi_n(\theta) \approx 0$. We will prove that, under some regularity conditions, the rescaled residual $n^{1/2}(\widehat{\theta}_n - \theta_0)$ converges weakly to $N(0, I(S,\theta_0)^{-1})$ with

$$I(S,\theta_0) = 2\int_{\mathbb{R}} \frac{\dot{\sigma}(x;\theta_0)\dot{\sigma}(x;\theta_0)^T}{\sigma(x;\theta_0)^2}\mu_{S,\theta_0}(dx) \tag{3}$$

where μ_{S,θ_0} denotes the invariant law. The limit distribution is the same as that in the LAN theory developed by Gobet (2002) in the case where the drift coefficient S is *known*, and therefore we can conclude that our estimator is asymptotically optimal. A related work was done by Genon-Catalot and Jacod (1993) who considered the case $t_n^n \equiv 1$, and the limit is a mixed Gaussian distribution. As it is announced above, our result is a consequence of the fact that the influence of the drift is negligible even in the case $t_n^n \to \infty$ if we assume $n\Delta_n^2 \to 0$. (This fact itself is known to some researchers, for example, Iacus *et al.* (2009).)

We present our main result in Section 2, and the proofs of some key lemmas are given in Section 3. We mean by the phrase "the function f on $\mathbb{R}$ has a polynomial majorant" that there exists a constant $C_f > 0$ such that $|f(x)| \leq C_f(1+|x|)^{C_f}$. The d-dimensional Euclidean norm is denoted by $||\cdot||$.

2 Result

First let us list up some conditions on the drift and diffusion coefficients.

A1. Θ is a bounded subset of $\mathbb{R}^d$, on which the function $\theta \mapsto \sigma(x;\theta)$ is differentiable with the derivative $\dot{\sigma}(x;\theta) = (\frac{\partial}{\partial\theta_1}\sigma(x;\theta), \cdots, \frac{\partial}{\partial\theta_d}\sigma(x;\theta))^T$, and there exists a measurable function Λ on $\mathbb{R}$ with a polynomial majorant such that

$$\sigma(x;\theta)^2 - \sigma(x;\theta_0)^2 = 2\sigma(x;\theta_0)\dot{\sigma}(x;\theta_0)^T(\theta-\theta_0) + \Lambda(x)\epsilon(x;\theta,\theta_0),$$

where $\sup_{x\in\mathbb{R}}|\epsilon(x;\theta,\theta_0)| = o(||\theta-\theta_0||)$ as $\theta \to \theta_0$.

A2. There exists a constant $K > 0$ such that:

$$|S(x) - S(x')| \leq K|x-x'|;$$

$$\sup_{\theta\in\Theta}|\sigma(x;\theta) - \sigma(x';\theta)| \leq K|x-x'|;$$

$$\sup_{\theta\in\Theta}||\dot{\sigma}(x;\theta) - \dot{\sigma}(x';\theta)|| \leq K|x-x'|;$$

$$\inf_{\theta\in\Theta}\inf_{x\in\mathbb{R}}\sigma(x;\theta) > 0.$$

Moreover, the function $x \mapsto \sigma_{\theta_0}^2(x) = \sigma(x;\theta_0)^2$ is two times continuously differentiable with the derivatives that have polynomial majorants.

A3. There exists a measurable function Λ on $\mathbb{R}$ with a polynomial majorant such that:

$$|\sigma(x;\theta_1) - \sigma(x;\theta_2)| \leq \Lambda(x)||\theta_1-\theta_2||, \quad \forall\theta_1,\theta_2 \in \Theta;$$

$$||\dot{\sigma}(x;\theta_1) - \dot{\sigma}(x;\theta_2)|| \leq \Lambda(x)||\theta_1-\theta_2||, \quad \forall\theta_1,\theta_2 \in \Theta.$$

A4. $\sup_{s\in[0,\infty)} E_{S,\theta_0}|X_s|^k < \infty$ for every $k \geq 1$.

A5. Under (S,θ_0), the process X is ergodic with the invariant law μ_{S,θ_0}, in the sense that it holds for every μ_{S,θ_0}-integrable function f that $t^{-1}\int_0^t f(X_s)ds$ converges in probability to $\int_{\mathbb{R}} f(x)\mu_{S,\theta_0}(dx)$ as $t \to \infty$.

A6. The matrix $I(S,\theta_0)$ given by (3) is invertible.

A7. It holds for every $\varepsilon > 0$ that $\inf_{\theta:||\theta-\theta_0||>\varepsilon} ||\Psi(\theta)|| > 0$ where

$$\Psi(\theta) = \int_{\mathbb{R}} \frac{\dot{\sigma}(x;\theta)}{\sigma(x;\theta)^3}\{\sigma(x;\theta_0)^2 - \sigma(x;\theta)^2\}\mu_{S,\theta_0}(dx).$$

We take the following approach. First, we will show that $\sup_{\theta\in\Theta} ||\Psi_n(\theta) - \Psi_n^b(\theta)|| = o_P(n^{-1/2})$, where

$$\Psi_n^b(\theta) = \frac{1}{n}\sum_{i=1}^n \frac{\dot{\sigma}(X_{t_{i-1}^n};\theta)}{\sigma(X_{t_{i-1}^n};\theta)^3}\left\{\frac{\sigma(X_{t_{i-1}^n};\theta_0)^2|W_{t_i^n} - W_{t_{i-1}^n}|^2}{|t_i^n - t_{i-1}^n|} - \sigma(X_{t_{i-1}^n};\theta)^2\right\}.$$

The "compensator" of $\Psi_n^b(\theta)$ is

$$\widetilde{\Psi}_n(\theta) = \frac{1}{n}\sum_{i=1}^n \frac{\dot{\sigma}(X_{t_{i-1}^n};\theta)}{\sigma(X_{t_{i-1}^n};\theta)^3}\{\sigma(X_{t_{i-1}^n};\theta_0)^2 - \sigma(X_{t_{i-1}^n};\theta)^2\},$$

in the sense that the difference $\Psi_n^b(\theta) - \widetilde{\Psi}_n(\theta)$ is the sum of a martingale difference array. In this way we will be able to prove the following lemma, in which we denote by $C_b(\Theta)$ the space of bounded functions from Θ to $\mathbb{R}^d$ which are continuous with respect to the Euclidean metric and equip the space with the uniform metric.

Lemma 2.1 *$n^{1/2}(\Psi_n(\cdot) - \widetilde{\Psi}_n(\cdot))$ converges weakly in $C_b(\Theta)$ to the zero-mean Gaussian random field $G(\cdot)$ with the covariance*

$$EG(\theta_1)G(\theta_2)^T = 2\int_{\mathbb{R}} \frac{\dot{\sigma}(x;\theta_1)\dot{\sigma}(x;\theta_2)^T}{\sigma(x;\theta_1)^3\sigma(x;\theta_2)^3}\sigma(x;\theta_0)^4\mu_{S,\theta_0}(dx), \quad \forall\theta_1,\theta_2 \in \Theta.$$

We will also be able to show the "differentiability" of $\theta \mapsto \widetilde{\Psi}_n(\theta)$.

Lemma 2.2 *For any random sequence θ_n converging to θ_0 in probability, it holds that*

$$\widetilde{\Psi}_n(\theta_n) - \widetilde{\Psi}_n(\theta_0) = \widetilde{\Psi}_n(\theta_n) = -I(S,\theta_0)(\theta_n - \theta_0) + o_P(||\theta_n - \theta_0||).$$

Combining these two lemmas, we obtain the main result of the paper.

Theorem 2.3 *Assume (1), (2) and* **A1** *–* **A6**. *If the random sequence $\widehat{\theta}_n$ satisfies that $\Psi_n(\widehat{\theta}_n) = o_P(n^{-1/2})$ and that $||\widehat{\theta}_n - \theta_0|| = o_P(1)$, it holds that $n^{1/2}(\widehat{\theta}_n - \theta_0)$ converges weakly to $N(0, I(S,\theta_0)^{-1})$. If* **A7** *is also satisfied, the assumption that $||\widehat{\theta}_n - \theta_0|| = o_P(1)$ is automatically satisfied.*

Proof. Apply Theorem 2.1 of Nishiyama (2009a). The last assertion follows e.g. from Theorem 5.9 of van der Vaart (1998) by using the fact that $\sup_{\theta\in\Theta} ||\Psi_n(\theta) - \Psi(\theta)|| = o_P(1)$ (see Lemma 3.2 below). □

We remark that a consistent estimator for $I(S,\theta_0)$ can be also constructed without specifying any structure of S, as follows:

$$\widehat{I}_n = \frac{2}{n}\sum_{i=1}^n \frac{\dot{\sigma}(X_{t_{i-1}^n};\widehat{\theta}_n)\dot{\sigma}(X_{t_{i-1}^n};\widehat{\theta}_n)^T}{\sigma(X_{t_{i-1}^n};\widehat{\theta}_n)^2}.$$

Use Lemma 3.2 below for the proof of this claim.

3 Proofs of lemmas

The following lemmas are well known.

Lemma 3.1 *Let K be the constant in* **A2**.
(i) For every $k \geq 1$ there exists a constant $C_k > 0$ such that

$$E_{S,\theta_0}\left[\sup_{s\in[t_{i-1}^n,t_i^n]}|X_s - X_{t_{i-1}^n}|^k \middle| \mathcal{F}_{t_{i-1}^n}\right] \leq K^k C_k |t_i^n - t_{i-1}^n|^{k/2}(1+|X_{t_{i-1}^n}|)^k.$$

(ii) For any measurable function f with a polynomial majorant, there exists a constant $\mathcal{C}_{f,K} > 0$ such that if $|t_i^n - t_{i-1}^n| \leq 1$

$$E_{S,\theta_0}\left[\sup_{s\in[t_{i-1}^n,t_i^n]}|f(X_s)| \middle| \mathcal{F}_{t_{i-1}^n}\right] \leq \mathcal{C}_{f,K}(1+|X_{t_{i-1}^n}|)^{\mathcal{C}_{f,K}}.$$

Lemma 3.2 *Let $\{f(\cdot;\theta);\theta\in\Theta\}$ be a class of Lipschitz continuous functions on $\mathbb{R}$ each of which has a polynomial majorant, and suppose that there exists a measurable function Λ on $\mathbb{R}$ with a polynomial majorant such that*

$$|f(x;\theta_1) - f(x;\theta_2)| \leq \Lambda(x)||\theta_1 - \theta_2||, \quad \forall\theta_1,\theta_2\in\Theta.$$

Then, it holds that

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^n f(X_{t_{i-1}^n};\theta) - \frac{1}{t_n^n}\int_0^{t_n^n} f(X_s;\theta)ds\right| = o_P(1).$$

Proof of Lemma 2.1. Recall and put

$$\begin{aligned}
\Psi_n(\theta) &= \frac{1}{n}\sum_{i=1}^n \frac{\dot{\sigma}(X_{t_{i-1}^n};\theta)}{\sigma(X_{t_{i-1}^n};\theta)^3}\left\{\frac{|X_{t_i^n} - X_{t_{i-1}^n}|^2}{|t_i^n - t_{i-1}^n|} - \sigma(X_{t_{i-1}^n};\theta)^2\right\}, \\
\Psi_n^a(\theta) &= \frac{1}{n}\sum_{i=1}^n \frac{\dot{\sigma}(X_{t_{i-1}^n};\theta)}{\sigma(X_{t_{i-1}^n};\theta)^3}\left\{\frac{\left|\int_{t_{i-1}^n}^{t_i^n}\sigma(X_s;\theta_0)dW_s\right|^2}{|t_i^n - t_{i-1}^n|} - \sigma(X_{t_{i-1}^n};\theta)^2\right\}, \\
\Psi_n^b(\theta) &= \frac{1}{n}\sum_{i=1}^n \frac{\dot{\sigma}(X_{t_{i-1}^n};\theta)}{\sigma(X_{t_{i-1}^n};\theta)^3}\left\{\frac{\sigma(X_{t_{i-1}^n};\theta_0)^2|W_{t_i^n} - W_{t_{i-1}^n}|^2}{|t_i^n - t_{i-1}^n|} - \sigma(X_{t_{i-1}^n};\theta)^2\right\}.
\end{aligned}$$

Noting that $E[||W_{t_i^n} - W_{t_{i-1}^n}|^2 - |t_i^n - t_{i-1}^n||^2|\mathcal{F}_{t_{i-1}^n}] = 2|t_i^n - t_{i-1}^n|^2$, it follows e.g. from the Jain-Marcus central limit theorem for martingale difference arrays (see Nishiyama (1996, 2000)) that $n^{1/2}(\Psi_n^b(\cdot)-\widetilde{\Psi}_n(\cdot))$ converges weakly in $C_b(\Theta)$ to $G(\cdot)$. So it is sufficient to show that $\sup_\theta ||\Psi_n(\theta)-\Psi_n^a(\theta)|| = o_P(n^{-1/2})$ and $\sup_\theta ||\Psi_n^a(\theta)-\Psi_n^b(\theta)|| = o_P(n^{-1/2})$.

First observe that $\Psi_n(\theta)-\Psi_n^a(\theta) = A_n^1(\theta)+A_n^2(\theta)$ where

$$\begin{aligned} A_n^1(\theta) &= \frac{1}{n}\sum_{i=1}^n \frac{\dot\sigma(X_{t_{i-1}^n};\theta)}{\sigma(X_{t_{i-1}^n};\theta)^3} \frac{\left|\int_{t_{i-1}^n}^{t_i^n} S(X_s)ds\right|^2}{|t_i^n-t_{i-1}^n|}, \\ A_n^2(\theta) &= \frac{2}{n}\sum_{i=1}^n \frac{\dot\sigma(X_{t_{i-1}^n};\theta)}{\sigma(X_{t_{i-1}^n};\theta)^3} \frac{\int_{t_{i-1}^n}^{t_i^n} \sigma(X_s;\theta_0)dW_s \int_{t_{i-1}^n}^{t_i^n} S(X_s)ds}{|t_i^n-t_{i-1}^n|}. \end{aligned}$$

Since

$$\begin{aligned} E\left[\left|\int_{t_{i-1}^n}^{t_i^n} \sigma(X_s;\theta_0)ds\right|^2 \middle| \mathcal{F}_{t_{i-1}^n}\right] &\le |t_i^n-t_{i-1}^n| \int_{t_{i-1}^n}^{t_i^n} E\left[\sigma(X_s;\theta_0)^2|\mathcal{F}_{t_{i-1}^n}\right] ds \\ &\le |t_i^n-t_{i-1}^n|^2 C(1+|X_{t_{i-1}^n}|)^C \end{aligned}$$

where $C=\mathcal{C}_{\sigma_{\theta_0}^2,K}$ in the notation of Lemma 3.1 (ii), it is easy to see that $\sup_\theta ||A_n^1(\theta)|| = O_P(\Delta_n) = o_P(n^{-1/2})$. On the other hand, since

$$\begin{aligned} & E\left[\left|\int_{t_{i-1}^n}^{t_i^n} \sigma(X_s;\theta_0)dW_s \int_{t_{i-1}^n}^{t_i^n} (S(X_s)-S(X_{t_{i-1}^n})ds\right| \middle| \mathcal{F}_{t_{i-1}^n}\right] \\ \le\ & \sqrt{E\left[\left|\int_{t_{i-1}^n}^{t_i^n} \sigma(X_s;\theta_0)dW_s\right|^2 \middle| \mathcal{F}_{t_{i-1}^n}\right]} \sqrt{E\left[\left|\int_{t_{i-1}^n}^{t_i^n} S(X_s)-S(X_{t_{i-1}^n})ds\right|^2 \middle| \mathcal{F}_{t_{i-1}^n}\right]} \\ \le\ & \sqrt{E\left[\int_{t_{i-1}^n}^{t_i^n} \sigma(X_s;\theta_0)^2ds \middle| \mathcal{F}_{t_{i-1}^n}\right]} \sqrt{E\left[|t_i^n-t_{i-1}^n| \int_{t_{i-1}^n}^{t_i^n} |S(X_s)-S(X_{t_{i-1}^n})|^2ds \middle| \mathcal{F}_{t_{i-1}^n}\right]} \\ \le\ & \sqrt{\int_{t_{i-1}^n}^{t_i^n} C(1+|X_{t_{i-1}^n}|)^C ds} \sqrt{|t_i^n-t_{i-1}^n| \int_{t_{i-1}^n}^{t_i^n} K^2C|t_i^n-t_{i-1}^n|(1+|X_{t_{i-1}^n}|)^C ds} \\ =\ & CK|t_i^n-t_{i-1}^n|^2(1+|X_{t_{i-1}^n}|)^C \end{aligned}$$

where $C=\mathcal{C}_{\sigma_{\theta_0}^2,K}\vee C_2$ in the notation of Lemma 3.1, it holds that $\sup_\theta ||A_n^2(\theta)-A_n^3(\theta)|| = O_P(\Delta_n) = o_P(n^{-1/2})$, where

$$A_n^3(\theta) = \frac{2}{n}\sum_{i=1}^n \frac{\dot\sigma(X_{t_{i-1}^n};\theta)}{\sigma(X_{t_{i-1}^n};\theta)^3} S(X_{t_{i-1}^n}) \int_{t_{i-1}^n}^{t_i^n} \sigma(X_s;\theta_0)dW_s.$$

By the Jain-Marcus theorem we have that $n^{1/2}A_n^3(\cdot)$ converges weakly to zero in $C_b(\Theta)$, and thus $\sup_\theta ||n^{1/2}A_n^3(\theta)|| = o_P(1)$. Therefore we have shown that $\sup_\theta ||\Psi_n(\theta)-\Psi_n^a(\theta)|| = o_P(n^{-1/2})$.

Next observe that $\Psi_n^a(\theta) - \Psi_n^b(\theta) = B_n^1(\theta) + B_n^2(\theta) + B_n^3(\theta)$ where

$$
\begin{aligned}
B_n^1(\theta) &= \frac{1}{n}\sum_{i=1}^n \frac{\dot\sigma(X_{t_{i-1}^n};\theta)}{\sigma(X_{t_{i-1}^n};\theta)^3} \frac{\left|\int_{t_{i-1}^n}^{t_i^n}\sigma(X_s;\theta_0)dW_s\right|^2 - E\left[\left|\int_{t_{i-1}^n}^{t_i^n}\sigma(X_s;\theta_0)dW_s\right|^2\middle|\mathcal{F}_{t_{i-1}^n}\right]}{|t_i^n - t_{i-1}^n|},\\
B_n^2(\theta) &= \frac{1}{n}\sum_{i=1}^n \frac{\dot\sigma(X_{t_{i-1}^n};\theta)}{\sigma(X_{t_{i-1}^n};\theta)^3}\sigma(X_{t_{i-1}^n};\theta_0)^2 \frac{|W_{t_i^n} - W_{t_{i-1}^n}|^2 - |t_i^n - t_{i-1}^n|}{|t_i^n - t_{i-1}^n|},\\
B_n^3(\theta) &= \frac{1}{n}\sum_{i=1}^n \frac{\dot\sigma(X_{t_{i-1}^n};\theta)}{\sigma(X_{t_{i-1}^n};\theta)^3} \frac{E\left[\left|\int_{t_{i-1}^n}^{t_i^n}\sigma(X_s;\theta_0)dW_s\right|^2\middle|\mathcal{F}_{t_{i-1}^n}\right] - \sigma(X_{t_{i-1}^n};\theta_0)^2|t_i^n - t_{i-1}^n|}{|t_i^n - t_{i-1}^n|}\\
&= \frac{1}{n}\sum_{i=1}^n \frac{\dot\sigma(X_{t_{i-1}^n};\theta)}{\sigma(X_{t_{i-1}^n};\theta)^3} \frac{\int_{t_{i-1}^n}^{t_i^n}\left(E\left[\sigma(X_s;\theta_0)^2\middle|\mathcal{F}_{t_{i-1}^n}\right] - \sigma(X_{t_{i-1}^n};\theta_0)^2\right)ds}{|t_i^n - t_{i-1}^n|}.
\end{aligned}
$$

It follows from the Jain-Marcus theorem that $n^{1/2}B_n^1(\cdot)$ and $n^{1/2}B_n^2(\cdot)$ converge weakly to zero in $C_b(\Theta)$, thus $\sup_\theta ||B_n^1(\theta)|| = o_P(n^{-1/2})$ and $\sup_\theta ||B_n^2(\theta)|| = o_P(n^{-1/2})$. On the other hand, by Itô's formula we have

$$
\begin{aligned}
\left|E\left[\sigma(X_s;\theta_0)^2\middle|\mathcal{F}_{t_{i-1}^n}\right] - \sigma(X_{t_{i-1}^n};\theta_0)^2\right| &= \left|\int_{t_{i-1}^n}^s E\left[f(X_u)|\mathcal{F}_{t_{i-1}^n}\right]du\right|\\
&\le |s - t_{i-1}^n|C(1 + |X_{t_{i-1}^n}|)^C
\end{aligned}
$$

where $C = \mathcal{C}_{f,K}$ in the notation of Lemma 3.1 (ii) with $f = (\sigma_{\theta_0}^2)'S + \frac{1}{2}(\sigma_{\theta_0}^2)''\sigma_{\theta_0}^2$, thus $\sup_\theta ||B_n^3(\theta)|| = O_P(\Delta_n) = o_P(n^{-1/2})$. We therefore have proved that $\sup_\theta ||\Psi_n^a(\theta) - \Psi_n^b(\theta)|| = o_P(n^{-1/2})$, and the proof is completed. □

Proof of Lemma 2.2. Trivially, $\widetilde{\Psi}_n(\theta_0) = 0$. For any random sequence θ_n such that $||\theta_n - \theta_0|| = o_P(1)$, we have that

$$
\begin{aligned}
\widetilde{\Psi}_n(\theta_n) &= \frac{1}{n}\sum_{i=1}^n \frac{\dot\sigma(X_{t_{i-1}^n};\theta_n)}{\sigma(X_{t_{i-1}^n};\theta_n)^3}\{\sigma(X_{t_{i-1}^n};\theta_0)^2 - \sigma(X_{t_{i-1}^n};\theta_n)^2\}\\
&= \frac{1}{n}\sum_{i=1}^n \frac{\dot\sigma(X_{t_{i-1}^n};\theta_n)}{\sigma(X_{t_{i-1}^n};\theta_n)^3} 2\sigma(X_{t_{i-1}^n};\theta_0)\dot\sigma(X_{t_{i-1}^n};\theta_0)^T(\theta_0 - \theta_n) + o_P(||\theta_n - \theta_0||)\\
&= \frac{2}{n}\sum_{i=1}^n \frac{\dot\sigma(X_{t_{i-1}^n};\theta_0)\dot\sigma(X_{t_{i-1}^n};\theta_0)^T}{\sigma(X_{t_{i-1}^n};\theta_0)^2}(\theta_0 - \theta_n) + o_P(||\theta_n - \theta_0||)\\
&= \frac{2}{t_n^n}\int_0^{t_n^n} \frac{\dot\sigma(X_s;\theta_0)\dot\sigma(X_s;\theta_0)^T}{\sigma(X_s;\theta_0)^2}ds(\theta_0 - \theta_n) + o_P(||\theta_n - \theta_0||)\\
&= -2\int_{\mathbb{R}} \frac{\dot\sigma(x;\theta_0)\dot\sigma(x;\theta_0)^T}{\sigma(x;\theta_0)^2}\mu_{S,\theta_0}(dx)(\theta_n - \theta_0) + o_P(||\theta_n - \theta_0||).
\end{aligned}
$$

The proof is finished. □

Acknowledgements. I thank Professors Hiroki Masuda and Masayuki Uchida for their comments to an earlier version of this paper.

References

[1] Florens-Zmirou, D. (1989). Approximate discrete time schemes for statistics of diffusion processes. *Statistics.* **20** 547-557.

[2] Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Inst. H. Poincaré Probab. Statist.* **29** 119–151.

[3] Gobet, E. (2002). LAN property for ergodic diffusions with discrete observations. *Ann. Inst. H. Poincaré Probab. Statist.* **38** 711-737.

[4] Iacus, S., Uchida, M. and Yoshida, N. (2009). Parametric estimation for partially hidden diffusion processes sampled at discrete times. *Stochastic Process. Appl.* **119** 1580-1600.

[5] Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24** 211-229.

[6] Nishiyama, Y. (1996). A central limit theorem for ℓ^∞-valued martingale difference arrays and its application. *Preprint* **971**, Dept. of Mathematics, Utrecht Univ. Available at `http://www.ism.ac.jp/~nisiyama/`

[7] Nishiyama, Y. (2000). Weak convergence of some classes of martingales with jumps. *Ann. Probab.* **28** 685-712.

[8] Nishiyama, Y. (2009a). Asymptotic theory of semiparametric Z-estimators for stochastic processes with applications to ergodic diffusions and time series. *Ann. Statist.* **37** 3555-3579.

[9] Nishiyama, Y. (2009b). A note on semiparametric estimation for ergodic diffusion processes. *Submitted for publication.*

[10] van der Vaart, A.W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.

[11] Yoshida, N. (1992). Estimation for diffusion processes from discrete observations. *J. Multivariate Anal.* **41** 220-242.