Bernoulli **9**(6), 2003, 1051–1069

Small-diffusion asymptotics for discretely sampled stochastic differential equations

MICHAEL SØRENSEN[1] and MASAYUKI UCHIDA[2]

[1]*Department of Applied Mathematics and Statistics, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark. E-mail: michael@math.ku.dk*
[2]*Faculty of Mathematics, Kyushu University, Ropponmatsu, Fukuoka 810-8560, Japan. E-mail: uchida@math.kyushu-u.ac.jp*

The minimum-contrast estimation of drift and diffusion coefficient parameters for a multidimensional diffusion process with a small dispersion parameter ε based on a Gaussian approximation to the transition density is presented when the sample path is observed at equidistant times k/n, $k = 0, 1, \ldots, n$. We study asymptotic results for the minimum-contrast estimator as ε goes to 0 and n goes to ∞ simultaneously.

Keywords: diffusion process with small noise; discrete-time observation; minimum-contrast estimation; parametric inference

1. Introduction

In this paper we consider a family of d-dimensional diffusion processes defined by the stochastic differential equations

$$dX_t = b(X_t, \alpha)dt + \varepsilon\sigma(X_t, \beta)dw_t, \qquad t \in [0, 1], \varepsilon \in (0, 1], \tag{1}$$

$$X_0 = x_0,$$

where $(\alpha, \beta) \in \overline{\Theta}_\alpha \times \overline{\Theta}_\beta$, with Θ_α and Θ_β being open bounded convex subsets of $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively. Furthermore, x_0 and ε are known constants, b is an $\mathbb{R}^d$-valued function defined on $\mathbb{R}^d \times \overline{\Theta}_\alpha$, σ is an $(\mathbb{R}^d \otimes \mathbb{R}^r)$-valued function defined on $\mathbb{R}^d \times \overline{\Theta}_\beta$, 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 parameters α and β. Our data are discrete observations of X at n regularly spaced time points $t_k = k/n$ on the fixed interval [0, 1], that is, $(X_{t_k})_{0 \leq k \leq n}$. We are interested in estimating α and β based on these observations. The type of asymptotics we consider is when ε goes to 0 and n goes to ∞ simultaneously.

For the case where the whole path $X = \{X_t;\ t \in [0, 1]\}$ is observed, parametric inference for diffusion-type processes with small noise is well developed. The first-order asymptotic statistical theory has been studied mainly by Kutoyants (1984; 1994). As for higher-order asymptotics, Yoshida (1992a) showed the validity of asymptotic expansions for statistical estimators by means of Malliavin calculus with truncation; see also Yoshida (1993; 1996; 2003), Dermoune and Kutoyants (1995), Sakamoto and Yoshida (1996), and Uchida and

Yoshida (1999). In recent years, the more realistic case of parametric estimation for discretely observed diffusion processes has also been studied by many researchers; see Dacunha-Castelle and Florens-Zmirou (1986), Florens-Zmirou (1989), Yoshida (1992c), Genon-Catalot and Jacod (1993), Bibby and Sørensen (1995; 2001), Hansen and Scheinkman (1995), Kessler (1997; 2000), Sørensen (1997), Kessler and Sørensen (1999), Jacobsen (2001) and H. Sørensen (2001).

Although there have been many applications of small-diffusion asymptotics (for applications in mathematical finance, see Yoshida (1992b), Kim and Kunitomo (1999), Takahashi (1999), Kunitomo and Takahashi (2001), Uchida and Yoshida (2001)), very little work has been done on small-noise asymptotics for estimation for diffusion processes from discrete-time observations. Genon-Catalot (1990) and Laredo (1990) studied the efficient estimation of drift parameters of a diffusion process with small noise from discrete observations under the assumptions that diffusion coefficients are known and the asymptotics is when $\varepsilon \to 0$ and $n \to \infty$. Sørensen (2000) presented martingale estimation functions for discretely observed diffusion processes with small noise, and showed consistency and asymptotic normality of the estimators of drift and diffusion coefficient parameters when $\varepsilon \to 0$ and n is fixed. Following on from the three papers last mentioned, our goal is to obtain a consistent, asymptotically normal and asymptotically efficient estimator of (α, β) in our setting.

This paper is organized as follows. In Section 2, we introduce a contrast function based on a Gaussian approximation to the transition density and state several preliminary lemmas. Section 3 presents our main result about the consistency, asymptotic normality and asymptotic efficiency of the minimum-contrast estimator obtained from the contrast function constructed in Section 2. Section 4 is devoted to proving the results stated in the previous sections.

2. The contrast function and preliminary lemmas

We begin by describing the notation and assumptions used in this paper. Suppose that the parameter and the parameter space can be decomposed as follows: $\theta = (\alpha, \beta)$ and $\Theta = \Theta_\alpha \times \Theta_\beta$. Let α_0, β_0 and θ_0 denote the true values of α, β and θ, respectively. Let X_t^0 be the solution of the ordinary differential equation corresponding to $\varepsilon = 0$, i.e. $\mathrm{d}X_t^0 = b(X_t^0, \alpha_0)\mathrm{d}t$, $X_0^0 = x_0$. For a matrix A, $|A|^2 = \mathrm{tr}(AA^{\mathrm{T}})$. We denote by $\overline{C}_{\uparrow}^{k_1,k_2}(\mathbb{R}^d \times \Theta; \mathbb{R}^m)$ the space of all functions f satisfying the following two conditions:

(i) $f(x, \theta)$ is an $\mathbb{R}^m$-valued function on $\mathbb{R}^d \times \Theta$ that is continuously differentiable with respect to x and θ up to order k_1 and k_2 respectively;
(ii) for $|\mathbf{n}| = 0, 1, \ldots, k_1$ and $|\nu| = 0, 1, \ldots, k_2$, there exists $C > 0$ such that $\sup_{\theta\in\Theta}|\delta^\nu \partial^{\mathbf{n}} f| \leqslant C(1 + |x|)^C$ for all x.

Here $\mathbf{n} = (n_1, \ldots, n_d)$ and $\nu = (\nu_1, \ldots, \nu_l)$ are multi-indices, $l = \dim(\Theta)$, $|\mathbf{n}| = n_1 + \cdots + n_d$, $|\nu| = \nu_1 + \cdots + \nu_l$, $\partial^{\mathbf{n}} = \partial_1^{n_1} \cdots \partial_d^{n_d}$, $\partial_i = \partial/\partial x^i$, $i = 1, \ldots, d$, $\delta^\nu = \delta_1^{\nu_1} \cdots \delta_l^{\nu_l}$, $\delta_j = \partial/\partial\theta^j$, $j = 1, \ldots, l$. Note that ν and δ depend on Θ. For example, $\nu = (\nu_1, \ldots, \nu_p)$, $\delta_j = \partial/\partial\alpha^j$ for $\overline{\Theta}_\alpha$.

In this paper, we make the following assumptions:

(A1) Equation (1) has a unique strong solution in [0, 1].
(A2) For all $m > 0$, $\sup_t \mathrm{E}[|X_t|^m] < \infty$.
(A3) $b(x, \alpha) \in \bar{C}_{\uparrow}^{14,3}(\mathbb{R}^d \times \bar{\Theta}_\alpha; \mathbb{R}^d)$, $\sigma(x, \beta) \in \bar{C}_{\uparrow}^{14,3}(\mathbb{R}^d \times \bar{\Theta}_\beta; \mathbb{R}^d \otimes \mathbb{R}^r)$.
(A4) $\inf_{x,\beta} \det[\sigma\sigma^{\mathrm{T}}](x, \beta) > 0$, $[\sigma\sigma^{\mathrm{T}}]^{-1}(x, \beta) \in \bar{C}_{\uparrow}^{1,3}(\mathbb{R}^d \times \bar{\Theta}_\beta; \mathbb{R}^d \otimes \mathbb{R}^d)$.
(A5) $\alpha \neq \alpha_0 \Rightarrow b(X_t^0, \alpha) \neq b(X_t^0, \alpha_0)$ for at least one value of t, and $\beta \neq \beta_0 \Rightarrow \sigma\sigma^{\mathrm{T}}(X_t^0, \beta) \neq \sigma\sigma^{\mathrm{T}}(X_t^0, \beta_0)$ for at least one value of t.

Remark 1. (i) For (A1), there are several well-known types of sufficient conditions for the existence and uniqueness of a solution of equation (1). For more details, see Ikeda and Watanabe (1989, Chapter IV).

(ii) To obtain the results in this paper, assumptions (A3) and (A4) can be relaxed. By using a 'classical' localization argument, they can be replaced by mild regularity conditions about b and σ in the neighbourhood of the path of X_t^0.

Moreover, the following conditions for ε and n are assumed:

(B1) $\lim_{\varepsilon\to 0, n\to\infty} (\varepsilon n)^{-1} = 0$.
(B2) $\lim_{\varepsilon\to 0, n\to\infty} (\varepsilon\sqrt{n})^{-1} < \infty$.

Let P_θ be the law of the solution of (1), and L_θ the infinitesimal generator of the diffusion (1):

$$L_\theta f(x) = \sum_{i=1}^d b^i(x, \alpha)\partial_i f(x) + \frac{1}{2}\varepsilon^2 \sum_{i,j=1}^d [\sigma\sigma^{\mathrm{T}}]^{i,j}(x, \beta)\partial_i\partial_j f(x).$$

In order to construct the contrast function, it is natural to consider a Gaussian approximation to the transition density in the same way as in Kessler (1997). Using Lemma 1 in Florens-Zmirou (1989), we obtain the contrast function

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

where

$$P_k(\alpha) = X_{t_k} - X_{t_{k-1}} - \frac{1}{n} b(X_{t_{k-1}}, \alpha), \qquad \Xi_k(\beta) = [\sigma\sigma^{\mathrm{T}}](X_{t_k}, \beta).$$

Let R denote a function $(0, 1] \times \mathbb{R}^d \to \mathbb{R}$ for which there exists a constant C such that $|R(a, x)| \leqslant aC(1 + |x|)^C$ for all a, x. We define $\mathcal{G}_k^n = \sigma(w_s;\, s \leqslant t_k)$, $B_k^i(\alpha_0, \alpha) = b^i(X_{t_k}, \alpha_0) - b^i(X_{t_k}, \alpha)$, and $B(x, \alpha_0, \alpha) = b(x, \alpha_0) - b(x, \alpha)$. Moreover, in order to formulate the preliminary lemmas given later, we need the following functions and notation. For Lemma 4, we define

$$U_1(\alpha, \alpha_0, \beta) = \int_0^1 B^{\mathrm{T}}(X_s^0, \alpha_0, \alpha)[\sigma\sigma^{\mathrm{T}}]^{-1}(X_s^0, \beta)B(X_s^0, \alpha_0, \alpha)\mathrm{d}s,$$

$$U_2(\alpha, \beta, \beta_0) = \int_0^1 \log\det[\sigma\sigma^{\mathrm{T}}](X_s^0, \beta)\mathrm{d}s + \int_0^1 \mathrm{tr}[[\sigma\sigma^{\mathrm{T}}](X_s^0, \beta_0)[\sigma\sigma^{\mathrm{T}}]^{-1}(X_s^0, \beta)]\mathrm{d}s$$
$$+ M^2\int_0^1 B^{\mathrm{T}}(X_s^0, \alpha_0, \alpha)[\sigma\sigma^{\mathrm{T}}]^{-1}(X_s^0, \beta)B(X_s^0, \alpha_0, \alpha)\mathrm{d}s,$$

where $M = \lim_{\varepsilon\to 0, n\to\infty}(\varepsilon\sqrt{n})^{-1}$. Note that U_2 is only well defined under assumption (B2). For Lemma 5, let

$$C_{\varepsilon,n}(\theta_0) = \begin{pmatrix} \varepsilon^2\left(\dfrac{\partial^2}{\partial\alpha_i\partial\alpha_j}U_{\varepsilon,n}(\theta_0)\right)_{1\leqslant i,j\leqslant p} & \varepsilon\dfrac{1}{\sqrt{n}}\left(\dfrac{\partial^2}{\partial\alpha_i\partial\beta_j}U_{\varepsilon,n}(\theta_0)\right)_{1\leqslant i\leqslant p, 1\leqslant j\leqslant q} \\ \varepsilon\dfrac{1}{\sqrt{n}}\left(\dfrac{\partial^2}{\partial\beta_i\partial\alpha_j}U_{\varepsilon,n}(\theta_0)\right)_{1\leqslant i\leqslant q, 1\leqslant j\leqslant p} & \dfrac{1}{n}\left(\dfrac{\partial^2}{\partial\beta_i\partial\beta_j}U_{\varepsilon,n}(\theta_0)\right)_{1\leqslant i,j\leqslant q} \end{pmatrix}$$

and

$$I(\theta_0) = \begin{pmatrix} (I_b^{i,j}(\theta_0))_{1\leqslant i,j\leqslant p} & 0 \\ 0 & (I_\sigma^{i,j}(\theta_0))_{1\leqslant i,j\leqslant q} \end{pmatrix},$$

where

$$I_b^{i,j}(\theta_0) = \int_0^1 \left(\frac{\partial}{\partial\alpha_i}b(X_s^0, \alpha_0)\right)^{\mathrm{T}}[\sigma\sigma^{\mathrm{T}}]^{-1}(X_s^0, \beta_0)\left(\frac{\partial}{\partial\alpha_j}b(X_s^0, \alpha_0)\right)\mathrm{d}s,$$

$$I_\sigma^{i,j}(\theta_0) = \frac{1}{2}\int_0^1 \mathrm{tr}\left[\left(\frac{\partial}{\partial\beta_i}[\sigma\sigma^{\mathrm{T}}]\right)[\sigma\sigma^{\mathrm{T}}]^{-1}\left(\frac{\partial}{\partial\beta_j}[\sigma\sigma^{\mathrm{T}}]\right)[\sigma\sigma^{\mathrm{T}}]^{-1}(X_s^0, \beta_0)\right]\mathrm{d}s.$$

For Lemma 6, define

$$\Lambda_{\varepsilon,n} = \begin{pmatrix} -\varepsilon\left(\dfrac{\partial}{\partial\alpha_i}U_{\varepsilon,n}(\theta_0)\right)_{1\leqslant i\leqslant p} \\ -\dfrac{1}{\sqrt{n}}\left(\dfrac{\partial}{\partial\beta_j}U_{\varepsilon,n}(\theta_0)\right)_{1\leqslant j\leqslant q} \end{pmatrix}.$$

Lemma 1. *Suppose that assumptions* (A1)–(A3) *hold. Then,*

$$\mathrm{E}_{\theta_0}[P^i_k(\alpha_0)|\mathcal{G}^n_{k-1}] = R\left(\frac{1}{n^2}, X_{t_{k-1}}\right),$$

$$\mathrm{E}_{\theta_0}[P^{i_1}_k(\alpha_0)P^{i_2}_k(\alpha_0)|\mathcal{G}^n_{k-1}] = \frac{\varepsilon^2}{n}\Xi^{i_1 i_2}_{k-1}(\beta_0) + R\left(\frac{\varepsilon^2}{n^2}, X_{t_{k-1}}\right) + R\left(\frac{1}{n^3}, X_{t_{k-1}}\right),$$

$$\mathrm{E}_{\theta_0}[P^{i_1}_k(\alpha_0)P^{i_2}_k(\alpha_0)P^{i_3}_k(\alpha_0)|\mathcal{G}^n_{k-1}] = R\left(\frac{\varepsilon^4}{n^2}, X_{t_{k-1}}\right) + R\left(\frac{\varepsilon^2}{n^3}, X_{t_{k-1}}\right) + R\left(\frac{1}{n^4}, X_{t_{k-1}}\right),$$

$$\mathrm{E}_{\theta_0}\left[\prod_{j=1}^4 P^{i_j}_k(\alpha_0)|\mathcal{G}^n_{k-1}\right] = \frac{\varepsilon^4}{n^2}\left\{\Xi^{i_1 i_2}_{k-1}\Xi^{i_3 i_4}_{k-1}(\beta_0) + \Xi^{i_1 i_3}_{k-1}\Xi^{i_2 i_4}_{k-1}(\beta_0) + \Xi^{i_1 i_4}_{k-1}\Xi^{i_2 i_3}_{k-1}(\beta_0)\right\}$$

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

Lemma 2. *Let* $f \in \overline{C}^{1,1}_{\uparrow}(\mathbb{R}^d \times \overline{\Theta}; \mathbb{R})$. *Assume* (A1)–(A3). *Then, under* P_{θ_0},

(i)

$$\frac{1}{n}\sum_{k=1}^n f(X_{t_{k-1}}, \theta) \to \int_0^1 f(X^0_s, \theta)\mathrm{d}s$$

as $\varepsilon \to 0$ *and* $n \to \infty$, *uniformly in* $\theta \in \overline{\Theta}$, *and*

(ii)

$$\sum_{k=1}^n f(X_{t_{k-1}}, \theta)P^i_k(\alpha_0) \to 0$$

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

Lemma 3. *Let* $f \in \overline{C}^{1,1}_{\uparrow}(\mathbb{R}^d \times \overline{\Theta}; \mathbb{R})$. *Assume* (A1)–(A3) *and* (B1). *Then the following holds:*

(i) *Under* P_{θ_0}, *as* $\varepsilon \to 0$ *and* $n \to \infty$,

$$\varepsilon^{-2}\sum_{k=1}^n f(X_{t_{k-1}}, \theta)P^i_k P^j_k(\alpha_0) \to \int_0^1 f(X^0_s, \theta)[\sigma\sigma^{\mathrm{T}}]^{ij}(X^0_s, \beta_0)\mathrm{d}s,$$

uniformly in $\theta \in \overline{\Theta}$.

(ii) *Moreover, if assumption* (B2) *holds, then, under* P_{θ_0}, *as* $\varepsilon \to 0$ *and* $n \to \infty$,

$$\varepsilon^{-2}\sum_{k=1}^n f(X_{t_{k-1}}, \theta)P^i_k P^j_k(\alpha) \to \int_0^1 f(X^0_s, \theta)[\sigma\sigma^{\mathrm{T}}]^{ij}(X^0_s, \beta_0)\mathrm{d}s,$$

$$+ M^2\int_0^1 f(X^0_s, \theta)B^i B^j(X^0_s, \alpha_0, \alpha)\mathrm{d}s,$$

uniformly in $\theta \in \overline{\Theta}$, *where* $M = \lim_{\varepsilon\to 0, n\to\infty}(\varepsilon\sqrt{n})^{-1}$.

Lemma 4. *Assume* (A1)–(A4). *Then the following holds:*

(i) *Under* P_{θ_0}, *as* $\varepsilon \to 0$ *and* $n \to \infty$,

$$\sup_{\theta \in \overline{\Theta}} |\varepsilon^2 \{U_{\varepsilon,n}(\alpha, \beta) - U_{\varepsilon,n}(\alpha_0, \beta)\} - U_1(\alpha, \alpha_0, \beta)| \to 0.$$

(ii) *Moreover, suppose that* (B2) *holds. Then, under* P_{θ_0}, *as* $\varepsilon \to 0$ *and* $n \to \infty$,

$$\sup_{\theta \in \overline{\Theta}} \left| \frac{1}{n} U_{\varepsilon,n}(\alpha, \beta) - U_2(\alpha, \beta, \beta_0) \right| \to 0.$$

Lemma 5. *Assume* (A1)–(A4) *and* (B2). *Then, under* P_{θ_0}, *as* $\varepsilon \to 0$ *and* $n \to \infty$,

(i)

$$C_{\varepsilon,n}(\theta_0) \to 2I(\theta_0),$$

(ii)

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

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

Lemma 6. *Assume* (A1)–(A4) *and* (B2). *Then*

$$\Lambda_{\varepsilon,n} \to N(0, 4I(\theta_0))$$

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

3. Main result

Let $\hat{\theta}_{\varepsilon,n} = (\hat{\alpha}_{\varepsilon,n}, \hat{\beta}_{\varepsilon,n})$ be a minimum-contrast estimator defined by

$$U_{\varepsilon,n}(\hat{\theta}_{\varepsilon,n}) = \inf_{\theta \in \overline{\Theta}} U_{\varepsilon,n}(\theta). \tag{2}$$

Our main theorem is as follows.

Theorem 1. *Assume* (A1)–(A5) *and* (B2). *Then,*

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

in P_{θ_0}*-probability as* $\varepsilon \to 0$ *and* $n \to \infty$. *Moreover, if* $\theta_0 \in \Theta$ *and* $I(\theta_0)$ *is positive definite,*

$$\begin{pmatrix} \varepsilon^{-1}(\hat{\alpha}_{\varepsilon,n} - \alpha_0) \\ \sqrt{n}(\hat{\beta}_{\varepsilon,n} - \beta_0) \end{pmatrix} \to N(0, I(\theta_0)^{-1})$$

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

Remark 2. (i) Let $P_n^{\alpha,\beta}$ be the restriction of $P_{\alpha,\beta}$ to $\mathcal{F}_n = \sigma(X_{t_k} : 0 \leqslant k \leqslant n)$. In the same way as in Gobet (2001; 2002), under regularity conditions, we can obtain the local asymptotic normality for the likelihoods as follows. For every $u \in \mathbb{R}^p$ and $v \in \mathbb{R}^q$, under P_{θ_0},

$$\log\left(\frac{\mathrm{d}P_n^{\alpha_0+\varepsilon u,\beta_0+v/\sqrt{n}}}{\mathrm{d}P_n^{\alpha_0,\beta_0}}\right)((X_{t_k})_{0\leqslant k\leqslant n}) \to \begin{pmatrix} u \\ v \end{pmatrix}^{\mathrm{T}} \mathcal{N} - \frac{1}{2}\begin{pmatrix} u \\ v \end{pmatrix}^{\mathrm{T}} I(\theta_0)\begin{pmatrix} u \\ v \end{pmatrix}$$

in distribution as $\varepsilon \to 0$ and $n \to \infty$, where $\mathcal{N}$ is a centred Gaussian variable with covariance matrix $I(\theta_0)$. For details, see Uchida (2002). If $I(\theta_0)$ is non-singular, it follows from minimax theorems that $I(\theta_0)^{-1}$ gives the lower bound for the asymptotic variance of regular estimators. This, together with Theorem 1, shows that the estimator given by (2) is asymptotically efficient.

(ii) It is worth mentioning that the estimators of the drift and diffusion coefficient parameters in Theorem 1 are asymptotically independent.

(iii) Note also that when $(\varepsilon\sqrt{n})^{-1} \to 0$ the rate of convergence is different for drift and diffusion coefficient parameters. The estimator of the diffusion coefficient parameter converges more quickly than the estimator of the drift parameter because it utilizes information about the diffusion coefficient in the fine structure of the continuous sample path.

When $\sigma(x, \beta) = \sigma(x)$, Theorem 1 holds under assumption (B1) instead of (B2). Let $C_\uparrow^k(\mathbb{R}^d; \mathbb{R}^d \otimes \mathbb{R}^r)$ be the set of all functions f of class $C^k(\mathbb{R}^d; \mathbb{R}^d \otimes \mathbb{R}^r)$ such that f and its first k derivatives have polynomial growth. Instead of assumptions (A3)–(A5), we make the following assumptions:

(A3′) $b(x, \alpha) \in \overline{C}_\uparrow^{4,3}(\mathbb{R}^d \times \overline{\Theta}_\alpha; \mathbb{R}^d)$, $\sigma(x) \in C_\uparrow^4(\mathbb{R}^d; \mathbb{R}^d \otimes \mathbb{R}^r)$.
(A4′) $\inf_x \det[\sigma\sigma^{\mathrm{T}}](x) > 0$, $[\sigma\sigma^{\mathrm{T}}]^{-1}(x) \in C_\uparrow^1(\mathbb{R}^d; \mathbb{R}^d \otimes \mathbb{R}^d)$.
(A5′) $\alpha \neq \alpha_0 \Rightarrow b(X_t^0, \alpha) \neq b(X_t^0, \alpha_0)$ for at least one value of t.

Set

$$\tilde{I}_b(\alpha_0) = (\tilde{I}_b^{i,j}(\alpha_0))_{1\leqslant i,j\leqslant p}, \qquad \tilde{I}_b^{i,j}(\alpha_0) = \int_0^1 \left(\frac{\partial}{\partial\alpha_i} b(X_s^0, \alpha_0)\right)^{\mathrm{T}} [\sigma\sigma^{\mathrm{T}}]^{-1}(X_s^0)\left(\frac{\partial}{\partial\alpha_j} b(X_s^0, \alpha_0)\right)\mathrm{d}s.$$

We consider the contrast function

$$\tilde{U}_{\varepsilon,n}(\alpha) = \varepsilon^{-2} n \sum_{k=1}^n P_k^{\mathrm{T}}(\alpha)[\sigma\sigma^{\mathrm{T}}]^{-1}(X_{t_{k-1}})P_k(\alpha),$$

and let $\hat{\alpha}_{\varepsilon,n}$ be a minimum-contrast estimator defined by

$$\tilde{U}_{\varepsilon,n}(\hat{\alpha}_{\varepsilon,n}) = \inf_{\alpha\in\overline{\Theta}_\alpha} \tilde{U}_{\varepsilon,n}(\alpha).$$

Corollary 1. *Suppose* $\sigma(x, \beta) = \sigma(x)$*, and assume* (A1), (A2), (A3′)–(A5′) *and* (B1). *Then*

$$\hat{\alpha}_{\varepsilon,n} \to \alpha_0$$

in P_{α_0}-probability, as $\varepsilon \to 0$ and $n \to \infty$. Moreover, if $\alpha_0 \in \Theta_\alpha$ and $\tilde{I}_b(\alpha_0)$ is positive definite, then

$$\varepsilon^{-1}(\hat{\alpha}_{\varepsilon,n} - \alpha_0) \to N(0, \tilde{I}_b(\alpha_0)^{-1})$$

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

4. Proofs

Proof of Lemma 1. We can proceed in the same way as for Lemma 7 in Kessler (1997). For details, see Sørensen and Uchida (2002).

Proof of Lemma 2. (i) In view of Theorem B in Genon-Catalot (1990) (cf. Theorem 1.3 in Azencott 1982), $\sup_{t \leq 1}|f(X_t, \theta) - f(X_t^0, \theta)| = o_p(1)$ for all θ. Thus, under P_{θ_0}, as $\varepsilon \to 0$ and $n \to \infty$,

$$\frac{1}{n}\sum_{k=1}^{n} f(X_{t_{k-1}}, \theta) \to \int_0^1 f(X_s^0, \theta)\mathrm{d}s.$$

Moreover, it follows from the assumption on f and (A2) that

$$\sup_{\varepsilon,n} \mathrm{E}_{\theta_0}\left[\sup_{\theta}\left|\frac{\partial}{\partial\theta}\left(\frac{1}{n}\sum_{k=1}^{n} f(X_{t_{k-1}}, \theta)\right)\right|\right] < \infty.$$

Therefore, the family of distributions of $(1/n)\sum_{k=1}^{n} f(X_{t_{k-1}}, \cdot)$ on the Banach space $C(\overline{\Theta})$ with the supremum norm is tight.

(ii) Let $\xi_k^i(\theta) = f(X_{t_{k-1}}, \theta)P_k^i(\alpha_0)$. From Lemmas 1 and 2 (i), under P_{θ_0}, as $\varepsilon \to 0$ and $n \to \infty$,

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

$$\sum_{k=1}^{n} \mathrm{E}[(\xi_k^i(\theta))^2|\mathcal{G}_{k-1}^n] = \sum_{k=1}^{n}\left\{R\left(\frac{\varepsilon^2}{n}, X_{t_{k-1}}\right) + R\left(\frac{1}{n^3}, X_{t_{k-1}}\right)\right\} \to 0.$$

It follows from Lemma 9 in Genon-Catalot and Jacod (1993) that $\sum_{k=1}^{n}\xi_k^i(\theta) \to 0$ in P_{θ_0}-probability as $\varepsilon \to 0$ and $n \to \infty$. Henceforth, let C be a generic positive constant independent of ε, n and, in some cases, other variables (see Yoshida 1992c; or Kessler 1997). Moreover, we may write C_m if it depends on an integer m. In order to prove the tightness of $\sum_{k=1}^{n}\xi_k^i(\cdot)$, it is enough to show the following inequalities (cf. Theorem 20 in Appendix I in Ibragimov and Has'minskii 1981; or Lemma 3.1 in Yoshida 1990):

$$\mathrm{E}_{\theta_0}\left[\left(\sum_{k=1}^{n}\xi_k^i(\theta)\right)^{2l}\right] \leqslant C, \tag{3}$$

$$\mathrm{E}_{\theta_0}\left[\left(\sum_{k=1}^{n}\xi_k^i(\theta_2)-\sum_{k=1}^{n}\xi_k^i(\theta_1)\right)^{2l}\right] \leqslant C|\theta_2-\theta_1|^{2l}, \tag{4}$$

for θ, θ_1, $\theta_2 \in \overline{\Theta}$, where $l > (p+q)/2$. We define $A^i_{k,1}(\theta)$, $A^i_{k,2}(\theta)$ and $A^i_{k,3}(\theta)$ by

$$\begin{aligned}
\xi_k^i(\theta) &= f(X_{t_{k-1}},\theta)\int_{t_{k-1}}^{t_k} b^i(X_s,\alpha_0)\mathrm{d}s + \varepsilon f(X_{t_{k-1}},\theta)\int_{t_{k-1}}^{t_k}\sum_{j=1}^{r}\sigma^{ij}(X_s,\beta_0)\mathrm{d}w_s^j \\
&\quad -\frac{1}{n}f(X_{t_{k-1}},\theta)b^i(X_{t_{k-1}},\alpha_0) \\
&=: A^i_{k,1}(\theta)+A^i_{k,2}(\theta)-A^i_{k,3}(\theta),
\end{aligned}$$

such that

$$\begin{aligned}
\mathrm{E}_{\theta_0}\left[\left|\sum_{k=1}^{n}A^i_{k,1}(\theta)\right|^{2l}\right] &\leqslant n^{2l-1}\sum_{k=1}^{n}\mathrm{E}_{\theta_0}\left[\left(\int_{t_{k-1}}^{t_k}|f(X_{t_{k-1}},\theta)b^i(X_s,\alpha_0)|\mathrm{d}s\right)^{2l}\right] \\
&\leqslant \sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}\mathrm{E}_{\theta_0}\left[|f(X_{t_{k-1}},\theta)|^{2l}\,\mathrm{E}_{\theta_0}[|b^i(X_s,\alpha_0)|^{2l}|\mathcal{G}^n_{k-1}]\right]\mathrm{d}s \\
&\leqslant n\cdot\frac{1}{n}\cdot C,
\end{aligned}$$

where the last estimate is based on Lemma 6 in Kessler (1997),

$$\begin{aligned}
\mathrm{E}_{\theta_0}\left[\left|\sum_{k=1}^{n}A^i_{k,2}(\theta)\right|^{2l}\right] &\leqslant C_{2l}\varepsilon^{2l}\mathrm{E}_{\theta_0}\left[\left(\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}f(X_{t_{k-1}},\theta)^2[\sigma\sigma^{\mathrm{T}}]^{ii}(X_s,\beta_0)\mathrm{d}s\right)^{l}\right] \\
&\leqslant C_{2l}\varepsilon^{2l}\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}\mathrm{E}_{\theta_0}\left[f(X_{t_{k-1}},\theta)^{2l}\mathrm{E}_{\theta_0}[([\sigma\sigma^{\mathrm{T}}]^{ii}(X_s,\beta_0))^l|\mathcal{G}^n_{k-1}]\right]\mathrm{d}s \\
&\leqslant C_{2l}\varepsilon^{2l}C,
\end{aligned}$$

where the first estimate is based on the Burkholder–Davis–Gundy inequality, and

$$\begin{aligned}
\mathrm{E}_{\theta_0}\left[\left|\sum_{k=1}^{n}A^i_{k,3}(\theta)\right|^{2l}\right] &\leqslant \frac{1}{n}\sum_{k=1}^{n}\mathrm{E}_{\theta_0}[|f(X_{t_{k-1}},\theta)b^i(X_{t_{k-1}},\alpha_0)|^{2l}] \\
&\leqslant C.
\end{aligned}$$

We thus deduce inequality (3). We obtain inequality (4) in the same way. This completes the proof. □

Proof of Lemma 3. (i) From Lemma 1,

$$\sum_{k=1}^{n} \mathrm{E}_{\theta_0}[\varepsilon^{-2} f(X_{t_{k-1}}, \theta) P_k^i P_k^j(\alpha_0) | \mathcal{G}_{k-1}^n] = \frac{1}{n} \sum_{k=1}^{n} f(X_{t_{k-1}}, \theta)[\sigma\sigma^{\mathrm{T}}]^{ij}(X_{t_{k-1}}, \beta_0) + \sum_{k=1}^{n} \left\{ R\left(\frac{1}{n^2}, X_{t_{k-1}}\right) + R\left(\frac{\varepsilon^{-2}}{n^3}, X_{t_{k-1}}\right) \right\},$$

$$\sum_{k=1}^{n} \mathrm{E}_{\theta_0}[\varepsilon^{-4} f(X_{t_{k-1}}, \theta)^2 (P_k^i P_k^j(\alpha_0))^2 | \mathcal{G}_{k-1}^n] = \sum_{k=1}^{n} \left\{ R\left(\frac{1}{n^2}, X_{t_{k-1}}\right) + R\left(\frac{\varepsilon^{-2}}{n^4}, X_{t_{k-1}}\right) + R\left(\frac{\varepsilon^{-4}}{n^5}, X_{t_{k-1}}\right) \right\}.$$

From Lemma 2(i) and (B1), under P_{θ_0}, as $\varepsilon \to 0$ and $n \to \infty$,

$$\sum_{k=1}^{n} \mathrm{E}_{\theta_0}[\varepsilon^{-2} f(X_{t_{k-1}}, \theta) P_k^i P_k^j(\alpha_0) | \mathcal{G}_{k-1}^n] \to \int_0^1 f(X_s^0, \theta)[\sigma\sigma^{\mathrm{T}}]^{ij}(X_s^0, \beta_0) \mathrm{d}s,$$

$$\sum_{k=1}^{n} \mathrm{E}_{\theta_0}[\varepsilon^{-4} f(X_{t_{k-1}}, \theta)^2 (P_k^i P_k^j(\alpha_0))^2 | \mathcal{G}_{k-1}^n] \to 0.$$

Thus, it follows from Lemma 9 in Genon-Catalot and Jacod (1993) that, under P_{θ_0}, as $\varepsilon \to 0$ and $n \to \infty$,

$$\varepsilon^{-2} \sum_{k=1}^{n} f(X_{t_{k-1}}, \theta) P_k^i P_k^j(\alpha_0) \to \int_0^1 f(X_s^0, \theta)[\sigma\sigma^{\mathrm{T}}]^{ij}(X_s^0, \beta_0) \mathrm{d}s.$$

For tightness of the family of distributions of $\varepsilon^{-2} \sum_{k=1}^{n} f(X_{t_{k-1}}, \cdot) P_k^i P_k^j(\alpha_0)$, we use the fact that

$$
\begin{aligned}
&\sup_{\varepsilon,n} \mathrm{E}_{\theta_0}\left[\sup_{\theta}\left|\varepsilon^{-2}\sum_{k=1}^{n}\frac{\partial}{\partial\theta}f(X_{t_{k-1}},\theta)P_k^iP_k^j(\alpha_0)\right|\right] \\
&\leqslant \sup_{\varepsilon,n} \mathrm{E}_{\theta_0}\left[\frac{\varepsilon^{-2}}{2}\sup_{\theta}\sum_{k=1}^{n}\left|\frac{\partial}{\partial\theta}f(X_{t_{k-1}},\theta)\right|\mathrm{E}_{\theta_0}[(P_k^i(\alpha_0))^2+(P_k^j(\alpha_0))^2|\mathcal{G}_{k-1}^n]\right] \\
&\leqslant \frac{1}{2}\sup_{\varepsilon,n} \mathrm{E}_{\theta_0}\left[\sum_{k=1}^{n}\sup_{\theta}\left|\frac{\partial}{\partial\theta}f(X_{t_{k-1}},\theta)\right|\left\{\frac{1}{n}([\sigma\sigma^{\mathrm{T}}]^{ii}(X_{t_{k-1}},\beta_0)+[\sigma\sigma^{\mathrm{T}}]^{jj}(X_{t_{k-1}},\beta_0))\right.\right. \\
&\quad \left.\left.+R\left(\frac{1}{n^2},X_{t_{k-1}}\right)+R\left(\frac{\varepsilon^{-2}}{n^3},X_{t_{k-1}}\right)\right\}\right] \\
&< \infty,
\end{aligned}
$$

where the last estimate is based on $\lim_{\varepsilon\to 0, n\to\infty}(\varepsilon n)^{-1}=0$.

(ii) Noting that

$$
\begin{aligned}
P_k^iP_k^j(\alpha) &= P_k^iP_k^j(\alpha_0)+\frac{1}{n}P_k^i(\alpha_0)B_{k-1}^j(\alpha_0,\alpha)+\frac{1}{n}P_k^j(\alpha_0)B_{k-1}^i(\alpha_0,\alpha) \\
&\quad +\frac{1}{n^2}B_{k-1}^iB_{k-1}^j(\alpha_0,\alpha),
\end{aligned}
$$

it follows from Lemmas 2 and 3(i) and (B2) that under P_{θ_0}, as $\varepsilon\to 0$ and $n\to\infty$,

$$
\begin{aligned}
&\varepsilon^{-2}\sum_{k=1}^{n}f(X_{t_{k-1}},\theta)P_k^iP_k^j(\alpha) \\
&= \varepsilon^{-2}\sum_{k=1}^{n}f(X_{t_{k-1}},\theta)P_k^iP_k^j(\alpha_0)+\frac{1}{n^2}\varepsilon^{-2}\sum_{k=1}^{n}f(X_{t_{k-1}},\theta)B_{k-1}^iB_{k-1}^j(\alpha_0,\alpha) \\
&\quad +\frac{1}{n}\varepsilon^{-2}\sum_{k=1}^{n}f(X_{t_{k-1}},\theta)\left\{P_k^i(\alpha_0)B_{k-1}^j(\alpha_0,\alpha)+P_k^j(\alpha_0)B_{k-1}^i(\alpha_0,\alpha)\right\} \\
&\to \int_0^1 f(X_s^0,\theta)[\sigma\sigma^{\mathrm{T}}]^{ij}(X_s^0,\beta_0)\mathrm{d}s+M^2\int_0^1 f(X_s^0,\theta)B^iB^j(X_s^0,\alpha_0,\alpha)\mathrm{d}s
\end{aligned}
$$

uniformly in $\theta\in\overline{\Theta}$, where $M=\lim_{\varepsilon\to 0,n\to\infty}(\varepsilon\sqrt{n})^{-1}$. This completes the proof. □

Proof of Lemma 4. (i) A simple computation yields

$$\varepsilon^2\{U_{\varepsilon,n}(\alpha,\beta)-U_{\varepsilon,n}(\alpha_0,\beta)\} = n\sum_{k=1}^{n}(P_k(\alpha)-P_k(\alpha_0))^{\mathrm{T}}\Xi_{k-1}^{-1}(\beta)(P_k(\alpha)+P_k(\alpha_0))$$

$$= \sum_{k=1}^{n}(b(X_{t_{k-1}},\alpha)-b(X_{t_{k-1}},\alpha_0))^{\mathrm{T}}\Xi_{k-1}^{-1}(\beta)$$

$$\times\left(2\left\{X_{t_k}-X_{t_{k-1}}-\frac{1}{n}b(X_{t_{k-1}},\alpha_0)\right\}.\right.$$

$$\left.+\frac{1}{n}\{b(X_{t_{k-1}},\alpha_0)-b(X_{t_{k-1}},\alpha)\}\right).$$

From Lemma 2, under P_{θ_0}, as $\varepsilon\to 0$ and $n\to\infty$,

$$\varepsilon^2\{U_{\varepsilon,n}(\alpha,\beta)-U_{\varepsilon,n}(\alpha_0,\beta)\}\to U_1(\alpha,\alpha_0,\beta)$$

uniformly in $\theta\in\overline{\Theta}$.

(ii) It follows from Lemmas 2(i) and 3(ii) that under P_{θ_0}, as $\varepsilon\to 0$ and $n\to\infty$,

$$\frac{1}{n}U_{\varepsilon,n}(\alpha,\beta)\to U_2(\alpha,\beta,\beta_0)$$

uniformly in $\theta\in\overline{\Theta}$. This completes the proof. □

Proof of Lemma 5. We first consider the uniform convergence of $C_{\varepsilon,n}(\theta)$. From Lemma 2, under P_{θ_0}, as $\varepsilon\to 0$ and $n\to\infty$,

$$\varepsilon^2\frac{\partial^2}{\partial\alpha_i\partial\alpha_j}U_{\varepsilon,n}(\theta)\to -2\int_0^1\left(\frac{\partial^2}{\partial\alpha_i\partial\alpha_j}b(X_s^0,\alpha)\right)^{\mathrm{T}}[\sigma\sigma^{\mathrm{T}}]^{-1}(X_s^0,\beta)B(X_s^0,\alpha_0,\alpha)\mathrm{d}s \tag{5}$$

$$+2\int_0^1\left(\frac{\partial}{\partial\alpha_i}b(X_s^0,\alpha)\right)^{\mathrm{T}}[\sigma\sigma^{\mathrm{T}}]^{-1}(X_s^0,\beta)\left(\frac{\partial}{\partial\alpha_j}b(X_s^0,\alpha)\right)\mathrm{d}s,$$

$$\varepsilon\frac{1}{\sqrt{n}}\frac{\partial^2}{\partial\alpha_i\partial\beta_j}U_{\varepsilon,n}(\theta)\to -2M\int_0^1\left(\frac{\partial}{\partial\alpha_i}b(X_s^0,\alpha)\right)^{\mathrm{T}}\left(\frac{\partial}{\partial\beta_j}[\sigma\sigma^{\mathrm{T}}]^{-1}(X_s^0,\beta)\right)B(X_s^0,\alpha_0,\alpha)\mathrm{d}s, \tag{6}$$

$$\frac{1}{n}\frac{\partial^2}{\partial\beta_i\partial\beta_j}U_{\varepsilon,n}(\theta)\to\int_0^1\frac{\partial^2}{\partial\beta_i\partial\beta_j}\log\det[\sigma\sigma^{\mathrm{T}}](\mathrm{X}_{\mathrm{s}}^0,\beta)\mathrm{ds} \tag{7}$$

$$+\int_0^1\mathrm{tr}\left[[\sigma\sigma^{\mathrm{T}}](X_s^0,\beta_0)\left(\frac{\partial^2}{\partial\beta_i\partial\beta_j}[\sigma\sigma^{\mathrm{T}}]^{-1}(X_s^0,\beta)\right)\right]\mathrm{d}s$$

$$+M^2\int_0^1\mathrm{tr}\left[BB^{\mathrm{T}}(X_s^0,\alpha_0,\alpha)\left(\frac{\partial^2}{\partial\beta_i\partial\beta_j}[\sigma\sigma^{\mathrm{T}}]^{-1}(X_s^0,\beta)\right)\right]\mathrm{d}s$$

uniformly in $\theta \in \overline{\Theta}$. Now (i) follows from (5), (6) and (7). Next, by the assumptions (A3) and (A4), the limits of (5), (6) and (7) are continuous with respect to θ, which completes the proof of (ii). □

Proof of Lemma 6. We set

$$
\begin{aligned}
-\varepsilon \frac{\partial}{\partial \alpha_i} U_{\varepsilon,n}(\theta_0) &= \sum_{k=1}^{n} 2\varepsilon^{-1} \sum_{l_1=1}^{d} \left[\left(\frac{\partial}{\partial \alpha_i} b(X_{t_{k-1}}, \alpha_0) \right)^{\mathrm{T}} \Xi_{k-1}^{-1}(\beta_0) \right]^{l_1} P_k^{l_1}(\alpha_0) \\
&=: \sum_{k=1}^{n} \xi_k^i(\theta_0), \\
-\frac{1}{\sqrt{n}} \frac{\partial}{\partial \beta_j} U_{\varepsilon,n}(\theta_0) &= -\sum_{k=1}^{n} \frac{1}{\sqrt{n}} \frac{\partial}{\partial \beta_j} \log \det \Xi_{\mathrm{k}-1}(\beta_0) \\
&\quad - \sum_{k=1}^{n} \varepsilon^{-2} \sqrt{n} \sum_{l_1,l_2=1}^{d} \left(\frac{\partial}{\partial \beta_j} \Xi_{k-1}^{-1}(\beta_0) \right)^{l_1 l_2} P_k^{l_1} P_k^{l_2}(\alpha_0) \\
&=: \sum_{k=1}^{n} \eta_k^j(\theta_0).
\end{aligned}
$$

In view of Theorems 3.2 and 3.4 in Hall and Heyde (1980), it is sufficient to show that under P_{θ_0}, as $\varepsilon \to 0$ and $n \to \infty$,

$$\sum_{k=1}^{n} \mathrm{E}_{\theta_0}[\xi_k^i(\theta_0) | \mathcal{G}_{k-1}^n] \to 0, \tag{8}$$

$$\sum_{k=1}^{n} \mathrm{E}_{\theta_0}[\eta_k^j(\theta_0) | \mathcal{G}_{k-1}^n] \to 0, \tag{9}$$

$$\sum_{k=1}^{n} \mathrm{E}_{\theta_0}[\xi_k^{i_1} \xi_k^{i_2}(\theta_0) | \mathcal{G}_{k-1}^n] \to 4 I_b^{i_1 i_2}(\theta_0), \tag{10}$$

$$\sum_{k=1}^{n} \mathrm{E}_{\theta_0}[\eta_k^{j_1} \eta_k^{j_2}(\theta_0) | \mathcal{G}_{k-1}^n] \to 4 I_\sigma^{j_1 j_2}(\theta_0), \tag{11}$$

$$\sum_{k=1}^{n} \mathrm{E}_{\theta_0}[\xi_k^i \eta_k^j(\theta_0) | \mathcal{G}_{k-1}^n] \to 0, \tag{12}$$

$$\sum_{k=1}^{n} \mathrm{E}_{\theta_0}[(\xi_k^i(\theta_0))^4 | \mathcal{G}_{k-1}^n] \to 0, \tag{13}$$

$$\sum_{k=1}^{n} \mathrm{E}_{\theta_0}[(\eta_k^j(\theta_0))^4 | \mathcal{G}_{k-1}^n] \to 0. \tag{14}$$

Using Lemma 1, we obtain

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

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

$$\sum_{k=1}^{n} \mathrm{E}_{\theta_0}[\xi_k^{i_1}\xi_k^{i_2}(\theta_0)|\mathcal{G}_{k-1}^n] = 4\frac{1}{n}\sum_{k=1}^{n} \frac{\partial}{\partial \alpha_{i_1}} b(X_{t_{k-1}}, \alpha_0)^{\mathrm{T}} \Xi_{k-1}^{-1}(\beta_0) \frac{\partial}{\partial \alpha_{i_2}} b(X_{t_{k-1}}, \alpha_0) + \sum_{k=1}^{n} \left\{ R\left(\frac{1}{n^2}, X_{t_{k-1}}\right) + R\left(\frac{\varepsilon^{-2}}{n^3}, X_{t_{k-1}}\right) \right\} \to 4I_b^{i_1 i_2}(\theta_0),$$

$$\sum_{k=1}^{n} \mathrm{E}_{\theta_0}[\eta_k^{j_1}\eta_k^{j_2}(\theta_0)|\mathcal{G}_{k-1}^n] = \frac{1}{n}\sum_{k=1}^{n} \left\{ 2\mathrm{tr}\left[\left(\frac{\partial}{\partial \beta_{j_1}} \Xi_{k-1}\right) \Xi_{k-1}^{-1} \left(\frac{\partial}{\partial \beta_{j_2}} \Xi_{k-1}\right) \Xi_{k-1}^{-1}(\beta_0) \right] \right\} + \sum_{k=1}^{n} \left\{ R\left(\frac{1}{n^2}, X_{t_{k-1}}\right) + R\left(\frac{\varepsilon^{-2}}{n^3}, X_{t_{k-1}}\right) + R\left(\frac{\varepsilon^{-4}}{n^5}, X_{t_{k-1}}\right) \right\} \to 4I_\sigma^{j_1 j_2}(\theta_0),$$

$$\sum_{k=1}^{n} \mathrm{E}_{\theta_0}[\xi_k^i\eta_k^j(\theta_0)|\mathcal{G}_{k-1}^n] = \sum_{k=1}^{n} \left\{ R\left(\frac{\varepsilon}{n\sqrt{n}}, X_{t_{k-1}}\right) + R\left(\frac{\varepsilon^{-1}}{n^2\sqrt{n}}, X_{t_{k-1}}\right) + R\left(\frac{\varepsilon^{-3}}{n^3\sqrt{n}}, X_{t_{k-1}}\right) \right\} \to 0$$

$$\sum_{k=1}^{n} \mathrm{E}_{\theta_0}[(\xi_k^i(\theta_0))^4|\mathcal{G}_{k-1}^n] = \sum_{k=1}^{n} \left\{ R\left(\frac{1}{n^2}, X_{t_{k-1}}\right) + R\left(\frac{\varepsilon^{-2}}{n^4}, X_{t_{k-1}}\right) + R\left(\frac{\varepsilon^{-4}}{n^5}, X_{t_{k-1}}\right) \right\} \to 0,$$

in P_{θ_0}-probability, as $\varepsilon \to 0$ and $n \to \infty$, thus proving (8)–(13).

To prove (14), we first obtain several estimates as follows:

$$(\eta_k^j(\theta_0))^4 \leqslant 2^3 \left[\frac{1}{n^2} \left(\frac{\partial}{\partial \beta_j} \log \det \Xi_{k-1}(\beta_0) \right)^4 \right.$$

$$\left. + \varepsilon^{-8} n^2 (2d)^3 \sum_{l_1, l_2 = 1}^{d} \left[\left(\frac{\partial}{\partial \beta_j} \Xi_{k-1}^{-1}(\beta_0) \right)^{l_1 l_2} \right]^4 (P_k^{l_1} P_k^{l_2}(\alpha_0))^4 \right],$$

$$\mathrm{E}_{\theta_0}[(P_k^{l_1} P_k^{l_2})^4(\alpha_0) | \mathcal{G}_{k-1}^n] \leqslant 3^3 \{ \mathrm{E}_{\theta_0}[((X_{t_k} - X_{t_{k-1}})^{l_1} (X_{t_k} - X_{t_{k-1}})^{l_2})^4 | \mathcal{G}_{k-1}^n]$$

$$+ \frac{1}{n^4} (b^{l_1}(X_{t_{k-1}}, \alpha_0))^4 \, \mathrm{E}_{\theta_0}[((X_{t_k} - X_{t_{k-1}})^{l_2})^4 | \mathcal{G}_{k-1}^n]$$

$$+ \frac{1}{n^4} (b^{l_2}(X_{t_{k-1}}, \alpha_0))^4 \, \mathrm{E}_{\theta_0}[(P_k^{l_1})^4(\alpha_0) | \mathcal{G}_{k-1}^n] \}.$$

By using a version of Lemma 1, we have

$$\mathrm{E}_{\theta_0}[((X_{t_k} - X_{t_{k-1}})^{l_1})^8 | \mathcal{G}_{k-1}^n] = R\left(\frac{\varepsilon^8}{n^4}, X_{t_{k-1}} \right) + R\left(\frac{\varepsilon^6}{n^5}, X_{t_{k-1}} \right) + R\left(\frac{\varepsilon^4}{n^6}, X_{t_{k-1}} \right) \quad (15)$$

$$+ R\left(\frac{\varepsilon^2}{n^7}, X_{t_{k-1}} \right) + R\left(\frac{1}{n^8}, X_{t_{k-1}} \right).$$

It then follows from (15) and Lemma 1 that

$$\sum_{k=1}^{n} \mathrm{E}_{\theta_0}[(\eta_k^j(\theta_0))^4 | \mathcal{G}_{k-1}^n] \leqslant \sum_{k=1}^{n} \left\{ R\left(\frac{1}{n^2}, X_{t_{k-1}} \right) + R\left(\frac{\varepsilon^{-2}}{n^3}, X_{t_{k-1}} \right) + R\left(\frac{\varepsilon^{-4}}{n^4}, X_{t_{k-1}} \right) \right.$$

$$\left. + R\left(\frac{\varepsilon^{-6}}{n^5}, X_{t_{k-1}} \right) + R\left(\frac{\varepsilon^{-8}}{n^6}, X_{t_{k-1}} \right) \right\}$$

$$\to 0,$$

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

Proof of Theorem 1. First of all, in order to prove consistency of $\hat{\theta}_{\varepsilon,n}$, we note that in view of the compactness of $\overline{\Theta}$, there exists a subsequence (ε_k, n_k) such that $\hat{\theta}_{\varepsilon_k,n_k}$ goes to a limit $\theta_\infty = (\alpha_\infty, \beta_\infty) \in \overline{\Theta}$.

Next, it follows from Lemma 4(i) and the continuity of $U_1(\alpha, \alpha_0, \beta)$ with respect to (α, β) that under P_{θ_0}, as $\varepsilon_k \to 0$ and $n_k \to \infty$,

$$\varepsilon_k^2 U_{\varepsilon_k,n_k}(\hat{\alpha}_{\varepsilon_k,n_k}, \hat{\beta}_{\varepsilon_k,n_k}) - \varepsilon_k^2 U_{\varepsilon_k,n_k}(\alpha_0, \hat{\beta}_{\varepsilon_k,n_k}) \to U_1(\alpha_\infty, \alpha_0, \beta_\infty). \quad (16)$$

From the definition of the minimum-contrast estimator $\hat{\theta}_{\varepsilon,n}$ and $\alpha_0 \in \overline{\Theta}_\alpha$,

$$\varepsilon_k^2 U_{\varepsilon_k,n_k}(\hat{\alpha}_{\varepsilon_k,n_k}, \hat{\beta}_{\varepsilon_k,n_k}) - \varepsilon_k^2 U_{\varepsilon_k,n_k}(\alpha_0, \hat{\beta}_{\varepsilon_k,n_k}) \leqslant 0. \quad (17)$$

By (16), (17), (A4) and (A5), we obtain $\alpha_\infty = \alpha_0$. We have thus deduced that any convergent subsequence of $\hat{\alpha}_{\varepsilon,n}$ goes to α_0. Thus, the consistency of $\hat{\alpha}_{\varepsilon,n}$ is proved.

For the consistency of $\hat{\beta}_{\varepsilon,n}$, using Lemma 4(ii) and the continuity of $U_2(\alpha, \beta, \beta_0)$ with respect to (α, β), under P_{θ_0}, we have

$$\frac{1}{n_k} U_{\varepsilon_k,n_k}(\hat{\alpha}_{\varepsilon_k,n_k}, \hat{\beta}_{\varepsilon_k,n_k}) \to U_2(\alpha_0, \beta_\infty, \beta_0), \tag{18}$$

as $\varepsilon_k \to 0$ and $n_k \to \infty$, where we note that $\alpha_{\varepsilon_k,n_k}$ tends to α_0 by the previous paragraph. Moreover, in view of the definition of $\hat{\theta}_{\varepsilon,n}$ and $\beta_0 \in \overline{\Theta}_\beta$,

$$\frac{1}{n_k} U_{\varepsilon_k,n_k}(\hat{\alpha}_{\varepsilon_k,n_k}, \hat{\beta}_{\varepsilon_k,n_k}) \leqslant \frac{1}{n_k} U_{\varepsilon_k,n_k}(\hat{\alpha}_{\varepsilon_k,n_k}, \beta_0). \tag{19}$$

It follows from (18) and (19) that $U_2(\alpha_0, \beta_\infty, \beta_0) \leqslant U_2(\alpha_0, \beta_0, \beta_0)$. Moreover, by a version of Lemma 17 in Genon-Catalot and Jacod (1993),

$$\log\det[\sigma\sigma^{\mathrm{T}}](X_t^0, \beta_0) + d \leqslant \log\det[\sigma\sigma^{\mathrm{T}}](X_t^0, \beta_\infty) + \mathrm{tr}\big[[\sigma\sigma^{\mathrm{T}}](X_t^0, \beta_0)[\sigma\sigma^{\mathrm{T}}]^{-1}(X_t^0, \beta_\infty)\big]$$

with equality if and only if $[\sigma\sigma^{\mathrm{T}}](X_t^0, \beta_\infty) = [\sigma\sigma^{\mathrm{T}}](X_t^0, \beta_0)$. Hence $U_2(\alpha_0, \beta_\infty, \beta_0) \geqslant U_2(\alpha_0, \beta_0, \beta_0)$. Thus (A5), together with the above inequalities for U_2, implies $\beta_\infty = \beta_0$. Therefore, noting that any convergent subsequence of $\hat{\beta}_{\varepsilon,n}$ tends to β_0, the consistency of $\hat{\beta}_{\varepsilon,n}$ is proved.

Finally, we prove the asymptotic normality of $\hat{\theta}_{\varepsilon,n}$. Let $B(\theta_0; \rho) = \{\theta : |\theta - \theta_0| \leqslant \rho\}$. It follows from $\theta_0 \in \Theta$ that for sufficiently small $\rho > 0$, $B(\theta_0; \rho) \subset \Theta$. By Taylor's formula, if $\hat{\theta}_{\varepsilon,n} \in B(\theta_0; \rho)$,

$$D_{\varepsilon,n} S_{\varepsilon,n} = \Lambda_{\varepsilon,n},$$

where

$$D_{\varepsilon,n} = \int_0^1 C_{\varepsilon,n}\Big(\theta_0 + u(\hat{\theta}_{\varepsilon,n} - \theta_0)\Big)\mathrm{d}u, \qquad S_{\varepsilon,n} = (\varepsilon^{-1}(\hat{\alpha}_{\varepsilon,n} - \alpha_0)^{\mathrm{T}}, \sqrt{n}(\hat{\beta}_{\varepsilon,n} - \beta_0)^{\mathrm{T}})^{\mathrm{T}}.$$

It follows from the consistency of $\hat{\theta}_{\varepsilon,n}$ that for sufficiently small $\rho > 0$,

$$P_{\theta_0}[\hat{\theta}_{\varepsilon,n} \in \Theta] \geqslant P_{\theta_0}[|\hat{\theta}_{\varepsilon,n} - \theta_0| < \rho] \to 1,$$

as $\varepsilon \to 0$ and $n \to \infty$. Moreover, by the consistency of $\hat{\theta}_{\varepsilon,n}$, there exists a sequence $\{B(\theta_0; \eta_{\varepsilon,n})\}$ such that $\eta_{\varepsilon,n} \to 0$ and $P_{\theta_0}[\hat{\theta}_{\varepsilon,n} \in B(\theta_0; \eta_{\varepsilon,n})] \to 1$, as $\varepsilon \to 0$ and $n \to \infty$. Since we obtain

$$P_{\theta_0}[\hat{\theta}_{\varepsilon,n} \in \Theta^c \cup B(\theta_0; \eta_{\varepsilon,n})^c] \leqslant P_{\theta_0}[\hat{\theta}_{\varepsilon,n} \in \Theta^c] + P_{\theta_0}[\hat{\theta}_{\varepsilon,n} \in B(\theta_0; \eta_{\varepsilon,n})^c] \to 0,$$

as $\varepsilon \to 0$ and $n \to \infty$, we have $1_{\{\hat{\theta}_{\varepsilon,n} \in \Theta^c \cup B(\theta_0; \eta_{\varepsilon,n})^c\}} \to 0$ in P_{θ_0}-probability as $\varepsilon \to 0$ and $n \to \infty$. By Lemma 5(ii), letting $R_n = D_{\varepsilon,n} - C_{\varepsilon,n}(\theta_0)$,

$$|R_{\varepsilon,n}| \cdot 1_{\{\hat{\theta}_{\varepsilon,n} \in \Theta \cap B(\theta_0; \eta_{\varepsilon,n})\}} \leqslant \sup_{\theta \in B(\theta_0; \eta_{\varepsilon,n})} |C_{\varepsilon,n}(\theta) - C_{\varepsilon,n}(\theta_0)| \to 0$$

in P_{θ_0}-probability, as $\varepsilon \to 0$ and $n \to \infty$. Thus, under P_{θ_0}, as $\varepsilon \to 0$ and $n \to \infty$, we have $R_{\varepsilon,n} \to 0$. By using Lemma 5(i), $D_{\varepsilon,n} \to 2I(\theta_0)$ in P_{θ_0}-probability, as $\varepsilon \to 0$ and $n \to \infty$.

Let $\Gamma(\theta)$ be the limit of $C_{\varepsilon,n}(\theta)$ as $\varepsilon \to 0$ and $n \to \infty$. For details of $\Gamma(\theta)$, see (5), (6) and (7) in the proof of Lemma 5. Note that $\Gamma(\theta)$ is continuous with respect to θ. Since $I(\theta_0)$ is positive definite, there exists a positive constant C such that $\inf_{|x|=1}|I(\theta_0)x| > 2C$. For such $C > 0$, there exist $N_1(C) > 0$ and $N_2(C) > 0$ such that for any $\varepsilon < N_1(C)$ ($\varepsilon > 0$) and $n > N_2(C)$, and for any $\delta \in [0, 1]$, $B(\theta_0; \eta_{\varepsilon,n}) \subset \Theta$ and $|\Gamma(\theta_0 + \delta\eta_{\varepsilon,n}) - \Gamma(\theta_0)| < C/2$, where $\eta_{\varepsilon,n} \to 0$ as $\varepsilon \to 0$ and $n \to \infty$. For such $C > 0$, let $\mathcal{C}_{\varepsilon,n}$ be the set defined by

$$\mathcal{C}_{\varepsilon,n} = \left\{ \sup_{\theta\in\overline{\Theta}} |C_{\varepsilon,n}(\theta) - \Gamma(\theta)| < \frac{C}{2},\ \hat{\theta}_{\varepsilon,n} \in B(\theta_0; \eta_{\varepsilon,n}) \right\}.$$

For any $\varepsilon < N_1(C)$ ($\varepsilon > 0$) and $n > N_2(C)$, and for any $|u| < 1$, one has, on $\mathcal{C}_{\varepsilon,n}$,

$$\begin{aligned}
\sup_{|x|=1}|(-D_n + \Gamma(\theta_0))x| &\leqslant \sup_{|x|=1}\left|\left(-D_n + \int_0^1 \Gamma(\theta_0 + u(\hat{\theta}_{\varepsilon,n} - \theta_0))\mathrm{d}u\right)x\right| \\
&\quad + \sup_{|x|=1}\left|\left(\Gamma(\theta_0) - \int_0^1 \Gamma(\theta_0 + u(\hat{\theta}_{\varepsilon,n} - \theta_0))\mathrm{d}u\right)x\right| \\
&\leqslant \sup_{|\theta-\theta_0|<\eta_{\varepsilon,n}} |C_{\varepsilon,n}(\theta) - \Gamma(\theta)| + \frac{C}{2} \\
&< C.
\end{aligned}$$

Hence, for any $\varepsilon < N_1(C)$ ($\varepsilon > 0$) and $n > N_2(C)$, we obtain, on $\mathcal{C}_{\varepsilon,n}$,

$$\begin{aligned}
\inf_{|x|=1} |D_{\varepsilon,n}x| &\geqslant \inf_{|x|=1} |\Gamma(\theta_0)x| - \sup_{|x|=1} |(-D_{\varepsilon,n} + \Gamma(\theta_0))x| \\
&> C.
\end{aligned}$$

Let $\mathcal{D}_{\varepsilon,n} = \{D_{\varepsilon,n} \text{ is invertible}\}$. It then follows that for any $\varepsilon < N_1(C)$ ($\varepsilon > 0$) and $n > N_2(C)$, $P_{\theta_0}[\mathcal{D}_{\varepsilon,n}] \geqslant P_{\theta_0}[\mathcal{C}_{\varepsilon,n}]$. Since it follows from (5), (6) and (7) that under (B2), $P_{\theta_0}[\mathcal{C}_{\varepsilon,n}] \to 1$ as $\varepsilon \to 0$ and $n \to \infty$, we have $P_{\theta_0}[\mathcal{D}_{\varepsilon,n}] \to 1$ as $\varepsilon \to 0$ and $n \to \infty$.

Let $\mathcal{E}_{\varepsilon,n} = \{\hat{\theta}_{\varepsilon,n} \in \Theta\} \cap \mathcal{D}_{\varepsilon,n}$, and $E_{\varepsilon,n} = D_{\varepsilon,n}$ on $\mathcal{E}_{\varepsilon,n}$ and $E_{\varepsilon,n} = J_{p+q}$ on $\mathcal{E}^c_{\varepsilon,n}$, where J_{p+q} is the $(p+q) \times (p+q)$ identity matrix. Note that $P_{\theta_0}[\mathcal{E}_{\varepsilon,n}] \to 1$ as $\varepsilon \to 0$ and $n \to \infty$. Since $|E_{\varepsilon,n} - 2I(\theta_0)|1_{\mathcal{E}_{\varepsilon,n}} \leqslant |D_{\varepsilon,n} - 2I(\theta_0)|$ and $1_{\mathcal{E}_{\varepsilon,n}} \to 1$ in P_{θ_0}-probability as $\varepsilon \to 0$ and $n \to \infty$, we have $E_{\varepsilon,n} \to 2I(\theta_0)$ in P_{θ_0}-probability as $\varepsilon \to 0$ and $n \to \infty$. Noting that $S_{\varepsilon,n}1_{\mathcal{E}_{\varepsilon,n}} = E^{-1}_{\varepsilon,n}D_{\varepsilon,n}S_{\varepsilon,n}1_{\mathcal{E}_{\varepsilon,n}} = E^{-1}_{\varepsilon,n}\Lambda_{\varepsilon,n}1_{\mathcal{E}_{\varepsilon,n}}$ and by Lemma 6, $S_{\varepsilon,n}1_{\mathcal{E}_n} \to N(0, I(\theta_0)^{-1})$ in distribution as $\varepsilon \to 0$ and $n \to \infty$. Thus, again using the fact that under P_{θ_0}, as $\varepsilon \to 0$ and $n \to \infty$, $1_{\mathcal{E}_{\varepsilon,n}} \to 1$, we complete the proof. □

Proof of Corollary 1. When $\sigma(x, \beta) = \sigma(x)$, it is easy to show that Lemmas 4(i), 5 and 6 hold under the assumptions (A1), (A2), (A3′), (A4′) and (B1). In the same way as for the proof of Theorem 1, we deduce the result. □

Acknowledgements

The authors wish to thank the referees for valuable comments. The research was carried out during a year-long visit by Masayuki Uchida at the University of Copenhagen in 2001–2002 under the programme for Overseas Researchers of the Japan Ministry of Education, Culture, Sports, Science and Technology. Masayuki Uchida is grateful to the Department of Statistics and Operations Research, University of Copenhagen for their kind hospitality. The research of Michael Sørensen was supported by MaPhySto – Centre for Mathematical Physics and Stochastics, funded by a grant from the Danish National Research Foundation, and by the European Commission through the Human Potential Programme.

References

Azencott, R. (1982) Formule de Taylor stochastique et développement asymptotique d'intégrales de Feynmann. In J. Azéma and M. Yor (eds), *Séminaire de Probabilités XVI. Supplément: Géométrie Différentielle Stochastique*, Lecture Notes in Math. 921, pp. 237–285. Berlin: Springer-Verlag.

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

Bibby, B.M. and Sørensen, M. (2001) Simplified estimating functions for diffusion models with a high-dimensional parameter. *Scand. J. Statist*., **28**, 99–112.

Dacunha-Castelle, D. and Florens-Zmirou, D. (1986) Estimation of the coefficient of a diffusion from discrete observations. *Stochastics*, **19**, 263–284.

Dermoune, A. and Kutoyants, Yu.A. (1995) Expansion of distribution function of maximum likelihood estimate for misspecified diffusion type observations. *Stochastics Stochastic 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. H. Poincaré Probab. Statist*., **29**, 119–151.

Gobet, E. (2001) Local asymptotic mixed normality property for elliptic diffusion: A Malliavin calculus approach. *Bernoulli*, **7**, 899–912.

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

Hall, P. and Heyde, C. (1980) *Martingale Limit Theory and Its Applications*. New York: Academic Press.

Hansen, L.P. and Scheinkman, J.A. (1995) Back to the future: generating moment implications for continuous-time Markov processes. *Econometrica*, **63**, 767–804.

Ibragimov, I.A. and Has'minskii, R.Z. (1981) *Statistical Estimation*. New York: Springer-Verlag.

Ikeda, N. and Watanabe, S. (1989) *Stochastic Differential Equations and Diffusion Processes* (2nd edition). Tokyo: North-Holland/Kodansha.

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

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.

Kessler, M. and Sørensen, M. (1999) Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli*, **5**, 299–314.

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

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*. Berlin: Heldermann.

Kutoyants, Yu.A. (1994) *Identification of Dynamical Systems with Small Noise*. Dordrecht: Kluwer.

Laredo, C. (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, H. (2001) Discretely observed diffusions: Approximation of the continuous-time score function. *Scand. J. Statist.*, **28**, 113–121.

Sørensen, M. (1997) Estimating functions for discretely observed diffusions: A review. In I.V. Basawa, V.P. Godambe and R.L. Taylor (eds), *Selected Proceedings of the Symposium on Estimating Functions*, IMS Lecture Notes Monogr. Ser. 32, pp. 305–325. Hayward, CA: Institute of Mathematical Statistics.

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. (2002) Small diffusion asymptotics for discretely sampled stochastic differential equations. Preprint no. 2002-3, Department of Statistics and Operations Research, University of Copenhagen.

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

Uchida, M. and Yoshida, N. (1999) Information criteria for small diffusions via the theory of Malliavin–Watanabe. *Statist. Inference Stochastic Process*. To appear.

Uchida, M. and Yoshida, N. (2001) Asymptotic expansion for small diffusions applied to option pricing. *Statist. Inference Stochastic Process*. To appear.

Uchida, M. (2002) Local asymptotic normality for discretely observed diffusion processes with small noise. In preparation.

Yoshida, N. (1990) Asymptotic behavior of M-estimator and related random field for diffusion process. *Ann. Inst. Statist. Math.*, **42**, 221–251.

Yoshida, N. (1992a) Asymptotic expansion of maximum likelihood estimators for small diffusions via the theory of Malliavin–Watanabe. *Probab. Theory Related 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 Related 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 March 2002 and revised April 2003