MINIMUM CONTRAST ESTIMATION FOR DISCRETELY OBSERVED DIFFUSION PROCESSES WITH SMALL DISPERSION PARAMETER

Uchida, Masayuki
Faculty of Mathematics, Kyushu University

https://doi.org/10.5109/12577

出版情報：Bulletin of informatics and cybernetics. 36, pp.35-49, 2004-12. Research Association of Statistical Sciences

MINIMUM CONTRAST ESTIMATION FOR DISCRETELY OBSERVED DIFFUSION PROCESSES WITH SMALL DISPERSION PARAMETER

by

Masayuki Uchida

Reprinted from the Bulletin of Informatics and Cybernetics
Research Association of Statistical Sciences, Vol.36

FUKUOKA, JAPAN
2004

Bulletin of Informatics and Cybernetics, Vol. 36, 2004

MINIMUM CONTRAST ESTIMATION FOR DISCRETELY OBSERVED DIFFUSION PROCESSES WITH SMALL DISPERSION PARAMETER

By

Masayuki Uchida*

Abstract

The parametric estimation of both drift and diffusion coefficient parameters for d-dimensional diffusion processes with small dispersion parameter ε is stated when the data are discretely observed at equidistant time points $k/n,\ k = 0, 1, \cdots, n$. Using the contrast function based on a Gaussian approximation to the transition density, we present asymptotic properties for the minimum contrast estimator as ε tends to 0 and n tends to ∞ simultaneously.

Key Words and Phrases: Discrete time sampling, parametric inference, stochastic differential equation.

1. Introduction

In this paper, we consider a family of d-dimensional diffusion processes with small dispersion parameter defined by the stochastic differential equations

$$dX_t \quad = \quad b(X_t, \theta)dt + \varepsilon\sigma(X_t, \theta)dw_t,\ t \in [0, 1],\ \varepsilon \in (0, 1], \quad X_0 = x_0, \tag{1}$$

where $\theta \in \bar{\Theta}$, Θ is an open bounded convex subset of $\mathbf{R}^p$, x_0 and ε are known constants, b is an $\mathbf{R}^d$-valued function defined on $\mathbf{R}^d \times \bar{\Theta}$, σ is an $\mathbf{R}^d \otimes \mathbf{R}^r$-valued function defined on $\mathbf{R}^d \times \bar{\Theta}$ and w is an r-dimensional standard Wiener process. We assume that the drift b and the diffusion coefficient σ are known apart from the parameter θ. The data are discretely observed at the points of time $t_k = k/n,\ k = 0, 1, \cdots, n$, on the interval $[0, 1]$, that is, $(X_{t_k})_{0\le k\le n}$. The asymptotics considered is when ε tends to 0 and n tends to ∞ simultaneously.

Small dispersion asymptotics for diffusion processes and their applications are well-developed. Most of the researches were focused on the case when the whole path is completely observed. For details, see Kutoyants (1984, 1994), Yoshida (1992a, 1993, 1996, 2001), Dermoune and Kutoyants (1995), Sakamoto and Yoshida (1996), Uchida and Yoshida (2004). For applications to mathematical finance, see Yoshida (1992b), Kim and Kunitomo (1999), Takahashi (1999), Kunitomo and Takahashi (2001), Takahashi and Yoshida (2004), Uchida and Yoshida (2004).

* Faculty of Mathematics, Kyushu University Ropponmatsu 4–2–1 Chuo-Ku Fukuoka 810-8560 Japan. tel +81–92–726–4785 uchida@math.kyushu-u.ac.jp

On the other hand, there are not so many studies on parametric estimation for diffusion processes with small dispersion parameter from discrete observations. Genon-Catalot (1990) and Laredo (1990) obtained asymptotically efficient estimators of drift parameters for discretely observed diffusion processes with small dispersion parameter under the assumption that the diffusion coefficient function is known. Uchida (2004) studied the efficient estimation under a general assumption on ε and n when the diffusion coefficient function is known. Sørensen (2000) considered martingale estimating functions for diffusion processes with small dispersion parameter. He also showed the consistency and asymptotic normality of estimators of drift and diffusion coefficient parameters under the condition that the sample size of discrete observations is fixed. Recently, Sørensen and Uchida (2003) studied consistent, asymptotically normal and asymptotically efficient estimators of parameters which appear in the drift and the diffusion coefficient separately. Their diffusion model is somewhat special and is defined by

$$dX_t \quad = \quad b(X_t, \alpha)dt + \varepsilon\sigma(X_t, \beta)dw_t, \; t \in [0, 1], \; \varepsilon \in (0, 1], \quad X_0 = x_0. \tag{2}$$

They pointed out that when $(\varepsilon\sqrt{n})^{-1} \to 0$, the rate of convergence for the estimator of the drift parameter α is different from that of diffusion coefficient parameter β. For more details of the results for the model (2), see Sørensen and Uchida (2003).

Although two models (1) and (2) look quite similar, there is an obvious difference in their parameterization. Note that the model (1) includes the model (2) but its inversion does not generally hold true. In order to obtain asymptotic properties of an estimator for the model (1), we cannot use the results in Sørensen and Uchida (2003). Thus, we discuss the estimation for the diffusion model (1) whose drift and diffusion coefficient may have the same parameter. The purpose of this paper is to show that a minimum contrast estimator obtained from a contrast function based on a Gaussian approximation to the transition density is consistent and asymptotically normal.

This paper is organized as follows: In Section 2, several notations and assumptions are introduced. Section 3 presents our main result. The consistency and asymptotic normality of the minimum contrast estimator are stated. Section 4 gives two examples and the asymptotic behaviour of our estimators through simulations. Section 5 is devoted to prove the asymptotic results in Section 3.

2. Notations and assumptions

Let θ_0 denote the true value of θ. Let X_t^0 be the solution of the ordinary differential equation: $dX_t^0 = b(X_t^0, \theta_0)dt$, $X_0^0 = x_0$. We denote by C a generic positive constant independent of n and other variables in some cases (see Yoshida (1992c) and Kessler (1997)). Moreover we may write C_m if it depends on an integer m. A^* denotes the transpose of the matrix A and $|A|^2 = \mathrm{tr}(AA^*)$. Let $\bar{C}_{\uparrow}^{\infty}(\mathbf{R}^d \times \Theta; \mathbf{R}^m)$ be the space of all functions f satisfying the following two conditions: (i) $f(x, \theta)$ is an $\mathbf{R}^m$-valued function on $\mathbf{R}^d \times \Theta$ and smooth in (x, θ), (ii) for $|\mathbf{n}| \geq 0, |\nu| \geq 0$ there exists $C > 0$ such that $\sup_{\theta \in \Theta} |\delta^\nu \partial^{\mathbf{n}} f| \leq C(1 + |x|)^C$ for all x, where $\mathbf{n} = (n_1, \cdots, n_d)$ and $\nu = (\nu_1, \cdots, \nu_p)$ are multi-indices, $|\mathbf{n}| = n_1 + \cdots + n_d$, $|\nu| = \nu_1 + \cdots + \nu_p$, $\partial^{\mathbf{n}} = \partial_1^{n_1} \cdots \partial_d^{n_d}$, $\partial_i = \partial/\partial x^i$, $i = 1, \cdots, d$, $\delta^\nu = \delta_1^{\nu_1} \cdots \delta_p^{\nu_p}$, $\delta_j = \partial/\partial\theta^j$, $j = 1, \cdots, p$.

We make the following assumption on the model (1).

Assumption 2.1. *(i) Equation (1) has a unique strong solution on* $[0,1]$. *(ii) For all* $m>0$, $\sup_t E[|X_t|^m]<\infty$. *(iii)* $b(x,\theta)\in \bar{C}^{\infty}_{\uparrow}(\mathbf{R}^d\times\bar{\Theta};\mathbf{R}^d)$, $\sigma(x,\theta)\in \bar{C}^{\infty}_{\uparrow}(\mathbf{R}^d\times\bar{\Theta};\mathbf{R}^d\otimes\mathbf{R}^r)$. *(iv)* $\inf_{x,\theta}\det[\sigma\sigma^*](x,\theta)>0$, $[\sigma\sigma^*]^{-1}(x,\theta)\in \bar{C}^{\infty}_{\uparrow}(\mathbf{R}^d\times\bar{\Theta};\mathbf{R}^d\otimes\mathbf{R}^d)$.

Moreover, we consider the following assumption for ε and n.

Assumption 2.2. $\lim_{\varepsilon\to 0,n\to\infty}(\varepsilon\sqrt{n})^{-1}=M$, *where* $M<\infty$.

Assumptions 2.1 and 2.2 will be made throughout this paper, while the following Assumption 2.3 will be needed as an identifiability assumption in order to obtain our main theorem in Section 3. (at least consistent estimators).

Assumption 2.3. $b(X_t^0,\theta)=b(X_t^0,\theta_0)$, $\sigma\sigma^*(X_t^0,\theta)=\sigma\sigma^*(X_t^0,\theta_0)\Rightarrow\theta=\theta_0$.

Let P_θ be the law of the solution of (1). Set $\Xi_k(\theta)=[\sigma\sigma^*](X_{t_k},\theta)$ and $B(x,\theta_0,\theta)=b(x,\theta_0)-b(x,\theta)$. We define $\mathcal{I}(\theta_0)=\Big(\mathcal{I}_b^{i,j}(\theta_0)+\mathcal{I}_\sigma^{i,j}(\theta_0)\Big)_{1\le i,j\le p}$, where

$$\begin{aligned}\mathcal{I}_b^{i,j}(\theta_0) &= M^2\int_0^1\left(\frac{\partial}{\partial\theta_i}b(X_s^0,\theta_0)\right)^*[\sigma\sigma^*]^{-1}(X_s^0,\theta_0)\left(\frac{\partial}{\partial\theta_j}b(X_s^0,\theta_0)\right)ds,\\ \mathcal{I}_\sigma^{i,j}(\theta_0) &= \frac{1}{2}\int_0^1\mathrm{tr}\left[\left(\frac{\partial}{\partial\theta_i}[\sigma\sigma^*]\right)[\sigma\sigma^*]^{-1}\left(\frac{\partial}{\partial\theta_j}[\sigma\sigma^*]\right)[\sigma\sigma^*]^{-1}(X_s^0,\theta_0)\right]ds.\end{aligned}$$

Set

$$\begin{aligned}U(\theta,\theta_0) &= \int_0^1\log\det[\sigma\sigma^*](X_s^0,\theta)ds+\int_0^1\mathrm{tr}\left[[\sigma\sigma^*](X_s^0,\theta_0)[\sigma\sigma^*]^{-1}(X_s^0,\theta)\right]ds\\ &\quad +M^2\int_0^1 B^*(X_s^0,\theta_0,\theta)[\sigma\sigma^*]^{-1}(X_s^0,\theta)B(X_s^0,\theta_0,\theta)ds.\end{aligned}$$

3. The minimum contrast estimator

In order to obtain the minimum contrast estimator, we construct the contrast function based on a Gaussian approximation to the transition density in the same way as in Kessler (1997). From Lemma 1 in Florens-Zmirou (1989), we have the following contrast function.

$$U_{\varepsilon,n}(\theta)=\sum_{k=1}^n\{\log\det\Xi_{k-1}(\theta)+\varepsilon^{-2}nP_k^*(\theta)\Xi_{k-1}(\theta)^{-1}P_k(\theta)\},$$

where $P_k(\theta)=X_{t_k}-X_{t_{k-1}}-b(X_{t_{k-1}},\theta)/n$.

Proposition 3.1. *Suppose that Assumptions 2.1 and 2.2 hold true. Then, in* P_{θ_0}-*probability, as* $\varepsilon\to 0$ *and* $n\to\infty$,

$$\sup_{\theta\in\bar{\Theta}}\left|\frac{1}{n}U_{\varepsilon,n}(\theta)-U(\theta,\theta_0)\right|\to 0.$$

Proposition 3.2. *Suppose that Assumptions 2.1 and 2.2 hold true. Then, in P_{θ_0}-probability, as $\varepsilon \to 0$ and $n \to \infty$,*
(i)

$$C_{\varepsilon,n}(\theta_0) := \left(\frac{1}{n}\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}U_{\varepsilon,n}(\theta_0)\right)_{1\le i,j\le p}\right) \to 2\mathcal{I}(\theta_0),$$

(ii)

$$\sup_{|\theta|\le\eta_{\varepsilon,n}} |C_{\varepsilon,n}(\theta_0+\theta) - C_{\varepsilon,n}(\theta_0)| \to 0,$$

where $\eta_{\varepsilon,n} \to 0$.

Proposition 3.3. *Suppose that Assumptions 2.1 and 2.2 hold true. Then*

$$\Lambda_{\varepsilon,n} := \left(-\frac{1}{\sqrt{n}}\left(\frac{\partial}{\partial\theta_i}U_{\varepsilon,n}(\theta_0)\right)_{1\le j\le p}\right) \to N\left(0, 4\mathcal{I}(\theta_0)\right)$$

in distribution, under P_{θ_0}, as $\varepsilon \to 0$ and $n \to \infty$.

Let $\hat{\theta}_{\varepsilon,n}$ be a minimum contrast estimator defined by

$$U_{\varepsilon,n}(\hat{\theta}_{\varepsilon,n}) = \inf_{\theta\in\bar{\Theta}} U_{\varepsilon,n}(\theta). \tag{3}$$

Our main theorem is as follows.

Theorem 3.4. *Suppose that Assumptions 2.1, 2.2 and 2.3 hold true. Then,*

$$\hat{\theta}_{\varepsilon,n} \to \theta_0$$

in P_{θ_0}-probability as $\varepsilon \to 0$ and $n \to \infty$. Moreover, if $\mathcal{I}(\theta_0)$ is non-singular, then

$$\sqrt{n}(\hat{\theta}_{\varepsilon,n} - \theta_0) \to N\left(0, \mathcal{I}(\theta_0)^{-1}\right)$$

in distribution, under P_{θ_0}, as $\varepsilon \to 0$ and $n \to \infty$.

Remark. (i) From the proof of the consistency in Theorem 3.4, it can be immediately shown that the consistency of $\hat{\theta}_{\varepsilon,n}$ holds true under $\theta_0 \in \bar{\Theta}$ instead of $\theta_0 \in \Theta$. (ii) To obtain Theorem 3.4, we can relax (iii)-(iv) in Assumption 2.1. Using a "classical" localization argument, we can replace them by mild regularity conditions about b and σ near the neighborhood of the path of X_t^0. (iii) Using approximate martingale estimating functions, we can derive estimators with the same properties as Theorem 3.4. For details, see Uchida (2003).

4. Examples

In this section, we study the behavior of our estimators in two examples through simulations. In both examples, for each $\varepsilon = 0.1, 0.05, 0.01, 0.005$ and $n = 50, 100, 500, 1000$, we simulated 1000 independent sample paths with $\theta = \theta_0$ (true parameter value) and the initial value x_0. The simulations were done with the Euler-Maruyama scheme, see Kloeden and Platen (1992), Deelstra and Delbaen (1999) and Kanagawa and Ogawa (2001). For each sample path, the minimum contrast estimators $\hat{\theta}_{\varepsilon,n}$ were calculated. For the resulting 1000 values of estimators, the mean and the standard deviation of the estimators were computed. The means should be compared to the true parameter values, while the standard deviations can be compared to the theoretical values given by Theorem 3.4.

4.1. The Pedersen-Bibby-Sørensen-Kessler model

We first consider the diffusion model with small dispersion parameter given by the one dimensional stochastic differential equation

$$dX_t \quad = \quad -\theta X_t dt + \varepsilon\sqrt{\theta + X_t^2} dw_t,\ t \in [0,1],\ \varepsilon \in (0,1], \quad X_0 = x_0, \tag{4}$$

where $\theta > 0$, x_0 and ε are known constants. This model for $\varepsilon = 1$ was originally proposed by A. R. Pedersen and was studied further by Bibby and Sørensen (1995) and by Kessler (2000).

The contrast function that yields an estimator for θ is

$$U_{\varepsilon,n}(\theta) = \sum_{k=1}^{n} \left\{ \log(\theta + X_{t_{k-1}}^2) + \varepsilon^{-2} n \frac{(X_{t_k} - X_{t_{k-1}} + \frac{1}{n}\theta X_{t_{k-1}})^2}{(\theta + X_{t_{k-1}}^2)} \right\}.$$

In Table 1 below, we set the parameter value $\theta = 1$ and the initial value $x_0 = 0.5$. The small dispersion asymptotics with decreasing step size gives a very good approximation to the standard deviation of $\hat{\theta}_{\varepsilon,n}$ and small bias of $\hat{\theta}_{\varepsilon,n}$ in all cases. Therefore, we conclude that our estimator has a good approximation to the true parameter in this example.

Table 1: (PBSK model) The mean and standard deviation of the estimator $\hat{\theta}_{\varepsilon,n}$ determined from 1000 independent simulated sample paths for $\theta = 1$, $x_0 = 0.5$.

n	ε	sim. mean	sim. s.d.	theor. s.d.
50	0.1	0.982204	0.165193	0.182485
	0.05	0.988577	0.117725	0.130733
	0.01	0.990805	0.031522	0.032110
	0.005	0.990693	0.016179	0.016184
100	0.1	0.988577	0.135620	0.140629
	0.05	0.993116	0.106856	0.112477
	0.01	0.995769	0.031742	0.031775
	0.005	0.995655	0.016359	0.016141
500	0.1	0.998824	0.068914	0.068223
	0.05	0.999258	0.065236	0.064106
	0.01	0.999873	0.030709	0.029428
	0.005	0.999691	0.016483	0.015806
1000	0.1	1.000425	0.049423	0.048783
	0.05	1.000531	0.047909	0.047209
	0.01	1.000527	0.028125	0.027116
	0.005	1.000257	0.016043	0.015416

4.2. The generalized Cox-Ingersoll-Ross model

The second example is the following diffusion model with small dispersion parameter defined by the one dimensional stochastic differential equation

$$dX_t \;=\; (\alpha X_t^{2\gamma-1} + \beta X_t)dt + \varepsilon X_t^{\gamma} dw_t,\ t \in [0,1],\ \varepsilon \in (0,1], \quad X_0 = x_0, \tag{5}$$

where $\alpha, \beta \in \mathbf{R}$, $\gamma \neq 1$, x_0 and ε are known constants. This is a version of the generalized Cox-Ingersoll-Ross model introduced by Jacobsen (2001). For more details of the generalized Cox-Ingersoll-Ross model, see Jacobsen (2001, 2002).

The contrast function that yields estimators for α, β and γ is

$$U_{\varepsilon,n}(\alpha,\beta,\gamma) = \sum_{k=1}^{n} \left\{ \log X_{t_{k-1}}^{2\gamma} + \varepsilon^{-2} n \frac{\left[X_{t_k} - X_{t_{k-1}} - \frac{1}{n}(\alpha X_{t_{k-1}}^{2\gamma-1} + \beta X_{t_{k-1}}) \right]^2}{X_{t_{k-1}}^{2\gamma}} \right\}.$$

In Tables 2 and 3 below, the parameter values $\alpha = 1$, $\beta = -2$, $\gamma = \frac{1}{2}$ and the initial value $x_0 = 2$ are considered. The small dispersion asymptotics with decreasing step size gives a good approximation to the standard deviations of $\hat{\gamma}_{\varepsilon,n}$ in all cases, while this type of asymptotics gives reasonable values of the standard deviations of $\hat{\alpha}_{\varepsilon,n}$ and $\hat{\beta}_{\varepsilon,n}$ except that $n \leq 50$ and $\varepsilon \geq 0.1$. The biases of $\hat{\alpha}_{\varepsilon,n}$ and $\hat{\beta}_{\varepsilon,n}$ are small in all cases, whereas $\hat{\gamma}_{\varepsilon,n}$ has a considerable bias when $n \leq 100$ and $\varepsilon \geq 0.05$. In this example, we can conclude that our estimators have good approximations to the true parameters unless $n \leq 100$ and $\varepsilon \geq 0.05$.

5. Proofs

Let R denote a function $(0,1] \times \mathbf{R}^d \to \mathbf{R}$ for which there exists a constant C such that $|R(a,x)| \leq aC(1+|x|)^C$ for all a, x. Set $\mathcal{G}_k^n = \sigma(w_s; s \leq t_k)$.

In order to show Propositions 3.1, 3.2 and 3.3, we will need the following three lemmas. For their proofs, see the proofs of Lemmas 1, 2 and 3 in Sørensen and Uchida (2003).

Lemma 5.1. *Suppose that (i)–(iii) in Assumption 2.1 hold true. Then*

$$E_{\theta_0}[P_k^i(\theta_0)|\mathcal{G}_{k-1}^n] \;=\; R\left(\frac{1}{n^2}, X_{t_{k-1}}\right), \tag{6}$$

$$E_{\theta_0}[P_k^{i_1}(\theta_0)P_k^{i_2}(\theta_0)|\mathcal{G}_{k-1}^n] \;=\; \frac{\varepsilon^2}{n}\Xi_{k-1}^{i_1 i_2}(\theta_0) + R\left(\frac{\varepsilon^2}{n^2}, X_{t_{k-1}}\right) + R\left(\frac{1}{n^3}, X_{t_{k-1}}\right), \tag{7}$$

$$\begin{aligned} E_{\theta_0}[P_k^{i_1}(\theta_0)P_k^{i_2}(\theta_0)P_k^{i_3}(\theta_0)|\mathcal{G}_{k-1}^n] \;=\; & R\left(\frac{\varepsilon^4}{n^2}, X_{t_{k-1}}\right) + R\left(\frac{\varepsilon^2}{n^3}, X_{t_{k-1}}\right) \\ & + R\left(\frac{1}{n^4}, X_{t_{k-1}}\right), \end{aligned} \tag{8}$$

$$\begin{aligned} E_{\theta_0}\left[\prod_{j=1}^{4} P_k^{i_j}(\theta_0)|\mathcal{G}_{k-1}^n\right] \;=\; & \frac{\varepsilon^4}{n^2}\left\{\Xi_{k-1}^{i_1 i_2}\Xi_{k-1}^{i_3 i_4}(\theta_0) + \Xi_{k-1}^{i_1 i_3}\Xi_{k-1}^{i_2 i_4}(\theta_0) \right. \\ & \left. + \Xi_{k-1}^{i_1 i_4}\Xi_{k-1}^{i_2 i_3}(\theta_0)\right\} \end{aligned} \tag{9}$$

Table 2: (G-CIR model) The mean and standard deviation of the estimators $\hat{\alpha}_{\varepsilon,n}$ and $\hat{\beta}_{\varepsilon,n}$ determined from 1000 independent simulated sample paths for $\alpha = 1$, $\beta = -2$, $\gamma = \frac{1}{2}$, $x_0 = 2$.

		$\hat{\alpha}$			$\hat{\beta}$		
n	ε	sim. mean	sim. s.d.	theor. s.d.	sim. mean	sim. s.d.	theor. s.d.
50	0.1	0.992437	0.786250	0.649709	-1.988823	0.762534	0.614513
	0.05	0.978790	0.499941	0.539011	-1.969653	0.479606	0.514339
	0.01	0.986401	0.207842	0.221533	-1.969399	0.200037	0.212198
	0.005	0.985218	0.106752	0.117400	-1.966752	0.102461	0.112468
100	0.1	1.071125	0.528710	0.524346	-2.066707	0.500329	0.492208
	0.05	1.011104	0.396940	0.416521	-2.007135	0.378784	0.396113
	0.01	0.998738	0.200012	0.206986	-1.991175	0.191955	0.198230
	0.005	0.991891	0.108053	0.115054	-1.983417	0.103439	0.110215
500	0.1	1.074976	0.422729	0.388852	-2.073341	0.392405	0.358438
	0.05	1.030149	0.272470	0.247245	-2.029448	0.257438	0.231460
	0.01	1.003538	0.145504	0.146954	-2.003185	0.139485	0.140551
	0.005	1.000715	0.094932	0.100372	-1.999986	0.090897	0.096117
1000	0.1	1.077343	0.389544	0.367892	-2.075833	0.360138	0.337478
	0.05	1.030990	0.224366	0.213694	-2.030512	0.210320	0.198372
	0.01	1.007646	0.112055	0.116839	-2.007784	0.107286	0.111574
	0.005	1.004152	0.081759	0.088241	-2.004244	0.078307	0.084465

$$+R\left(\frac{\varepsilon^4}{n^3}, X_{t_{k-1}}\right) + R\left(\frac{\varepsilon^2}{n^4}, X_{t_{k-1}}\right) + R\left(\frac{1}{n^5}, X_{t_{k-1}}\right).$$

LEMMA 5.2. *Let $f \in \bar{C}^{\infty}_{\uparrow}(\mathbf{R}^d \times \bar{\Theta}; \mathbf{R})$. Suppose that (i)–(iii) in Assumption 2.1 hold true. Then, in P_{θ_0}-probability, as $\varepsilon \to 0$ and $n \to \infty$,*
(i)

$$\sup_{\theta\in\bar{\Theta}} \left| \frac{1}{n}\sum_{k=1}^{n} f(X_{t_{k-1}}, \theta) - \int_0^1 f(X_s^0, \theta) ds \right| \to 0,$$

(ii)

$$\sup_{\theta\in\bar{\Theta}} \left| \sum_{k=1}^{n} f(X_{t_{k-1}}, \theta) P_k^i(\theta_0) \right| \to 0.$$

LEMMA 5.3. *Let $f \in \bar{C}^{\infty}_{\uparrow}(\mathbf{R}^d \times \bar{\Theta}; \mathbf{R})$. Suppose that (i)–(iii) in Assumption 2.1 hold true and that $\lim_{\varepsilon\to 0, n\to\infty}(\varepsilon n)^{-1} = 0$. Then, in P_{θ_0}-probability, as $\varepsilon \to 0$ and $n \to \infty$,*

Table 3: (G-CIR model) The mean and standard deviation of the estimator $\hat{\gamma}_{\varepsilon,n}$ determined from 1000 independent simulated sample paths for $\alpha = 1$, $\beta = -2$, $\gamma = \frac{1}{2}$, $x_0 = 2$.

		$\hat{\gamma}$		
n	ε	sim. mean	sim. s.d.	theor. s.d.
50	0.1	0.355375	0.285371	0.306180
	0.05	0.372587	0.266187	0.284125
	0.01	0.477732	0.111643	0.121770
	0.005	0.495062	0.058836	0.064620
100	0.1	0.442129	0.209536	0.219472
	0.05	0.435458	0.195823	0.210907
	0.01	0.481312	0.105910	0.113568
	0.005	0.493049	0.059536	0.063300
500	0.1	0.486971	0.095367	0.099254
	0.05	0.487301	0.095008	0.098423
	0.01	0.490580	0.073839	0.079485
	0.005	0.495407	0.050871	0.055021
1000	0.1	0.496954	0.067205	0.070282
	0.05	0.497049	0.066935	0.069986
	0.01	0.496799	0.056946	0.062111
	0.005	0.498125	0.043591	0.048154

(i)

$$\varepsilon^{-2}\sum_{k=1}^{n} f(X_{t_{k-1}},\theta)P_k^i P_k^j(\theta_0) \to \int_0^1 f(X_s^0,\theta)[\sigma\sigma^*]^{ij}(X_s^0,\theta_0)ds$$

uniformly in $\theta \in \bar{\Theta}$. Moreover, if Assumption 2.2 holds true, then, in P_{θ_0}-probability,
(ii)

$$\begin{aligned}\varepsilon^{-2}\sum_{k=1}^{n} f(X_{t_{k-1}},\theta)P_k^i P_k^j(\theta) \quad &\to \quad \int_0^1 f(X_s^0,\theta)[\sigma\sigma^*]^{ij}(X_s^0,\theta_0)ds \\ &\qquad +M^2\int_0^1 f(X_s^0,\theta)B^i B^j(X_s^0,\theta_0,\theta)ds\end{aligned}$$

uniformly in $\theta \in \bar{\Theta}$ as $\varepsilon \to 0$ and $n \to \infty$.

Proof of Proposition 3.1. It follows from Lemmas 5.2 and 5.3 that in P_{θ_0}-probability,

$$\frac{1}{n}U_{\varepsilon,n}(\theta) \to U(\theta,\theta_0)$$

uniformly in $\theta \in \bar{\Theta}$ as $\varepsilon \to 0$ and $n \to \infty$. This completes the proof. □

Proof of Proposition 3.2. We first consider the uniform convergence of $C_{\varepsilon,n}(\theta)$. An easy computation implies $\frac{1}{n}\frac{\partial^2}{\partial\theta_i\partial\theta_j}U_{\varepsilon,n}(\theta) = U^{ij}_{1,\varepsilon,n}(\theta) + U^{ij}_{2,\varepsilon,n}(\theta) + U^{ij}_{3,\varepsilon,n}(\theta)$, where

$$\begin{aligned}
U^{ij}_{1,\varepsilon,n}(\theta) &= -\frac{2}{\varepsilon^2 n}\sum_{k=1}^n \left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}b(X_{t_{k-1}},\theta)\right)^* \Xi^{-1}_{k-1}(\theta)\left(P_k(\theta_0)+\frac{1}{n}B(X_{t_{k-1}},\theta_0,\theta)\right)\\
&\quad +\frac{2}{\varepsilon^2 n^2}\sum_{k=1}^n \left(\frac{\partial}{\partial\theta_i}b(X_{t_{k-1}},\theta)\right)^* \Xi^{-1}_{k-1}(\theta)\left(\frac{\partial}{\partial\theta_j}b(X_{t_{k-1}},\theta)\right),\\
U^{ij}_{2,\varepsilon,n}(\theta) &= -\frac{2}{\varepsilon^2 n}\sum_{k=1}^n \left(\frac{\partial}{\partial\theta_i}b(X_{t_{k-1}},\theta)\right)^* \frac{\partial}{\partial\theta_j}(\Xi^{-1}_{k-1}(\theta))\left(P_k(\theta_0)+\frac{1}{n}B(X_{t_{k-1}},\theta_0,\theta)\right)\\
&\quad -\frac{2}{\varepsilon^2 n}\sum_{k=1}^n \left(\frac{\partial}{\partial\theta_j}b(X_{t_{k-1}},\theta)\right)^* \frac{\partial}{\partial\theta_i}(\Xi^{-1}_{k-1}(\theta))\left(P_k(\theta_0)+\frac{1}{n}B(X_{t_{k-1}},\theta_0,\theta)\right),\\
U^{ij}_{3,\varepsilon,n}(\theta) &= \frac{1}{n}\sum_{k=1}^n\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log\det\Xi_{k-1}(\theta)\right) + \frac{1}{\varepsilon^2}\sum_{k=1}^n P_k(\theta)^*\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\Xi^{-1}_{k-1}(\theta)\right)P_k(\theta).
\end{aligned}$$

By Lemma 5.2, one has that in P_{θ_0}-probability, as $\varepsilon\to 0$ and $n\to\infty$,

$$\begin{aligned}
U^{ij}_{1,\varepsilon,n}(\theta) &\to -2M^2\int_0^1\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}b(X^0_s,\theta)\right)^*[\sigma\sigma^*]^{-1}(X^0_s,\theta)B(X^0_s,\theta_0,\theta)ds \qquad (10)\\
&\quad +2M^2\int_0^1\left(\frac{\partial}{\partial\theta_i}b(X^0_s,\theta)\right)^*[\sigma\sigma^*]^{-1}(X^0_s,\theta)\left(\frac{\partial}{\partial\theta_j}b(X^0_s,\theta)\right)ds\\
&\quad =: U^{ij}_1(\theta),\\
U^{ij}_{2,\varepsilon,n}(\theta) &\to -2M^2\int_0^1\left(\frac{\partial}{\partial\theta_i}b(X^0_s,\theta)\right)^*\left(\frac{\partial}{\partial\theta_j}[\sigma\sigma^*]^{-1}(X^0_s,\theta)\right)B(X^0_s,\theta_0,\theta)ds \qquad (11)\\
&\quad -2M^2\int_0^1\left(\frac{\partial}{\partial\theta_j}b(X^0_s,\theta)\right)^*\left(\frac{\partial}{\partial\theta_i}[\sigma\sigma^*]^{-1}(X^0_s,\theta)\right)B(X^0_s,\theta_0,\theta)ds\\
&\quad =: U^{ij}_2(\theta),\\
U^{ij}_{3,\varepsilon,n}(\theta) &\to \int_0^1\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log\det[\sigma\sigma^*](X^0_s,\theta)ds \qquad (12)\\
&\quad +\int_0^1\mathrm{tr}\left[[\sigma\sigma^*](X^0_s,\theta_0)\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}[\sigma\sigma^*]^{-1}(X^0_s,\theta)\right)\right]ds\\
&\quad +M^2\int_0^1\mathrm{tr}\left[BB^*(X^0_s,\theta_0,\theta)\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}[\sigma\sigma^*]^{-1}(X^0_s,\theta)\right)\right]ds\\
&\quad =: U^{ij}_3(\theta)
\end{aligned}$$

uniformly in $\theta\in\bar{\Theta}$. By (10), (11) and (12), we complete the proof of (i).

Next, by (iii) and (iv) in Assumption 2.1, $U^{ij}_1(\theta)$, $U^{ij}_2(\theta)$ and $U^{ij}_3(\theta)$ are continuous with respect to θ, which completes the proof of (ii). □

Proof of Proposition 3.3. We define $\xi^i_k(\theta_0)$ and $\eta^i_k(\theta_0)$ as follows:

$$-\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta_i}U_{\varepsilon,n}(\theta_0)$$

$$
\begin{aligned}
&= \left\{\sum_{k=1}^{n}\frac{2}{\varepsilon^2\sqrt{n}}\sum_{l_1=1}^{d}\left[\left(\frac{\partial}{\partial\theta_i}b(X_{t_{k-1}},\theta_0)\right)^*\Xi_{k-1}^{-1}(\theta_0)\right]^{l_1}P_k^{l_1}(\theta_0)\right\}\\
&\quad+\left\{-\sum_{k=1}^{n}\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta_i}\log\det\Xi_{k-1}(\theta_0)-\sum_{k=1}^{n}\frac{\sqrt{n}}{\varepsilon^2}\sum_{l_1,l_2=1}^{d}\left(\frac{\partial}{\partial\theta_i}\Xi_{k-1}^{-1}(\theta_0)\right)^{l_1l_2}P_k^{l_1}P_k^{l_2}(\theta_0)\right\}\\
&=: \sum_{k=1}^{n}\xi_k^i(\theta_0)+\sum_{k=1}^{n}\eta_k^i(\theta_0).
\end{aligned}
$$

By Theorems 3.2 and 3.4 of Hall and Heyde (1980), it suffices to show that

$$
\sum_{k=1}^{n}E_{\theta_0}[\xi_k^i(\theta_0)+\eta_k^i(\theta_0)|\mathcal{G}_{k-1}^n]\to 0, \tag{13}
$$

$$
\sum_{k=1}^{n}E_{\theta_0}[(\xi_k^{i_1}+\eta_k^{i_1})(\xi_k^{i_2}+\eta_k^{i_2})(\theta_0)|\mathcal{G}_{k-1}^n]\to 4\mathcal{I}^{i_1i_2}(\theta_0), \tag{14}
$$

$$
\sum_{k=1}^{n}E_{\theta_0}[(\xi_k^i(\theta_0))^4+(\eta_k^i(\theta_0))^4|\mathcal{G}_{k-1}^n]\to 0 \tag{15}
$$

in P_{θ_0}-probability, as $\varepsilon\to 0$ and $n\to\infty$.

Proof of (13). It follows from Lemma 5.1–(6) that

$$
\sum_{k=1}^{n}E_{\theta_0}[\xi_k^i(\theta_0)|\mathcal{G}_{k-1}^n]=\sum_{k=1}^{n}R\left(\frac{1}{\varepsilon^2n^2\sqrt{n}},X_{t_{k-1}}\right)\to 0
$$

in P_{θ_0}-probability, as $\varepsilon\to 0$ and $n\to\infty$. By Lemma 5.1–(7), we obtain that

$$
\begin{aligned}
\sum_{k=1}^{n}E_{\theta_0}[\eta_k^i(\theta_0)|\mathcal{G}_{k-1}^n] &= \sum_{k=1}^{n}\left\{-\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta_i}\log\det\Xi_{k-1}(\theta_0)\right.\\
&\quad -\frac{1}{\sqrt{n}}\mathrm{tr}\left[\Xi_{k-1}(\theta_0)\left(\frac{\partial}{\partial\theta_i}\Xi_{k-1}^{-1}(\theta_0)\right)\right]\\
&\quad \left.+R\left(\frac{1}{n\sqrt{n}},X_{t_{k-1}}\right)+R\left(\frac{1}{\varepsilon^2n^2\sqrt{n}},X_{t_{k-1}}\right)\right\}\\
&= \sum_{k=1}^{n}\left\{R\left(\frac{1}{n\sqrt{n}},X_{t_{k-1}}\right)+R\left(\frac{1}{\varepsilon^2n^2\sqrt{n}},X_{t_{k-1}}\right)\right\}\to 0
\end{aligned}
$$

in P_{θ_0}-probability, as $\varepsilon\to 0$ and $n\to\infty$. □

Proof of (14). By Lemma 5.1–(7), we have

$$
\begin{aligned}
&\sum_{k=1}^{n}E_{\theta_0}[\xi_k^{i_1}\xi_k^{i_2}(\theta_0)|\mathcal{G}_{k-1}^n]\\
&= \frac{4}{\varepsilon^4n}\sum_{k=1}^{n}\sum_{l_1,l_2=1}^{d}\left[\left(\frac{\partial}{\partial\theta_{i_1}}b(X_{t_{k-1}},\theta_0)\right)^*\Xi_{k-1}^{-1}(\theta_0)\right]^{l_1}\left[\left(\frac{\partial}{\partial\theta_{i_2}}b(X_{t_{k-1}},\theta_0))\right)^*\Xi_{k-1}^{-1}(\theta_0)\right]^{l_2}
\end{aligned}
$$

$$
\begin{aligned}
&\times E_{\theta_0}[P_k^{l_1}P_k^{l_2}(\theta_0)|\mathcal{G}_{k-1}^n] \\
= \;& \frac{4}{\varepsilon^2 n^2}\sum_{k=1}^n \frac{\partial}{\partial\theta_{i_1}}b(X_{t_{k-1}},\theta_0)^*\Xi_{k-1}^{-1}(\theta_0)\frac{\partial}{\partial\theta_{i_2}}b(X_{t_{k-1}},\theta_0) \\
&+\sum_{k=1}^n\left\{R\left(\frac{1}{\varepsilon^2 n^3},X_{t_{k-1}}\right)+R\left(\frac{1}{\varepsilon^4 n^4},X_{t_{k-1}}\right)\right\} \\
\to \;& 4\mathcal{I}_b^{i_1 i_2}(\theta_0)
\end{aligned}
$$

in P_{θ_0}-probability as $\varepsilon\to 0$ and $n\to\infty$. Using Lemma 5.1–(7) and (9), one has

$$
\begin{aligned}
&\sum_{k=1}^n E_{\theta_0}[\eta_k^{i_1}\eta_k^{i_2}(\theta_0)|\mathcal{G}_{k-1}^n] \\
= \;& \frac{1}{n}\sum_{k=1}^n\left\{\mathrm{tr}\left[\frac{\partial}{\partial\theta_{i_1}}\Xi_{k-1}(\theta_0)\Xi_{k-1}^{-1}(\theta_0)\right]\mathrm{tr}\left[\frac{\partial}{\partial\theta_{i_2}}\Xi_{k-1}(\theta_0)\Xi_{k-1}^{-1}(\theta_0)\right]\right. \\
&-\mathrm{tr}\left[\frac{\partial}{\partial\theta_{i_1}}\Xi_{k-1}(\theta_0)\Xi_{k-1}^{-1}(\theta_0)\right]\mathrm{tr}\left[\frac{\partial}{\partial\theta_{i_2}}\Xi_{k-1}(\theta_0)\Xi_{k-1}^{-1}(\theta_0)\right] \\
&-\mathrm{tr}\left[\frac{\partial}{\partial\theta_{i_2}}\Xi_{k-1}(\theta_0)\Xi_{k-1}^{-1}(\theta_0)\right]\mathrm{tr}\left[\frac{\partial}{\partial\theta_{i_1}}\Xi_{k-1}(\theta_0)\Xi_{k-1}^{-1}(\theta_0)\right] \\
&+\mathrm{tr}\left[\left(\frac{\partial}{\partial\theta_{i_1}}\Xi_{k-1}(\theta_0)\right)\Xi_{k-1}^{-1}(\theta_0)\right]\mathrm{tr}\left[\left(\frac{\partial}{\partial\theta_{i_2}}\Xi_{k-1}(\theta_0)\right)\Xi_{k-1}^{-1}(\theta_0)\right] \\
&\left.+2\mathrm{tr}\left[\left(\frac{\partial}{\partial\theta_{i_1}}\Xi_{k-1}(\theta_0)\right)\Xi_{k-1}^{-1}(\theta_0)\left(\frac{\partial}{\partial\theta_{i_2}}\Xi_{k-1}(\theta_0)\right)\Xi_{k-1}^{-1}(\theta_0)\right]\right\} \\
&+\sum_{k=1}^n\left\{R\left(\frac{1}{n^2},X_{t_{k-1}}\right)+R\left(\frac{1}{\varepsilon^2 n^3},X_{t_{k-1}}\right)+R\left(\frac{1}{\varepsilon^4 n^4},X_{t_{k-1}}\right)\right\} \\
\to \;& 4\mathcal{I}_\sigma^{i_1 i_2}(\theta_0)
\end{aligned}
$$

in P_{θ_0}-probability as $\varepsilon\to 0$ and $n\to\infty$. By Lemma 5.1–(6) and (9), we obtain

$$
\begin{aligned}
&\sum_{k=1}^n E_{\theta_0}[\xi_k^i\eta_k^j(\theta_0)|\mathcal{G}_{k-1}^n] \\
= \;& -\frac{2}{\varepsilon^2 n}\sum_{k=1}^n\sum_{l_1=1}^d\left[\left(\frac{\partial}{\partial\theta_i}b(X_{t_{k-1}},\theta_0)\right)^*\Xi_{k-1}^{-1}(\theta_0)\right]^{l_1}\frac{\partial}{\partial\theta_j}\log\det\Xi_{k-1}(\theta_0)E_{\theta_0}[P_k^{l_1}(\theta_0)|\mathcal{G}_{k-1}^n] \\
&-\frac{2}{\varepsilon^4}\sum_{k=1}^n\sum_{l_1,l_2,l_3=1}^d\left[\left(\frac{\partial}{\partial\theta_i}b(X_{t_{k-1}},\theta_0)\right)^*\Xi_{k-1}^{-1}(\theta_0)\right]^{l_1}\left(\frac{\partial}{\partial\theta_j}\Xi_{k-1}^{-1}(\theta_0)\right)^{l_2 l_3} \\
&\times E_{\theta_0}[P_k^{l_1}P_k^{l_2}P_k^{l_3}(\theta_0)|\mathcal{G}_{k-1}^n] \\
= \;& \sum_{k=1}^n\left\{R\left(\frac{1}{n^3},X_{t_{k-1}}\right)+R\left(\frac{1}{\varepsilon^2 n^3},X_{t_{k-1}}\right)+R\left(\frac{1}{\varepsilon^4 n^4},X_{t_{k-1}}\right)\right\}\to 0
\end{aligned}
$$

in P_{θ_0}-probability as $\varepsilon\to 0$ and $n\to\infty$. □

Proof of (15). Using Lemma 5.1–(iv), one has

$$\begin{aligned}
&\sum_{k=1}^{n} E_{\theta_0}[(\xi_k^i(\theta_0))^4|\mathcal{G}_{k-1}^n] \\
= &\frac{16}{\varepsilon^8 n^2}\sum_{k=1}^{n}\sum_{l_1,l_2,l_3,l_4=1}^{d}\left[\left(\frac{\partial}{\partial\theta_i}b(X_{t_{k-1}},\theta_0)\right)^*\Xi_{k-1}^{-1}(\theta_0)\right]^{l_1}\left[\left(\frac{\partial}{\partial\theta_i}b(X_{t_{k-1}},\theta_0)\right)^*\Xi_{k-1}^{-1}(\theta_0)\right]^{l_2} \\
&\times\left[\left(\frac{\partial}{\partial\theta_i}b(X_{t_{k-1}},\theta_0)\right)^*\Xi_{k-1}^{-1}(\theta_0)\right]^{l_3}\left[\left(\frac{\partial}{\partial\theta_i}b(X_{t_{k-1}},\theta_0)\right)^*\Xi_{k-1}^{-1}(\theta_0)\right]^{l_4} \\
&\times E_{\theta_0}[P_k^{l_1}P_k^{l_2}P_k^{l_3}P_k^{l_4}(\theta_0)|\mathcal{G}_{k-1}^n] \\
= &\sum_{k=1}^{n}\left\{R\left(\frac{1}{\varepsilon^4 n^4},X_{t_{k-1}}\right)+R\left(\frac{1}{\varepsilon^6 n^6},X_{t_{k-1}}\right)+R\left(\frac{1}{\varepsilon^8 n^7},X_{t_{k-1}}\right)\right\}\to 0
\end{aligned}$$

in P_{θ_0}-probability as $\varepsilon\to 0$ and $n\to\infty$. We obtain

$$\begin{aligned}
(\eta_k^i(\theta_0))^4 \quad\le\quad & 2^3\left[\frac{1}{n^2}\left(\frac{\partial}{\partial\theta_i}\log\det\Xi_{k-1}(\theta_0)\right)^4\right. \\
&\left.+(2d)^3\frac{n^2}{\varepsilon^8}\sum_{l_1 l_2=1}^{d}\left[\left(\frac{\partial}{\partial\theta_i}\Xi_{k-1}^{-1}(\theta_0)\right)^{l_1 l_2}\right]^4(P_k^{l_1}P_k^{l_2}(\theta_0))^4\right].
\end{aligned}$$

In the same way as Lemma 5.1, we have

$$\begin{aligned}
E_{\theta_0}[(P_k^{l_1}P_k^{l_2})^4(\theta_0)|\mathcal{G}_{k-1}^n] \quad=\quad & R\left(\frac{\varepsilon^8}{n^4},X_{t_{k-1}}\right)+R\left(\frac{\varepsilon^6}{n^5},X_{t_{k-1}}\right)+R\left(\frac{\varepsilon^4}{n^6},X_{t_{k-1}}\right) \\
&+R\left(\frac{\varepsilon^2}{n^7},X_{t_{k-1}}\right)+R\left(\frac{1}{n^8},X_{t_{k-1}}\right).
\end{aligned}$$

Thus, one has

$$\begin{aligned}
\sum_{k=1}^{n} E_{\theta_0}[(\eta_k^i(\theta_0))^4|\mathcal{G}_{k-1}^n] \quad\le\quad & \sum_{k=1}^{n}\left\{R\left(\frac{1}{n^2},X_{t_{k-1}}\right)+R\left(\frac{1}{\varepsilon^2 n^3},X_{t_{k-1}}\right)+R\left(\frac{1}{\varepsilon^4 n^4},X_{t_{k-1}}\right)\right. \\
&\left.+R\left(\frac{1}{\varepsilon^6 n^5},X_{t_{k-1}}\right)+R\left(\frac{1}{\varepsilon^8 n^6},X_{t_{k-1}}\right)\right\}\to 0
\end{aligned}$$

in P_{θ_0}-probability as $\varepsilon\to 0$ and $n\to\infty$. This completes the proof. □

Proof of Theorem 3.4. We begin by showing the consistency of $\hat{\theta}_{\varepsilon,n}$. From a version of Lemma 17 in Genon-Catalot and Jacod (1993), one has

$$\log\det[\sigma\sigma^*](X_t^0,\theta)+\mathrm{tr}\left[[\sigma\sigma^*](X_t^0,\theta_0)[\sigma\sigma^*]^{-1}(X_t^0,\theta)\right]\ge\log\det[\sigma\sigma^*](X_t^0,\theta_0)+d$$

with equality if and only if $[\sigma\sigma^*](X_t^0,\theta)=[\sigma\sigma^*](X_t^0,\theta_0)$. By (iv) in Assumption 2.1, we obtain

$$\int_0^1 B^*(X_s^0,\theta_0,\theta)[\sigma\sigma^*]^{-1}(X_s^0,\theta)B(X_s^0,\theta_0,\theta)ds\ge 0$$

with equality if and only if $b(X_t^0, \theta) = b(X_t^0, \theta_0)$. Thus, it follows from Assumption 2.3 that $U(\theta, \theta_0) \geq U(\theta_0, \theta_0)$ with equality if and only if $\theta = \theta_0$. Therefore, for any $\eta > 0$,

$$\inf_{\theta:|\theta-\theta_0|\geq\eta} U(\theta, \theta_0) > U(\theta_0, \theta_0). \tag{16}$$

Moreover, it follows from the definition of $\hat{\theta}_{\varepsilon,n}$ and $\theta_0 \in \bar{\Theta}$ that for any $\eta > 0$,

$$P_{\theta_0}\left[\bar{U}_{\varepsilon,n}(\hat{\theta}_{\varepsilon,n}) \leq \bar{U}_{\varepsilon,n}(\theta_0) + \eta\right] \to 1 \tag{17}$$

as $\varepsilon \to 0$ and $n \to \infty$, where $\bar{U}_{\varepsilon,n}(\theta) = \frac{1}{n}U_{\varepsilon,n}(\theta)$. From (16), for every $\eta > 0$, there exists $\eta' > 0$ such that

$$\inf_{\theta:|\theta-\theta_0|\geq\eta} U(\theta, \theta_0) > U(\theta_0, \theta_0) + \eta'.$$

Furthermore, for every $\eta > 0$ there exists $\eta' > 0$ such that

$$|\hat{\theta}_{\varepsilon,n} - \theta_0| \geq \eta \Rightarrow U(\hat{\theta}_{\varepsilon,n}, \theta_0) \geq \inf_{\theta:|\theta-\theta_0|\geq\eta} U(\theta, \theta_0) > U(\theta_0, \theta_0) + \eta'.$$

Thus, one has

$$\begin{aligned}
P_{\theta_0}\left[|\hat{\theta}_{\varepsilon,n} - \theta_0| \geq \eta\right] &\leq P_{\theta_0}\left[U(\hat{\theta}_{\varepsilon,n}, \theta_0) > U(\theta_0, \theta_0) + \eta'\right] \\
&\leq P_{\theta_0}\left[\left|U(\hat{\theta}_{\varepsilon,n}, \theta_0) - \bar{U}_{\varepsilon,n}(\hat{\theta}_{\varepsilon,n})\right| \geq \frac{\eta'}{3}\right] \\
&\quad + P_{\theta_0}\left[\bar{U}_{\varepsilon,n}(\hat{\theta}_{\varepsilon,n}) - \bar{U}_{\varepsilon,n}(\theta_0) \geq \frac{\eta'}{3}\right] \\
&\quad + P_{\theta_0}\left[\left|\bar{U}_{\varepsilon,n}(\theta_0) - U(\theta_0, \theta_0)\right| \geq \frac{\eta'}{3}\right] \\
&\leq 2P_{\theta_0}\left[\sup_{\theta\in\bar{\Theta}}\left|\bar{U}_{\varepsilon,n}(\theta) - U(\theta, \theta_0)\right| \geq \frac{\eta'}{3}\right] \\
&\quad + P_{\theta_0}\left[\bar{U}_{\varepsilon,n}(\hat{\theta}_{\varepsilon,n}) \geq \bar{U}_{\varepsilon,n}(\theta_0) + \frac{\eta'}{3}\right] \\
&\to 0
\end{aligned}$$

as $\varepsilon \to 0$ and $n \to \infty$, where the last estimate is based on Proposition 3.1 and (17). This completes the proof of the consistency of $\hat{\theta}_{\varepsilon,n}$.

Using the consistency of $\hat{\theta}_{\varepsilon,n}$ and Propositions 3.2–3.3, we can show the asymptotic normality of $\hat{\theta}_{\varepsilon,n}$ along the same lines as the proof of the asymptotic normality of Theorem 1 in Sørensen and Uchida (2003). This completes the proof. □

Acknowledgement

The author wishes to thank the referee, the associate editor and the editor for their valuable comments. This work was in part supported by the Research Fund for Scientists of the Ministry of Education, Culture, Sports, Science and Technology, and by Cooperative Program of the Institute of Statistical Mathematics. The author would like to thank Professor Takashi Yanagawa for a lively discussion on diffusion models. His interest in statistical inference for semimartingales gave me a great encouragement.

References

Bibby, B. M. and Sørensen, M. (1995). Martingale estimating functions for discretely observed diffusion processes, *Bernoulli*, **1**, 17–39.

Deelstra, G. and Delbaen, F. (1999). Convergence of discreted stochastic (interest rate) processes with drift term, preprint.

Dermoune, A. and Kutoyants, Yu. A. (1995). Expansion of distribution function of maximum likelihood estimate for misspecified diffusion type observations, *Stoch. Stoch. Rep.*, **52**, 121–145.

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

Genon-Catalot, V. (1990). Maximum contrast estimation for diffusion processes from discrete observations, *Statistics*, **21**, 99–116.

Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multidimensional diffusion processes, *Ann. Inst. Henri Poincaré Probab. Statist.*, **29**, 119–151.

Hall, P. and Heyde, C. (1980). *Martingale limit theory and its applications*, Academic Press, New York.

Jacobsen, M. (2001). Discretely observed diffusions; classes of estimating functions and small Δ-optimality, *Scand. J. Statist.*, **28**, 123–150.

Jacobsen, M. (2002). Optimality and small Δ-optimality of martingale estimating functions, *Bernoulli*, **8**, 643–668.

Kanagawa, S. and Ogawa, S. (2001). Numerical solutions of stochastic differential equations. II. Applications, (Japanese) *Sugaku*, **53**, 125–138.

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

Kessler, M. (2000). Simple and explicit estimating functions for a discretely observed diffusion process, *Scand. J. Statist.*, **27**, 65–82.

Kim, Y. J. and Kunitomo, N. (1999). Pricing options under stochastic interest rates, *Asia Pacific Financial Markets*, **6**, 49–70.

Kloeden, P. E. and Platen, E. (1992). *Numerical solution of stochastic differential equations*, Springer-Verlag, New York.

Kunitomo, N. and Takahashi, A. (2001). The asymptotic expansion approach to the valuation of interest rate contingent claims, *Math. Finance*, **11**, 117–151.

Kutoyants, Yu. A. (1984). *Parameter estimation for stochastic processes*, Prakasa Rao, B.L.S. (ed.), Heldermann, Berlin.

Kutoyants, Yu. A. (1994). *Identification of dynamical systems with small noise*, Kluwer, Dordrecht.

Laredo, C. F. (1990). A sufficient condition for asymptotic sufficiency of incomplete observations of a diffusion process, *Ann. Statist.*, **18**, 1158–1171.

Sakamoto, Y. and Yoshida, N. (1996). Expansion of perturbed random variables based on generalized Wiener functionals, *J. Multivariate Anal.*, **59**, 34–59.

Sørensen, M. (2000). Small dispersion asymptotics for diffusion martingale estimating functions, Preprint No. 2000-2, Department of Statistics and Operations Research, University of Copenhagen.

Sørensen, M. and Uchida, M. (2003). Small diffusion asymptotics for discretely sampled stochastic differential equations, *Bernoulli*, **9**, 1051–1069.

Takahashi, A. (1999). An asymptotic expansion approach to pricing contingent claims, *Asia Pacific Financial Markets*, **6**, 115–151.

Takahashi, A. and Yoshida, N. (2004). Asymptotic expansion scheme for the optimal portfolio for investment, to appear in *Statist. Infer. Stochast. Process.*

Uchida, M. (2003). Estimation for dynamical systems with small noise from discrete observations, *J. Japan Statist. Soc.*, **33**, 157–167.

Uchida, M. (2004). Estimation for discretely observed small diffusions based on approximate martingale estimating functions, to appear in *Scand. J. Statist.*

Uchida, M. and Yoshida, N. (2004). Information criteria for small diffusions via the theory of Malliavin-Watanabe, *Statist. Infer. Stochast. Process.*, **7**, 35–67.

Uchida, M. and Yoshida, N. (2004). Asymptotic expansion for small diffusions applied to option pricing, to appear in *Statist. Infer. Stochast. Process.*

Yoshida, N. (1992a). Asymptotic expansion of maximum likelihood estimators for small diffusions via the theory of Malliavin-Watanabe, *Probab. Theory Relat. Fields*, **92**, 275–311.

Yoshida, N. (1992b). Asymptotic expansion for statistics related to small diffusions, *J. Japan Statist. Soc.*, **22**, 139–159.

Yoshida, N. (1992c). Estimation for diffusion processes from discrete observation, *J. Multivariate Anal.*, **41**, 220–242.

Yoshida, N. (1993). Asymptotic expansion of Bayes estimators for small diffusions, *Probab. Theory Relat. Fields*, **95**, 429–450.

Yoshida, N. (1996). Asymptotic expansions for perturbed systems on Wiener space: maximum likelihood estimators, *J. Multivariate Anal.*, **57**, 1–36.

Yoshida, N. (2003). Conditional expansions and their applications, *Stochastic Process. Appl.*, **107**, 53–81.

Received July 25, 2003
Revised April 7, 2004