ESAIM: PS 15 (2011) 270–290
DOI: 10.1051/ps/2010001

ESAIM: Probability and Statistics
www.esaim-ps.org

ESTIMATION FOR MISSPECIFIED ERGODIC DIFFUSION PROCESSES FROM DISCRETE OBSERVATIONS

Masayuki Uchida[1, 2] and Nakahiro Yoshida[3]

Abstract. The joint estimation of both drift and diffusion coefficient parameters is treated under the situation where the data are discretely observed from an ergodic diffusion process and where the statistical model may or may not include the true diffusion process. We consider the minimum contrast estimator, which is equivalent to the maximum likelihood type estimator, obtained from the contrast function based on a locally Gaussian approximation of the transition density. The asymptotic normality of the minimum contrast estimator is proved. In particular, the rate of convergence for the minimum contrast estimator of diffusion coefficient parameter in a misspecified model is different from the one in the correctly specified parametric model.

Mathematics Subject Classification. 62F12, 62M05, 60J60.

Received November 17, 2008. Revised September 30, 2009.

1. Introduction

Let X_t be a d-dimensional ergodic diffusion process defined by the stochastic differential equation

$$\mathrm{d}X_t = B(X_t)\mathrm{d}t + S(X_t)\mathrm{d}w_t, \quad t \in [0,T], \quad X_0 = \eta, \tag{1.1}$$

where B is an $\mathbf{R}^d$-valued function defined on $\mathbf{R}^d$, S is an $\mathbf{R}^d \otimes \mathbf{R}^r$-valued function defined on $\mathbf{R}^d$ and w is an r-dimensional standard Wiener process independent of X_0. We consider a family of parametric models defined by the stochastic differential equations

$$\mathrm{d}X_t = b(X_t,\alpha)\mathrm{d}t + \sigma(X_t,\beta)\mathrm{d}w_t, \quad t \in [0,T], \quad X_0 = \eta, \tag{1.2}$$

where $\theta = (\alpha,\beta) \in \Theta_\alpha \times \Theta_\beta = \Theta$ with Θ_α and Θ_β being compact convex subsets of $\mathbf{R}^p$ and $\mathbf{R}^q$, respectively. Furthermore, b is an $\mathbf{R}^d$-valued function defined on $\mathbf{R}^d \times \Theta_\alpha$ and σ is an $\mathbf{R}^d \otimes \mathbf{R}^r$-valued function defined on $\mathbf{R}^d \times \Theta_\beta$. The data are discrete observations $\mathbf{X}_n = (X_{t_k^n})_{0 \le k \le n}$ with $t_k^n = kh_n$, where h_n is the discretization step. We will treat asymptotics when $h_n \to 0$, $nh_n \to \infty$ and $nh_n^2 \to 0$ as $n \to \infty$.

Keywords and phrases. Diffusion process, misspecified model, discrete time observations, minimum contrast estimator, rate of convergence.

[1] Graduate School of Engineering Science, Osaka University Toyonaka, Osaka 560-8531, Japan; uchida@sigmath.es.osaka-u.ac.jp

[2] Japan Science and Technology Agency, CREST, Japan

[3] Graduate School of Mathematical Sciences, University of Tokyo 3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan

Article published by EDP Sciences

The parametric inference for correctly specified ergodic diffusion processes from discrete observations has been studied by many researchers, see Prakasa Rao [14,15], Florens-Zmirou [2], Yoshida [18,19], Bibby and Sørensen [1], Kessler [7] and references therein. Here the correctly specified diffusion model means that there exists a true parameter value $\theta_0 = (\alpha_0, \beta_0) \in \Theta_\alpha \times \Theta_\beta$ such that $b(x, \alpha_0) = B(x)$ and $[\sigma\sigma^\star](x, \beta_0) = [SS^\star](x)$ for all x, where $\star$ denotes the transpose. For both the estimator $\hat{\alpha}_n$ of the drift parameter α and the estimator $\hat{\beta}_n$ of the diffusion coefficient parameter β in the correctly specified case of discretely observed ergodic diffusion processes, Yoshida [18] showed that under some regularity conditions, $(\sqrt{nh_n}(\hat{\alpha}_n - \alpha_0), \sqrt{n}(\hat{\beta}_n - \beta_0))$ converges in distribution to a normal random variable, which means that the rate of convergence for $\hat{\alpha}_n$ is different from the one for $\hat{\beta}_n$, see also Kessler [7]. On the other hand, the parameter estimation for misspecified diffusion models has been mainly investigated for the case where the whole path $X = \{X_t; t \in [0, T]\}$ is observed, see McKeague [11], Yoshida [17] and Kutoyants [9]. They proved that under some regularity conditions, $\sqrt{T}(\hat{\alpha}_T - \alpha^*)$ converges in distribution to a normal random variable, where $\hat{\alpha}_T$ is the maximum likelihood estimator and α^* is the quasi-optimal parameter. Although there have been applications of parametric estimation for discretely observed misspecified diffusion models (for example, information criteria for selecting the best model among competing misspecified models, see Uchida and Yoshida [16]), there seems no theoretical work on discretely observed misspecified diffusion models to the authors' knowledge.

In this paper, we consider parametric estimation for misspecified models from the discrete observations $\mathbf{X}_n$. The contrast function based on a locally Gaussian approximation (the Euler-Maruyama approximation) is used and we treat the following two kinds of misspecified diffusion models: (i) completely misspecified model, which means that a family of drift functions $\{b(x, \alpha), \alpha \in \Theta_\alpha\}$ may or may not include $B(x)$, and for $j = 1, \dots, q$, $g_{\sigma,j}(x)$ is not identically equal to zero, where $g_{\sigma,j}(x) = \frac{1}{2}\mathrm{tr}\left\{ (\frac{\partial}{\partial \beta_j}[\sigma\sigma^\star]^{-1})(x, \beta^*) \left([SS^\star](x) - [\sigma\sigma^\star](x, \beta^*)\right) \right\}$ and β^* is a quasi-optimal parameter defined in Section 2 below, (ii) semi-misspecified model, that is, $\{b(x, \alpha), \alpha \in \Theta_\alpha\}$ may or may not include $B(x)$, and for $j = 1, \dots, q$, $g_{\sigma,j}(x) = 0$ for all x. In both cases, we show that the minimum contrast estimator has asymptotic normality. It is worth stating that the rate of convergence for the diffusion coefficient estimator in the completely misspecified case turns out to be $\sqrt{nh_n}$ while the one in the semi-misspecified case is still $\sqrt{n}$.

This paper is organized as follows. In Section 2, a contrast function based on a locally Gaussian approximation to the transition density is introduced, and consistency of the minimum contrast estimator obtained from the contrast function is stated. Moreover, for both the completely misspecified case and the semi-misspecified case, asymptotic normality of the minimum contrast estimator is presented. Section 3 gives two examples and simulation studies. Section 4 is devoted to the proofs of the results in Section 2. The conclusion of this paper and the discussion on the results are given in Section 5.

2. Minimum contrast estimators

2.1. Contrast function

Let $C_\uparrow^{k,l}(\mathbf{R}^d \times \Theta; \mathbf{R}^d)$ denote the space of all functions f satisfying the following conditions: (i) $f(x, \theta)$ is an $\mathbf{R}^d$-valued function on $\mathbf{R}^d \times \Theta$; (ii) $f(x, \theta)$ is continuously differentiable with respect to x up to order k for all θ, and their derivatives up to order k are of polynomial growth in x uniformly in θ; (iii) for $|\mathbf{n}| = 0, 1, \dots, k$, $\partial^{\mathbf{n}} f(x, \theta)$ is continuously differentiable with respect to θ up to order l for all x. Moreover, for $|\nu| = 1, \dots, l$ and $|\mathbf{n}| = 0, 1, \dots, k$, $\delta^\nu \partial^{\mathbf{n}} f(x, \theta)$ is of polynomial growth in x uniformly in θ. Here $\mathbf{n} = (n_1, \dots, n_d)$ and $\nu = (\nu_1, \dots, \nu_m)$ are multi-indices, $m = \dim(\Theta)$, $|\mathbf{n}| = n_1 + \ldots + n_d$, $|\nu| = \nu_1 + \ldots + \nu_m$, $\partial^{\mathbf{n}} = \partial_1^{n_1} \cdots \partial_d^{n_d}$, $\partial_i = \partial/\partial x_i$, and $\delta^\nu = \delta_{\theta_1}^{\nu_1} \cdots \delta_{\theta_m}^{\nu_m}$, $\delta_{\theta_i} = \partial/\partial\theta_i$. Let $C_\uparrow^k(\mathbf{R}^d; \mathbf{R}^d)$ be the space of all functions f satisfying that $f(x)$ is an $\mathbf{R}^d$-valued function on $\mathbf{R}^d$, $f(x)$ is continuously differentiable with respect to x up to order k and their derivatives up to order k are of polynomial growth in x. Let $\mathcal{F}_\uparrow(\mathbf{R}^d)$ be the space of all measurable functions f satisfying that $f(x)$ is an $\mathbf{R}$-valued function on $\mathbf{R}^d$ with polynomial growth in x. Let L be the infinitesimal generator of the diffusion (1.1): $L = \sum_{i=1}^d B^i(x)\partial_i + \frac{1}{2}\sum_{i,j=1}^d [SS^\star]^{ij}(x)\partial_i\partial_j$. Let $\to^p$ and $\to^d$ be the convergence in probability and the convergence in distribution, respectively.

In this paper, we make the following assumptions.

*A*1 (i) There exists $L_0 > 0$ such that for all x, y,

$$|B(x) - B(y)| + |S(x) - S(y)| \le L_0|x - y|.$$

(ii) $B \in C^2_\uparrow(\mathbf{R}^d; \mathbf{R}^d)$ and $S \in C^2_\uparrow(\mathbf{R}^d; \mathbf{R}^d \otimes \mathbf{R}^r)$.

(iii) There exists a unique invariant probability measure μ of X_t and for any $g \in \mathcal{F}_\uparrow(\mathbf{R}^d)$ satisfying $\int_{\mathbf{R}^d} |g(x)|\mu(\mathrm{d}x) < \infty$, as $T \to \infty$,

$$\frac{1}{T}\int_0^T g(X_t) \to^p \int_{\mathbf{R}^d} g(x)\mu(\mathrm{d}x).$$

(iv) $\sup_t E[|X_t|^p] < \infty$ for all $p > 0$.

*A*2 (i) There exists $L_1 > 0$ such that for all x, y,

$$\sup_\alpha |b(x,\alpha) - b(y,\alpha)| + \sup_\beta |\sigma(x,\beta) - \sigma(y,\beta)| \le L_1|x - y|.$$

(ii) $b \in C^{2,3}_\uparrow(\mathbf{R}^d \times \Theta_\alpha; \mathbf{R}^d)$ and $\sigma \in C^{2,3}_\uparrow(\mathbf{R}^d \times \Theta_\beta; \mathbf{R}^d \otimes \mathbf{R}^r)$.

(iii) There exists $c > 0$ such that $\inf_\beta \det[\sigma\sigma^\star](x,\beta) \ge \frac{1}{c(1+|x|^c)}$ for all x.

Remark 2.1. (i) As sufficient conditions for *A*1-(iii)–(iv), we make the following assumptions. *A*1-$(iii)'$: $\inf_x \det[SS^\star](x) > 0$, *A*1-$(iv)'$: there exist $f \in Q$ and positive constants c_1^* and c_2^* such that $Lf \le -c_1^* f + c_2^*$, where $Q = \left\{f \in C^2(\mathbf{R}^d; \mathbf{R}_+) | \lim_{|x|\to\infty} \frac{f(x)}{|x|^p} = +\infty \text{ for all } p \ge 0\right\}$, and *A*1-$(v)'$: $E[f(X_0)] < \infty$. It follows from *A*1-(i)–(ii) and *A*1-$(iii)'$–$(v)'$ that *A*1-(iii)–(iv) hold true, see Meyn and Tweedie [12] and Masuda [10]; (ii) as a sufficient condition for *A*1-$(iv)'$, we make the assumption *A*1-$(iv)''$: (a) there exist $c_0, K_0 > 0$ such that $x^\star B(x) \le -c_0|x|^2 + K_0$ for all x. (b) There exists $c_1 > 0$ such that $\sum_{i,j=1}^d [SS^\star]^{ij}(x)\lambda_i\lambda_j \le c_1|\lambda|^2$ for all x, λ. Then *A*1-$(iv)'$ holds true for the case that $f(x) = \exp\{c_2|x|^2\}$ for $c_2 \in (0, c_0/c_1)$, see Gobet [4]. (iii) For another sufficient condition for *A*1-(iii)–(iv), we can refer Kusuoka and Yoshida [8].

The contrast function is as follows:

$$u_n(\theta) \quad = \quad \frac{1}{2}\sum_{k=1}^n \left\{\log\det\Xi(X_{t_{k-1}^n},\beta) + \frac{1}{h_n}\Delta_k^\star(\alpha)\Xi^{-1}(X_{t_{k-1}^n},\beta)\Delta_k(\alpha)\right\},$$

where $\Xi(x,\beta) = [\sigma\sigma^\star](x,\beta)$, $\Delta_k(\alpha) = X_{t_k^n} - X_{t_{k-1}^n} - h_n b(X_{t_{k-1}^n},\alpha)$. Let $\hat\theta_n = (\hat\alpha_n, \hat\beta_n)$ be a minimum contrast estimator defined as

$$u_n(\hat\theta_n) = \inf_\theta u_n(\theta). \tag{2.1}$$

Let

$$D_1(\beta) \quad = \quad \frac{1}{2}\int_{\mathbf{R}^d} \left\{\log\det\Xi(x,\beta) + \mathrm{tr}\left([SS^\star](x)\Xi^{-1}(x,\beta)\right)\right\}\mu(\mathrm{d}x),$$

$$D_2(\alpha,\beta) \quad = \quad \frac{1}{2}\int_{\mathbf{R}^d} \left(B(x) - b(x,\alpha)\right)^\star \Xi^{-1}(x,\beta)\left(B(x) - b(x,\alpha)\right)\mu(\mathrm{d}x).$$

Set $\theta^* = (\alpha^*, \beta^*)$, where α^* and β^* are the quasi-optimal parameters defined by $\beta^* = \arg\inf_\beta D_1(\beta)$ and $\alpha^* = \arg\inf_\alpha D_2(\alpha, \beta^*)$. Suppose that $\theta^* \in \mathrm{Int}(\Theta)$.

In order to obtain the consistency of $\hat\theta_n$, we make the assumption as follows.

*A*3 (i) For any $\epsilon > 0$, $\inf_{|\beta-\beta^*|\geq\epsilon}[D_1(\beta) - D_1(\beta^*)] > 0$.
(ii) For any $\epsilon > 0$, $\inf_{|\alpha-\alpha^*|\geq\epsilon}[D_2(\alpha,\beta^*) - D_2(\alpha^*,\beta^*)] > 0$.
The result of the consistency is as follows.

Proposition 2.1. *Assume A*1*–A*3*. Then,* $\hat{\theta}_n \to^p \theta^*$ *as* $h_n \to 0$ *and* $nh_n \to \infty$*.*

2.2. Completely misspecified case

We consider the situation where (i) a family of drift functions $\{b(x,\alpha), \alpha \in \Theta_\alpha\}$ may or may not include $B(x)$, (ii) for $j = 1,\dots,q$, $g_{\sigma,j}(x)$ is not identically equal to zero, where

$$g_{\sigma,j}(x) = \frac{1}{2}\mathrm{tr}\left\{(\delta_{\beta_j}\Xi^{-1})(x,\beta^*)\left([SS^\star](x) - \Xi(x,\beta^*)\right)\right\}.$$

In order to state the sufficient condition for the asymptotic normality of $\hat{\theta}_n$, we set

$$J(\theta) = \begin{pmatrix} \left(J_{bb}^{ij}(\theta)\right)_{i,j=1,\dots,p} & \left(J_{b\sigma}^{ij}(\theta)\right)_{\substack{i=1,\dots,p\\ j=1,\dots,q}} \\ 0 & \left(J_{\sigma\sigma}^{ij}(\beta)\right)_{i,j=1,\dots,q} \end{pmatrix},$$

where

$$\begin{aligned} J_{bb}^{ij}(\theta) &= \int_{\mathbf{R}^d} (\delta_{\alpha_i}b)^\star\Xi^{-1}(x,\beta)(\delta_{\alpha_j}b)(x,\alpha)\mu(\mathrm{d}x) \\ &\quad - \int_{\mathbf{R}^d} (\delta_{\alpha_i}\delta_{\alpha_j}b)^\star(x,\alpha)\Xi^{-1}(x,\beta)\left\{B(x) - b(x,\alpha)\right\}\mu(\mathrm{d}x), \\ J_{b\sigma}^{ij}(\theta) &= -\int_{\mathbf{R}^d} (\delta_{\alpha_i}b)^\star(x,\alpha)(\delta_{\beta_j}\Xi^{-1})(x,\beta)(B(x) - b(x,\alpha))\mu(\mathrm{d}x), \\ J_{\sigma\sigma}^{ij}(\beta) &= \frac{1}{2}\int_{\mathbf{R}^d} \mathrm{tr}\left((\delta_{\beta_i}\Xi)\Xi^{-1}(\delta_{\beta_j}\Xi)\Xi^{-1}(x,\beta)\right)\mu(\mathrm{d}x) \\ &\quad + \frac{1}{2}\int_{\mathbf{R}^d} \mathrm{tr}\left((\delta_{\beta_i}\delta_{\beta_j}\Xi^{-1})(x,\beta)([SS^\star](x) - \Xi(x,\beta))\right)\mu(\mathrm{d}x). \end{aligned}$$

In addition to *A*1–*A*3, we need the following assumptions.
*A*4 There exist functions $G_{b,i}$, $G_{\sigma,j}$, $\partial_k G_{b,i}$, $\partial_k G_{\sigma,j} \in \mathcal{F}_\uparrow(\mathbf{R}^d)$, $i = 1,\dots,p$, $j = 1,\dots,q$, $k = 1,\dots,d$ such that

$$\begin{aligned} LG_{b,i}(x) &= (\delta_{\alpha_i}b)^\star(x,\alpha^*)\Xi^{-1}(x,\beta^*)\left\{B(x) - b(x,\alpha^*)\right\}, \\ LG_{\sigma,j}(x) &= g_{\sigma,j}(x). \end{aligned}$$

*A*5 $J(\theta^*)$ is invertible.

Remark 2.2. (i) For sufficient conditions satisfying that $G_{b,i}$, $G_{\sigma,j}$, $\partial_k G_{b,i}$, $\partial_k G_{\sigma,j} \in \mathcal{F}_\uparrow(\mathbf{R}^d)$ in *A*4, we can refer Pardoux and Veretennikov [13]. For example, the assumptions *A*1-(*i*)–(*ii*), *A*1-(*iii*)′, *A*1-(*iv*)″ in Remark 2.1 and *A*2 imply *A*4. (ii) In the case that $d = r = 1$ and $\mu(\mathrm{d}x) = v(x)\mathrm{d}x$, under mild regularity conditions, both $\partial_x G_{b,i}(x)$ and $\partial_x G_{\sigma,j}(x)$ have the following explicit forms:

$$\begin{aligned} \partial_x G_{b,i}(x) &= \frac{2}{S(x)^2 v(x)}\int_{-\infty}^x \frac{\{B(y) - b(y,\alpha^*)\}(\delta_{\alpha_i}b)(y,\alpha^*)}{\sigma^2(y,\beta^*)}\mu(\mathrm{d}y), \\ \partial_x G_{\sigma,j}(x) &= \frac{-2}{S(x)^2 v(x)}\int_{-\infty}^x \frac{\{S^2(y) - \sigma^2(y,\beta^*)\}(\delta_{\beta_j}\sigma)(y,\beta^*)}{\sigma^3(y,\beta^*)}\mu(\mathrm{d}y). \end{aligned}$$

Let

$$K = \begin{pmatrix} \left(K_{bb}^{ij}\right)_{i,j=1,\ldots,p} & \left(K_{b\sigma}^{ij}\right)_{\substack{i=1,\ldots,p \\ j=1,\ldots,q}} \\ \left(K_{b\sigma}^{ij}\right)^{\star}_{\substack{i=1,\ldots,p \\ j=1,\ldots,q}} & \left(K_{\sigma\sigma}^{ij}\right)_{i,j=1,\ldots,q} \end{pmatrix},$$

where

$$\begin{aligned} K_{bb}^{ij} &= \int_{\mathbf{R}^d} \left((\delta_{\alpha_i} b)^{\star}(x,\alpha^*)\Xi^{-1}(x,\beta^*) - (\partial_x G_{b,i})(x)\right)[SS^{\star}](x) \\ &\quad \times \left((\delta_{\alpha_j} b)^{\star}(x,\alpha^*)\Xi^{-1}(x,\beta^*) - (\partial_x G_{b,j})(x)\right)^{\star} \mu(\mathrm{d}x), \\ K_{b\sigma}^{ij} &= \int_{\mathbf{R}^d} \left((\delta_{\alpha_i} b)^{\star}(x,\alpha^*)\Xi^{-1}(x,\beta^*) - (\partial_x G_{b,i})(x)\right)[SS^{\star}](x)(\partial_x G_{\sigma,j})^{\star}(x)\mu(\mathrm{d}x), \\ K_{\sigma\sigma}^{ij} &= \int_{\mathbf{R}^d} (\partial_x G_{\sigma,i})(x)[SS^{\star}](x)(\partial_x G_{\sigma,j})^{\star}(x)\mu(\mathrm{d}x). \end{aligned}$$

The result of asymptotic normality is as follows.

Theorem 2.1. *Assume A1–A5. Then, as $h_n \to 0$, $nh_n \to \infty$ and $nh_n^2 \to 0$,*

$$\begin{pmatrix} \sqrt{nh_n}(\hat{\alpha}_n - \alpha^*) \\ \sqrt{nh_n}(\hat{\beta}_n - \beta^*) \end{pmatrix} \to^d N(0, J^{-1}(\theta^*)K(J^{\star})^{-1}(\theta^*)).$$

Remark 2.3. As seen from Theorem 2.1, if the asymptotic covariance matrix $J^{-1}(\theta^*)K(J^{\star})^{-1}(\theta^*)$ is non-degenerate, then the rate of convergence for the estimator $\hat{\beta}_n$ of the diffusion coefficient parameter β in the completely misspecified case is $\sqrt{nh_n}$, which is different from the one of the correctly specified parametric case. Meanwhile, the rate of convergence for the estimator $\hat{\alpha}_n$ of the drift parameter α in the completely misspecified case is the same as the one of the correctly specified parametric case. For the intuitive reason why the rate of convergence for $\hat{\beta}_n$ in the completely misspecified case is worse than the one of the correctly specified parametric case, see Section 5 below.

2.3. Semi-misspecified case

In this subsection, we treat the case that (i) a family of drift functions $\{b(x,\alpha), \alpha \in \Theta_\alpha\}$ may or may not include $B(x)$, (ii) for $j = 1,\ldots,q$, $g_{\sigma,j}(x) = 0$ for all x, where $g_{\sigma,j}(x)$ is defined in Section 2.2. We call it the semi-misspecified case. If a family of diffusion functions $\{[\sigma\sigma^{\star}](x,\beta), \beta \in \Theta_\beta\}$ includes $[SS^{\star}](x)$, that is, there exists a true parameter $\beta^* \in \Theta_\beta$ such that $[\sigma\sigma^{\star}](x,\beta^*) = [SS^{\star}](x)$ for all x, then the above condition (ii) is satisfied.

Let

$$\bar{J}(\theta) = \begin{pmatrix} \left(J_{bb}^{ij}(\theta)\right)_{i,j=1,\ldots,p} & 0 \\ 0 & \left(J_{\sigma\sigma}^{ij}(\beta)\right)_{i,j=1,\ldots,q} \end{pmatrix}$$

and

$$\bar{K} = \begin{pmatrix} \left(K_{bb}^{ij}\right)_{i,j=1,\ldots,p} & 0 \\ 0 & \left(\bar{K}_{\sigma\sigma}^{ij}\right)_{i,j=1,\ldots,q} \end{pmatrix},$$

where

$$\bar{K}_{\sigma\sigma}^{ij} = \frac{1}{2}\int_{\mathbf{R}^d} \mathrm{tr}\left((\delta_{\beta_i}\Xi^{-1})(x,\beta^*)[SS^{\star}](x)(\delta_{\beta_j}\Xi^{-1})(x,\beta^*)[SS^{\star}](x)\right)\mu(\mathrm{d}x).$$

We make the following assumptions.

A4′ There exist functions $G_{b,i}$, $\partial_k G_{b,i} \in \mathcal{F}_{\uparrow}(\mathbf{R}^d)$, $i = 1, \ldots, p$, $k = 1, \ldots, d$ such that

$$LG_{b,i}(x) = (\delta_{\alpha_i} b)^{\star}(x, \alpha^*)\Xi^{-1}(x, \beta^*)\left\{B(x) - b(x, \alpha^*)\right\}.$$

A5′ $\bar{J}(\theta^*)$ is invertible.

The result of asymptotic normality is as follows.

Theorem 2.2. *Assume A1–A3 and A4′–A5′. Then, as $h_n \to 0$, $nh_n \to \infty$ and $nh_n^2 \to 0$,*

$$\begin{pmatrix} \sqrt{nh_n}(\hat{\alpha}_n - \alpha^*) \\ \sqrt{n}(\hat{\beta}_n - \beta^*) \end{pmatrix} \to^d N(0, \bar{J}^{-1}(\theta^*)\bar{K}\bar{J}^{-1}(\theta^*)).$$

Remark 2.4. Following the proof of Lemma 6 below, we can show that Theorem 2.2 still holds true even if $\frac{1}{\sqrt{n}}\sum_{k=1}^{n} g_{\sigma,j}(X_{t_{k-1}^n}) = o_p(1)$ for $j = 1, \ldots, q$.

3. Examples

3.1. Completely misspecified case

As an example of the completely misspecified case, we consider the one-dimensional ergodic and stationary diffusion process

$$\mathrm{d}X_t = -\frac{1}{2}X_t\mathrm{d}t + \mathrm{d}w_t, \quad t \in [0, T], \quad X_0 \sim \mu, \tag{3.1}$$

where μ is the invariant distribution, $\mu(\mathrm{d}x) = \phi(x)\mathrm{d}x$ and $\phi(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}$. We assume the statistical model

$$\mathrm{d}X_t = -\alpha(X_t - 1)\mathrm{d}t + \frac{\beta}{\sqrt{1 + X_t^2}}\mathrm{d}w_t, \quad t \in [0, T], \quad X_0 \sim \mu, \tag{3.2}$$

where $\alpha, \beta > 0$. The contrast function for (3.2) with $\theta = (\alpha, \beta)$ is

$$u_n(\theta) = \frac{1}{2}\sum_{k=1}^{n}\left\{\log\frac{\beta^2}{1 + X_{t_{k-1}^n}^2} + \frac{\left\{X_{t_k^n} - X_{t_{k-1}^n} + h_n\alpha(X_{t_{k-1}^n} - 1)\right\}^2}{h_n\frac{\beta^2}{1 + X_{t_{k-1}^n}^2}}\right\}.$$

The quasi-optimal parameters for α and β are

$$\beta^* = \arg\inf_{\beta}\int_{\mathbf{R}}\left\{\log\frac{\beta^2}{1 + x^2} + \frac{1}{\frac{\beta^2}{1+x^2}}\right\}\mu(\mathrm{d}x) = \sqrt{2},$$

$$\alpha^* = \arg\inf_{\alpha}\int_{\mathbf{R}}\frac{\left\{-\frac{1}{2}x + \alpha(x - 1)\right\}^2}{\frac{(\beta^*)^2}{1+x^2}}\mu(\mathrm{d}x) = \frac{1}{3}.$$

The minimum contrast estimators of α and β are

$$\hat{\alpha}_n = -\frac{\sum_{k=1}^{n}(X_{t_{k-1}^n} - 1)(X_{t_k^n} - X_{t_{k-1}^n})(X_{t_{k-1}^n}^2 + 1)}{h_n\sum_{k=1}^{n}(X_{t_{k-1}^n} - 1)^2(X_{t_{k-1}^n}^2 + 1)}, \tag{3.3}$$

$$\hat{\beta}_n = \sqrt{\frac{1}{nh_n}\sum_{k=1}^{n}\left\{X_{t_k^n} - X_{t_{k-1}^n} + h_n\hat{\alpha}_n(X_{t_{k-1}^n} - 1)\right\}^2(X_{t_{k-1}^n}^2 + 1)}. \tag{3.4}$$

TABLE 1. The mean and s.d. of the estimators for 100 000 independent simulated sample paths. The quasi-optimal values for α and β are $\alpha^* = \frac{1}{3}$ and $\beta^* = \sqrt{2}$, respectively.

T	h_n	$\hat{\alpha}_n$		$\hat{\beta}_n$	
		mean	s.d.	mean	s.d.
50	1/200	0.3856	0.1212	1.4098	0.0982
70	1/200	0.3718	0.0995	1.4106	0.0832
90	1/200	0.3637	0.0865	1.4108	0.0743

Next, we calculate the asymptotic covariance matrix of $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n)$. An easy computation yields that

$$\begin{aligned} J_{bb}(\theta^*) &= \int_{\mathbf{R}} \frac{(x-1)^2}{\frac{(\beta^*)^2}{1+x^2}} \mu(\mathrm{d}x) = 3, \\ J_{b\sigma}(\theta^*) &= 2\int_{\mathbf{R}} \frac{-(x-1)(1+x^2)\left\{-\frac{1}{2}x + \alpha^*(x-1)\right\}}{(\beta^*)^3} \mu(\mathrm{d}x) = 0, \\ J_{\sigma\sigma}(\beta^*) &= 2\int_{\mathbf{R}} \frac{1}{(\beta^*)^2}\mu(\mathrm{d}x) + 3\int_{\mathbf{R}} \frac{1+x^2-(\beta^*)^2}{(\beta^*)^4}\mu(\mathrm{d}x) = 1. \end{aligned}$$

Moreover, since

$$\begin{aligned} \partial_x G_b(x) &= -\frac{2}{\phi(x)}\int_{-\infty}^{x} \frac{\left\{-\frac{1}{2}y + \alpha^*(y-1)\right\}(y-1)}{\frac{(\beta^*)^2}{1+y^2}} \phi(y)\mathrm{d}y = -\frac{1}{6}(x^3+x^2+2x+3), \\ \partial_x G_\sigma(x) &= -\frac{2}{\phi(x)}\int_{-\infty}^{x} \frac{1+y^2-(\beta^*)^2}{(\beta^*)^3}\phi(y)\mathrm{d}y = \frac{x}{\sqrt{2}}, \end{aligned}$$

one has that $K_{bb} = \frac{205}{36}, K_{b\sigma} = -\frac{7}{6\sqrt{2}}, K_{\sigma\sigma} = \frac{1}{2}$. Thus,

$$J^{-1}(\theta^*)KJ^{-1}(\theta^*) = \begin{pmatrix} \frac{205}{324} & -\frac{7}{18\sqrt{2}} \\ -\frac{7}{18\sqrt{2}} & \frac{1}{2} \end{pmatrix}.$$

Here we examine the asymptotic behaviour of the estimator $\hat{\theta}_n$ through the simulations, which were done for each $T = 50, 70, 90$ and $h_n = 1/200$. For the true model (3.1), 100 000 independent sample paths are generated by the exact simulation, and the mean and the standard deviation (s.d.) for the estimators (3.3) and (3.4) are computed and shown in Table 1 below.

In Table 1, $\hat{\beta}_n$ is unbiased in all cases, and $\sqrt{T}\times$(the sample s.d.) is close to the asymptotic s.d. of $\sqrt{T}(\hat{\beta}_n - \beta^*)$, which is equal to $\sqrt{0.5} \simeq 0.7071$ by Theorem 2.1. In special, $\sqrt{T} \times$ (the sample s.d.) $\simeq 0.7050$ when $T = 90$. We see that $\hat{\beta}_n$ gives good results in all cases. On the other hand, since $\hat{\alpha}_n$ has a bias even when $T = 90$, we will need to set that T is more than 90 in order to get a good estimate of α^*.

3.2. **Semi-misspecified case**

We consider the two-dimensional ergodic diffusion process

$$\mathrm{d}X_t = -AX_t\mathrm{d}t + \mathrm{d}w_t, \quad t \in [0,T], \quad X_0 = (1,-1)^\star, \tag{3.5}$$

where w is a two-dimensional standard Wiener process, $A = \begin{pmatrix} 1/2 & 1/4 \\ 1/4 & 1/2 \end{pmatrix}$, the invariant distribution is $\mu(\mathrm{d}x) = \frac{\sqrt{3}}{4\pi}\exp\{-V(x)\}\mathrm{d}x$ and $V(x) = \frac{1}{2}(x_1^2 + x_1x_2 + x_2^2)$. We assume the statistical model

$$\mathrm{d}X_t \quad = \quad -\alpha AX_t\mathrm{d}t + \sqrt{\beta}B\mathrm{d}w_t, \quad t \in [0,T], \quad X_0 = (1,-1)^\star, \tag{3.6}$$

where α and β are positive constants and $B = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$. Note that in this example, $[SS^\star](x) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\Xi(x,\beta) = \beta B^2$ and $\Xi^{-1}(x,\beta) = \frac{1}{\beta}C$, where $C = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$.

The contrast function for (3.6) with $\theta = (\alpha,\beta)$ is

$$u_n(\theta) = \frac{1}{2}\sum_{k=1}^{n}\left\{\log\beta^2 + \frac{1}{h_n\beta}(X_{t_k^n} - X_{t_{k-1}^n} + h_n\alpha AX_{t_{k-1}^n})^\star C(X_{t_k^n} - X_{t_{k-1}^n} + h_n\alpha AX_{t_{k-1}^n})\right\}.$$

The quasi-optimal parameters for α and β are

$$\begin{aligned} \beta^* &= \arg\inf_\beta \int_{\mathbf{R}^2}\left\{\log\beta^2 + \frac{3}{\beta}\right\}\mu(\mathrm{d}x) = \frac{3}{2}, \\ \alpha^* &= \arg\inf_\alpha \int_{\mathbf{R}^2}(-Ax + \alpha Ax)^\star C(-Ax + \alpha Ax)\mu(\mathrm{d}x) = 1, \end{aligned}$$

where we note that α^* is the true value of α. The minimum contrast estimators of α and β are

$$\hat{\alpha}_n \quad = \quad -\frac{\sum_{k=1}^n X_{t_{k-1}^n}^\star AC(X_{t_k^n} - X_{t_{k-1}^n})}{h_n\sum_{k=1}^n X_{t_{k-1}^n}^\star ACAX_{t_{k-1}^n}}, \tag{3.7}$$

$$\hat{\beta}_n \quad = \quad \frac{1}{2nh_n}\sum_{k=1}^{n}\Big(X_{t_k^n} - X_{t_{k-1}^n} + h_n\hat{\alpha}_nAX_{t_{k-1}^n}\Big)^\star C\Big(X_{t_k^n} - X_{t_{k-1}^n} + h_n\hat{\alpha}_nAX_{t_{k-1}^n}\Big). \tag{3.8}$$

It is easy to see that $\mathrm{tr}\left\{(\delta_\beta\Xi^{-1})(x,\beta^*)\left([SS^\star](x) - \Xi(x,\beta^*)\right)\right\} = -\frac{3}{(\beta^*)^2} + \frac{2}{\beta^*} = 0$, which means that this example is the semi-misspecified case. Furthermore, one has that

$$\begin{aligned} J_{bb}(\theta^*) &= \frac{1}{\beta^*}\int_{\mathbf{R}^2} x^\star ACAx\mu(\mathrm{d}x) = \frac{1}{2\beta^*}, \\ J_{\sigma\sigma}(\beta^*) &= \frac{1}{2(\beta^*)^2}\int_{\mathbf{R}^2}\mathrm{tr}(C^{-1}CC^{-1}C)\mu(\mathrm{d}x) = \frac{1}{(\beta^*)^2}, \\ K_{bb} &= \frac{1}{(\beta^*)^2}\int_{\mathbf{R}^2} x^\star AC^2Ax\mu(\mathrm{d}x) = \frac{1}{(\beta^*)^2}, \\ \bar{K}_{\sigma\sigma} &= \frac{1}{2(\beta^*)^4}\int_{\mathbf{R}^2}\mathrm{tr}(C^2)\mu(\mathrm{d}x) = \frac{7}{2(\beta^*)^4}. \end{aligned}$$

Thus, the asymptotic covariance matrix of $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n)$ is

$$\bar{J}^{-1}(\theta^*)\bar{K}\bar{J}^{-1}(\theta^*) = \begin{pmatrix} 4 & 0 \\ 0 & \frac{7}{2} \end{pmatrix}.$$

For the same T and h_n as the previous example, we examine the asymptotic behaviour of the estimator $\hat{\theta}_n$ through the simulations, and the mean and the s.d. for the estimators (3.7) and (3.8) are given in Table 2 below.

TABLE 2. The mean and s.d. of the estimators for 100000 independent simulated sample paths. The true value for α is $\alpha^*=1$ and the quasi-optimal value for β is $\beta^*=\frac{3}{2}$.

T	h_n	$\hat{\alpha}_n$		$\hat{\beta}_n$	
		mean	s.d.	mean	s.d.
50	1/200	1.0697	0.2893	1.4979	0.0186
70	1/200	1.0493	0.2425	1.4979	0.0157
90	1/200	1.0406	0.2130	1.4978	0.0139

In Table 2, $\hat{\alpha}_n$ has a little bit bias, but $\hat{\alpha}_n$ gives a good performance as compared with the one of the previous example, which may result from the fact that the drift function is correctly specified. $\hat{\beta}_n$ is unbiased in all cases, and when $T=90$, $\sqrt{n}\times$(the sample s.d.) $\simeq 1.8648$, while the asymptotic s.d. of $\sqrt{n}(\hat{\beta}_n-\beta^*)$ is approximately 1.8708 by Theorem 2.2. The simulation result shows that $\hat{\beta}_n$ works well as an estimator of β^* in all cases.

4. PROOFS

Let $\mathcal{G}^n_{k-1}$ denote the *history* up to the time t^n_{k-1}. Let R be a function $\Theta\times(0,1]\times\mathbf{R}^d\to\mathbf{R}$ for which there exists a constant C such that $|R(\theta,a,x)|\le aC(1+|x|)^C$ for all θ,a,x.

4.1. Proof of Proposition 2.1

In order to prove Proposition 2.1, the following lemmas are required.

Lemma 1. *Under A1–A2,*

(i)

$$E[\Delta_k^{i_1}(\alpha)|\mathcal{G}^n_{k-1}]=h_n(B^{i_1}(X_{t^n_{k-1}})-b^{i_1}(X_{t^n_{k-1}},\alpha))+R(\theta,h_n^2,X_{t^n_{k-1}}),$$

(ii)

$$E[\Delta_k^{i_1}(\alpha)\Delta_k^{i_2}(\alpha)|\mathcal{G}^n_{k-1}] = h_n[SS^\star]^{i_1i_2}(X_{t^n_{k-1}})+R(\theta,h_n^2,X_{t^n_{k-1}}),$$

(iii)

$$E[\Delta_k^{i_1}(\alpha)\Delta_k^{i_2}(\alpha)\Delta_k^{i_3}(\alpha)|\mathcal{G}^n_{k-1}]=R(\theta,h_n^2,X_{t^n_{k-1}}),$$

(iv)

$$\begin{aligned}E\left[\prod_{j=1}^4\Delta_k^{i_j}(\alpha)\,\middle|\,\mathcal{G}^n_{k-1}\right] &= h_n^2\Big\{[SS^\star]^{i_1i_2}[SS^\star]^{i_3i_4}(X_{t^n_{k-1}})+[SS^\star]^{i_1i_3}[SS^\star]^{i_2i_4}(X_{t^n_{k-1}})\\ &\quad+[SS^\star]^{i_1i_4}[SS^\star]^{i_2i_3}(X_{t^n_{k-1}})\Big\}+R(\theta,h_n^{5/2},X_{t^n_{k-1}}).\end{aligned}$$

Proof. In the same way as in the proof of Lemma 7 in Kessler [7], the Ito-Taylor expansion yields the results. This completes the proof. □

Lemma 2. *Let $f\in C^{1,1}_{\uparrow}(\mathbf{R}^d\times\Theta;\mathbf{R})$. Under A1–A2, as $h_n\to 0$ and $nh_n\to\infty$,*

(i)

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{k=1}^n f(X_{t^n_{k-1}},\theta)-\int_{\mathbf{R}^d}f(x,\theta)\mu(\mathrm{d}x)\right|\to^p 0,$$

(ii)

$$\sup_{\theta\in\Theta}\left|\frac{1}{nh_n}\sum_{k=1}^n f(X_{t^n_{k-1}},\theta)\Delta_k^{l_1}(\alpha)-\int_{\mathbf{R}^d}f(x,\theta)(B(x)-b(x,\alpha))^{l_1}\mu(\mathrm{d}x)\right|\to^p 0,$$

(iii)

$$\sup_{\theta\in\Theta}\left|\frac{1}{nh_n}\sum_{k=1}^{n} f(X_{t_{k-1}^n},\theta)\Delta_k^{l_1}(\alpha)\Delta_k^{l_2}(\alpha)-\int_{\mathbf{R}^d} f(x,\theta)[SS^\star]^{l_1l_2}(x)\mu(\mathrm{d}x)\right|\to^p 0.$$

Proof. (i) By the method used in the proof of Theorem 4.1 in Yoshida [17] or Lemma 8 in Kessler [7], we can show the result.
(ii) Let $\bar{\Delta}_k = X_{t_k^n} - X_{t_{k-1}^n} - h_n B(X_{t_{k-1}^n})$ and $\eta_k^{l_1}(\theta) = \frac{1}{nh_n} f(X_{t_{k-1}^n},\theta)\bar{\Delta}_k^{l_1}$. It is enough to prove that

$$\sup_{\theta\in\Theta}\left|\sum_{k=1}^n \eta_k^{l_1}(\theta)\right| \quad \to^p \quad 0. \tag{4.1}$$

By Lemma 1 and 2-(i), as $h_n\to 0$ and $nh_n\to\infty$,

$$\sum_{k=1}^n E\left[\eta_k^{l_1}(\theta)\middle|\mathcal{G}_{k-1}^n\right]\to^p 0,\quad \sum_{k=1}^n E\left[(\eta_k^{l_1}(\theta))^2\middle|\mathcal{G}_{k-1}^n\right]\to^p 0.$$

It follows from Lemma 9 of Genon-Catalot and Jacod [3] that $\sum_{k=1}^n \eta_k^{l_1}(\theta)\to^p 0$ for all θ. In order to prove the tightness of $\sum_{k=1}^n \eta_k^{l_1}(\cdot)$, it is sufficient to prove the following inequalities (cf. Theorem 20 in Appendix I of Ibragimov and Has'minskii [6] or Lemma 3.1 of Yoshida [17]:

$$E\left[\left(\sum_{k=1}^n \eta_k^{l_1}(\theta)\right)^{2l}\right] \quad\le\quad C, \tag{4.2}$$

$$E\left[\left(\sum_{k=1}^n \eta_k^{l_1}(\theta_2)-\sum_{k=1}^n \xi_k^{l_1}(\theta_1)\right)^{2l}\right] \quad\le\quad C|\theta_2-\theta_1|^{2l}, \tag{4.3}$$

for $\theta,\theta_1,\theta_2\in\Theta$, where $l>(p+q)/2$. We define $\xi_{k,1}^{l_1}(\theta)$ and $\xi_{k,2}^{l_1}(\theta)$ by

$$\begin{aligned}\eta_k^{l_1}(\theta) &= \frac{1}{nh_n} f(X_{t_{k-1}^n},\theta)\int_{t_{k-1}^n}^{t_k^n}(B^{l_1}(X_s)-B^{l_1}(X_{t_{k-1}^n}))\mathrm{d}s+\frac{1}{nh_n} f(X_{t_{k-1}^n},\theta)\int_{t_{k-1}^n}^{t_k^n}\sum_{j=1}^r S^{l_1j}(X_s)\mathrm{d}w_s^j\\ &=: \xi_{k,1}^{l_1}(\theta)+\xi_{k,2}^{l_1}(\theta).\end{aligned}$$

By the standard estimates, one has that

$$E\left[\left|\sum_{k=1}^n \xi_{k,1}^{l_1}(\theta)\right|^{2l}\right] \quad\le\quad \frac{1}{nh_n^{2l}}\sum_{k=1}^n E\left[\left(\int_{t_{k-1}^n}^{t_k^n}|f(X_{t_{k-1}^n},\theta)(B^{l_1}(X_s)-B^{l_1}(X_{t_{k-1}^n}))|\mathrm{d}s\right)^{2l}\right]\le C.$$

It follows from the Burkholder-Davis-Gundy inequality that

$$E\left[\left|\sum_{k=1}^n \xi_{k,2}^{l_1}(\theta)\right|^{2l}\right] \quad\le\quad \frac{C_1}{(nh_n)^{2l}}E\left[\left(\sum_{k=1}^n\int_{t_{k-1}^n}^{t_k^n} f(X_{t_{k-1}^n},\theta)^2[SS^\star]^{l_1l_1}(X_s)\mathrm{d}s\right)^l\right]\le C.$$

Therefore, we deduce the inequality (4.2). For the proof of (4.3), setting that $\bar{B}_{k-1}(s) = B(X_s) - B(X_{t_{k-1}^n})$ and $\tilde{\theta}_u = \theta_1 + u(\theta_2 - \theta_1)$, we first obtain that

$$E\left[\left(\sum_{k=1}^{n}\{\xi_{k,1}^{l_1}(\theta_2) - \xi_{k,1}^{l_1}(\theta_1)\}\right)^{2l}\right] \leq \frac{1}{nh_n^{2l}}\sum_{k=1}^{n} E\left[\left(\int_{t_{k-1}^n}^{t_k^n}\left|\{f(X_{t_{k-1}^n},\theta_2) - f(X_{t_{k-1}^n},\theta_1)\}\bar{B}_{k-1}^{l_1}(s)\right|\mathrm{d}s\right)^{2l}\right]$$
$$< C\sum_{j=1}^{p+q}|(\theta_2-\theta_1)^j|^{2l}.$$

Next, the Burkholder-Davis-Gundy inequality yields that

$$E\left[\left(\sum_{k=1}^{n}\{\xi_{k,2}^{l_1}(\theta_2) - \xi_{k,2}^{l_1}(\theta_1)\}\right)^{2l}\right] \leq \frac{1}{(nh_n)^{2l}}E\left[\left(\sum_{k=1}^{n}\int_{t_{k-1}^n}^{t_k^n}\{f(X_{t_{k-1}^n},\theta_2) - f(X_{t_{k-1}^n},\theta_1)\}^2[SS^\star]^{l_1l_1}(X_s)\mathrm{d}s\right)^{l}\right]$$
$$< C\sum_{j=1}^{p+q}|(\theta_2-\theta_1)^j|^{2l},$$

which completes the proof.
(iii) Since

$$\begin{aligned}\Delta_k^{l_1}(\alpha)\Delta_k^{l_2}(\alpha) &= \bar{\Delta}_k^{l_1}\bar{\Delta}_k^{l_2} + h_n\bar{\Delta}_k^{l_1}(B^{l_2}(X_{t_{k-1}^n}) - b^{l_2}(X_{t_{k-1}^n},\alpha)) \\ &\quad + h_n\bar{\Delta}_k^{l_2}(B^{l_1}(X_{t_{k-1}^n}) - b^{l_1}(X_{t_{k-1}^n},\alpha)) \\ &\quad + h_n^2(B^{l_1}(X_{t_{k-1}^n}) - b^{l_1}(X_{t_{k-1}^n},\alpha))(B^{l_2}(X_{t_{k-1}^n}) - b^{l_2}(X_{t_{k-1}^n},\alpha)),\end{aligned} \tag{4.4}$$

it is enough to prove that

$$\sup_{\theta\in\Theta}\left|\frac{1}{nh_n}\sum_{k=1}^{n} f(X_{t_{k-1}^n},\theta)\bar{\Delta}_k^{l_1}\bar{\Delta}_k^{l_2} - \int_{\mathbf{R}^d} f(x,\theta)[SS^\star]^{l_1l_2}(x)\mu(\mathrm{d}x)\right| \to^p 0.$$

From Lemmas 1 and 2-(i), as $h_n \to 0$ and $nh_n \to \infty$,

$$\begin{aligned}\sum_{k=1}^{n} E\left[\frac{1}{nh_n}f(X_{t_{k-1}^n},\theta)\bar{\Delta}_k^{l_1}\bar{\Delta}_k^{l_2}\,\Big|\,\mathcal{G}_{k-1}^n\right] &\to^p \int_{\mathbf{R}^d} f(x,\theta)[SS^\star]^{l_1l_2}(x)\mu(\mathrm{d}x), \\ \sum_{k=1}^{n} E\left[\left(\frac{1}{nh_n}f(X_{t_{k-1}^n},\theta)\bar{\Delta}_k^{l_1}\bar{\Delta}_k^{l_2}\right)^2\,\Big|\,\mathcal{G}_{k-1}^n\right] &\to^p 0.\end{aligned}$$

Lemma 9 of Genon-Catalot and Jacod [3] yields that for all θ,

$$\frac{1}{nh_n}\sum_{k=1}^{n} f(X_{t_{k-1}^n},\theta)\bar{\Delta}_k^{l_1}\bar{\Delta}_k^{l_2} \to^p \int_{\mathbf{R}^d} f(x,\theta)[SS^\star]^{l_1l_2}(x)\mu(\mathrm{d}x).$$

For tightness of the family of distributions of $\frac{1}{nh_n}\sum_{k=1}^n f(X_{t_{k-1}^n},\theta)\bar{\Delta}_k^{l_1}\bar{\Delta}_k^{l_2}$, one has

$$\sup_n E\left[\sup_{\theta\in\Theta}\left|\frac{1}{nh_n}\sum_{k=1}^n \frac{\partial}{\partial\theta} f(X_{t_{k-1}^n},\theta)\bar{\Delta}_k^{l_1}\bar{\Delta}_k^{l_2}\right|\right] \leq \sup_n E\left[\frac{1}{2nh_n}\sup_{\theta\in\Theta}\sum_{k=1}^n\left|\frac{\partial}{\partial\theta} f(X_{t_{k-1}^n},\theta)\right| E\left[(\bar{\Delta}_k^{l_1})^2+(\bar{\Delta}_k^{l_2})^2\Big|\mathcal{G}_{k-1}^n\right]\right] < \infty.$$

This completes the proof. □

Lemma 3. *Assume A1–A2. Then, as $h_n\to 0$ and $nh_n\to\infty$,*
(i)

$$\sup_{\theta\in\Theta}\left|\frac{1}{n}u_n(\alpha,\beta)-D_1(\beta)\right| \to^p 0,$$

(ii)

$$\sup_{\theta\in\Theta}\left|\frac{1}{nh_n}\{u_n(\alpha,\beta)-u_n(\alpha^*,\beta)\}-\{D_2(\alpha,\beta)-D_2(\alpha^*,\beta)\}\right| \to^p 0.$$

Proof. (i) Noting that

$$\frac{1}{n}u_n(\theta)=\frac{1}{2}\sum_{k=1}^n\left\{\frac{1}{n}\log\det\Xi(X_{t_{k-1}^n},\beta)+\frac{1}{nh_n}\sum_{l_1,l_2=1}^d(\Xi^{-1}(X_{t_{k-1}^n},\beta))^{l_1l_2}\Delta_k^{l_1}\Delta_k^{l_2}(\alpha)\right\},$$

one has the result by Lemmas 2-(i) and (iii).
(ii) By setting $\bar{b}_k(\alpha^*,\alpha)=b(X_{t_{k-1}^n},\alpha^*)-b(X_{t_{k-1}^n},\alpha)$,

$$\begin{aligned}&\frac{1}{nh_n}\{u_n(\alpha,\beta)-u_n(\alpha^*,\beta)\} \qquad (4.5)\\ =\ &\frac{1}{2nh_n}\sum_{k=1}^n\sum_{l_1,l_2=1}^d(\Xi^{-1}(X_{t_{k-1}^n},\beta))^{l_1l_2}\Big\{\bar{b}_k^{l_1}(\alpha^*,\alpha)\bar{\Delta}_k^{l_2}+\bar{b}_k^{l_2}(\alpha^*,\alpha)\bar{\Delta}_k^{l_1}\\ &+h_n(B^{l_1}(X_{t_{k-1}^n})-b^{l_1}(X_{t_{k-1}^n},\alpha))(B^{l_2}(X_{t_{k-1}^n})-b^{l_2}(X_{t_{k-1}^n},\alpha))\\ &-h_n(B^{l_1}(X_{t_{k-1}^n})-b^{l_1}(X_{t_{k-1}^n},\alpha^*))(B^{l_2}(X_{t_{k-1}^n})-b^{l_2}(X_{t_{k-1}^n},\alpha^*))\Big\}.\end{aligned}$$

It follows from Lemma 2-(i) and (4.1) that we have the result. This completes the proof. □

Proof of Proposition 2.1. . By *A*3-(*i*), we see that if $|\beta-\beta^*|\geq\epsilon$, then $D_1(\beta)>D_1(\beta^*)+\eta$ for some $\eta>0$. Thus, for any $\epsilon>0$,

$$\begin{aligned}P[|\hat{\beta}_n-\beta^*|\geq\epsilon] &\leq P\left[D_1(\hat{\beta}_n)>D_1(\beta^*)+\eta\right]\\ &\leq P\left[-\frac{1}{n}u_n(\hat{\alpha}_n,\hat{\beta}_n)+D_1(\hat{\beta}_n)>\frac{\eta}{3}\right]+P\left[\frac{1}{n}u_n(\hat{\alpha}_n,\hat{\beta}_n)-\frac{1}{n}u_n(\hat{\alpha}_n,\beta^*)>\frac{\eta}{3}\right]\\ &\quad +P\left[\frac{1}{n}u_n(\hat{\alpha}_n,\beta^*)-D_1(\beta^*)>\frac{\eta}{3}\right]\\ &\leq 2P\left[\sup_\theta\left|\frac{1}{n}u_n(\alpha,\beta)-D_1(\beta)\right|>\frac{\eta}{3}\right]+P\left[u_n(\hat{\alpha}_n,\hat{\beta}_n)>u_n(\hat{\alpha}_n,\beta^*)\right].\end{aligned}$$

It follows from Lemma 3-(i) and the definition of $\hat{\theta}_n$ that for any $\epsilon > 0$, $P[|\hat{\beta}_n - \beta^*| \geq \epsilon] \to 0$ as $h_n \to 0$ and $nh_n \to \infty$, which completes the proof of consistency of $\hat{\beta}_n$.

Next, we will show the consistency of $\hat{\alpha}_n$. Let $Y_n(\alpha,\beta) = u_n(\alpha,\beta) - u_n(\alpha^*,\beta)$ and $D(\alpha,\beta) = D_2(\alpha,\beta) - D_2(\alpha^*,\beta)$. Note that

$$\begin{aligned}\sup_\alpha \left|\frac{1}{nh_n}Y_n(\alpha,\hat{\beta}_n) - D(\alpha,\beta^*)\right| &\leq \sup_\alpha \left|\frac{1}{nh_n}\left\{Y_n(\alpha,\hat{\beta}_n) - Y_n(\alpha,\beta^*)\right\}\right| \\ &\quad + \sup_\alpha \left|\frac{1}{nh_n}Y_n(\alpha,\beta^*) - D(\alpha,\beta^*)\right|\end{aligned}$$

and it follows from (4.5), (4.1), Lemma 2-(i) and consistency of $\hat{\beta}_n$ that

$$\begin{aligned}\sup_\alpha \left|\frac{1}{nh_n}\left\{Y_n(\alpha,\hat{\beta}_n) - Y_n(\alpha,\beta^*)\right\}\right| &\leq \sup_\theta \left|\frac{1}{nh_n}\sum_{k=1}^n \sum_{l_1,l_2=1}^d (\Xi^{-1}(X_{t_{k-1}^n},\beta))^{l_1 l_2}\left\{\bar{b}_k^{l_1}(\alpha^*,\alpha)\bar{\Delta}_k^{l_2} + \bar{b}_k^{l_2}(\alpha^*,\alpha)\bar{\Delta}_k^{l_1}\right\}\right| \\ &\quad + |\hat{\beta}_n - \beta^*|\frac{1}{n}\sum_{k=1}^n \sum_{l_1,l_2=1}^d \sup_\beta \left|\delta_\beta(\Xi^{-1}(X_{t_{k-1}^n},\beta))^{l_1 l_2}\right| C(1+|X_{t_{k-1}^n}|)^C = o_p(1).\end{aligned}$$

Lemma 3-(ii) implies that as $h_n \to 0$ and $nh_n \to \infty$,

$$\sup_\alpha \left|\frac{1}{nh_n}Y_n(\alpha,\hat{\beta}_n) - D(\alpha,\beta^*)\right| \to^p 0. \tag{4.6}$$

Therefore, by the same argument as the proof of consistency of $\hat{\beta}_n$, for any $\epsilon > 0$, there exists a constant $\eta > 0$ such that

$$\begin{aligned}P[|\hat{\alpha}_n - \alpha^*| \geq \epsilon] &\leq P\left[D(\hat{\alpha}_n,\beta^*) > \eta\right] \\ &\leq P\left[-\frac{1}{nh_n}Y_n(\hat{\alpha}_n,\hat{\beta}_n) + D(\hat{\alpha}_n,\beta^*) > \frac{\eta}{2}\right] + P\left[\frac{1}{nh_n}Y_n(\hat{\alpha}_n,\hat{\beta}_n) > \frac{\eta}{2}\right] \\ &\leq P\left[\sup_\alpha\left|\frac{1}{nh_n}Y_n(\alpha,\hat{\beta}_n) - D(\alpha,\beta^*)\right| > \frac{\eta}{2}\right] + P\left[u_n(\hat{\alpha}_n,\hat{\beta}_n) > u_n(\alpha^*,\hat{\beta}_n)\right].\end{aligned}$$

From (4.6) and the definition of $\hat{\theta}_n$, one has that for any $\epsilon > 0$, $P[|\hat{\alpha}_n - \alpha^*| \geq \epsilon] \to 0$ as $h_n \to 0$ and $nh_n \to \infty$. This completes the proof. □

4.2. **Proofs of Theorems 2.1 and 2.2**

For the proofs of Theorems 2.1 and 2.2, we set that

$$C_n(\theta) = \begin{pmatrix} \left(\frac{1}{nh_n}(\delta_{\alpha_i}\delta_{\alpha_j}u_n)(\theta)\right)_{i,j=1,\dots,p} & \left(\frac{1}{nh_n}(\delta_{\alpha_i}\delta_{\beta_j}u_n)(\theta)\right)_{\substack{i=1,\dots,p\\ j=1,\dots,q}} \\ \left(\frac{1}{n}(\delta_{\alpha_j}\delta_{\beta_i}u_n)(\theta)\right)_{\substack{i=1,\dots,q\\ j=1,\dots,p}} & \left(\frac{1}{n}(\delta_{\beta_i}\delta_{\beta_j}u_n)(\theta)\right)_{i,j=1,\dots,q}\end{pmatrix},$$

$$\bar{C}_n(\theta) = \begin{pmatrix} \left(\frac{1}{nh_n}(\delta_{\alpha_i}\delta_{\alpha_j}u_n)(\theta)\right)_{i,j=1,\dots,p} & \left(\frac{1}{n\sqrt{h_n}}(\delta_{\alpha_i}\delta_{\beta_j}u_n)(\theta)\right)_{\substack{i=1,\dots,p\\ j=1,\dots,q}} \\ \left(\frac{1}{n\sqrt{h_n}}(\delta_{\alpha_j}\delta_{\beta_i}u_n)(\theta)\right)_{\substack{i=1,\dots,q\\ j=1,\dots,p}} & \left(\frac{1}{n}(\delta_{\beta_i}\delta_{\beta_j}u_n)(\theta)\right)_{i,j=1,\dots,q}\end{pmatrix}$$

and that

$$L_n = \begin{pmatrix} \left(-\frac{1}{\sqrt{nh_n}}(\delta_{\alpha_i} u_n)(\theta^*)\right)_{i=1,\dots,p} \\ \left(-\frac{\sqrt{h_n}}{\sqrt{n}}(\delta_{\beta_j} u_n)(\theta^*)\right)_{j=1,\dots,q} \end{pmatrix}, \quad \bar{L}_n = \begin{pmatrix} \left(-\frac{1}{\sqrt{nh_n}}(\delta_{\alpha_i} u_n)(\theta^*)\right)_{i=1,\dots,p} \\ \left(-\frac{1}{\sqrt{n}}(\delta_{\beta_j} u_n)(\theta^*)\right)_{j=1,\dots,q} \end{pmatrix}.$$

Lemma 4. *Assume A1–A2. Then, as $h_n \to 0$ and $nh_n \to \infty$,*
(i)

$$\sup_{\theta\in\Theta} |C_n(\theta) - J(\theta)| \to^p 0,$$

(ii)

$$\sup_{\theta\in\Theta} |\bar{C}_n(\theta) - \bar{J}(\theta)| \to^p 0.$$

Proof. An easy computation yields that

$$\begin{aligned}
\delta_{\alpha_i}\delta_{\alpha_j} u_n(\theta) &= -\sum_{k=1}^{n}\sum_{l_1=1}^{d}\Big\{\Big[(\delta_{\alpha_i}\delta_{\alpha_j} b)^\star(X_{t_{k-1}^n},\alpha)\Xi^{-1}(X_{t_{k-1}^n},\beta)\Big]^{l_1}\Delta_k^{l_1}(\alpha) \\
&\quad -h_n\Big[(\delta_{\alpha_j} b)^\star(X_{t_{k-1}^n},\alpha)\Xi^{-1}(X_{t_{k-1}^n},\beta)\Big]^{l_1}(\delta_{\alpha_i} b)^{l_1}(X_{t_{k-1}^n},\alpha)\Big\}, \\
\delta_{\alpha_i}\delta_{\beta_j} u_n(\theta) &= -\sum_{k=1}^{n}\sum_{l_1=1}^{d}\Big[(\delta_{\alpha_i} b)^\star(X_{t_{k-1}^n},\alpha)(\delta_{\beta_j}\Xi^{-1})(X_{t_{k-1}^n},\beta)\Big]^{l_1}\Delta_k^{l_1}(\alpha), \\
\delta_{\beta_i}\delta_{\beta_j} u_n(\theta) &= \frac{1}{2}\sum_{k=1}^{n}\Big\{\delta_{\beta_i}\delta_{\beta_j}\log\det\Xi(X_{t_{k-1}^n},\beta) \\
&\quad +h_n^{-1}\sum_{l_1,l_2=1}^{d}\Big(\delta_{\beta_i}\delta_{\beta_j}\Xi^{-1}(X_{t_{k-1}^n},\beta)\Big)^{l_1 l_2}\Delta_k^{l_1}(\alpha)\Delta_k^{l_2}(\alpha)\Big\}.
\end{aligned}$$

It follows from Lemma 2 that as $h_n \to 0$ and $nh_n \to \infty$, uniformly in θ,

$$\begin{aligned}
\frac{1}{nh_n}\delta_{\alpha_i}\delta_{\alpha_j} u_n(\theta) &\to^p -\int_{\mathbf{R}^d}\big\{(\delta_{\alpha_i}\delta_{\alpha_j} b)^\star(x,\alpha)\Xi^{-1}(x,\beta)(B(x)-b(x,\alpha)) \\
&\quad -(\delta_{\alpha_j} b)^\star(x,\alpha)\Xi^{-1}(x,\beta)(\delta_{\alpha_i} b)(x,\alpha)\big\}\,\mu(\mathrm{d}x), \\
\frac{1}{nh_n}\delta_{\alpha_i}\delta_{\beta_j} u_n(\theta) &\to^p -\int_{\mathbf{R}^d}(\delta_{\alpha_i} b)^\star(x,\alpha)(\delta_{\beta_j}\Xi^{-1})(x,\beta)(B(x)-b(x,\alpha))\mu(\mathrm{d}x), \\
\frac{1}{n\sqrt{h_n}}\delta_{\alpha_i}\delta_{\beta_j} u_n(\theta) &\to^p 0, \\
\frac{1}{n}\delta_{\beta_i}\delta_{\beta_j} u_n(\theta) &\to^p \frac{1}{2}\int_{\mathbf{R}^d}\big\{\delta_{\beta_i}\mathrm{tr}\left((\delta_{\beta_j}\Xi)\Xi^{-1}(x,\beta)\right)+\mathrm{tr}\left((\delta_{\beta_i}\delta_{\beta_j}\Xi^{-1})\Xi(x,\beta)\right) \\
&\quad +\mathrm{tr}\left((\delta_{\beta_i}\delta_{\beta_j}\Xi^{-1})(x,\beta)([SS^\star](x)-\Xi(x,\beta))\right)\big\}\,\mu(\mathrm{d}x),
\end{aligned}$$

which completes the proof. □

Lemma 5. *Assume A1–A2 and A4. Then, as $h_n \to 0$, $nh_n \to \infty$ and $nh_n^2 \to 0$,*

$$L_n \to^d N(0, K).$$

Proof. In order to obtain the result, it is sufficient to show that

$$-\frac{1}{\sqrt{nh_n}}\delta_{\alpha_i}u_n(\theta^*) = \frac{1}{\sqrt{T}}\int_0^T(\delta_{\alpha_i}b)^\star(X_t,\alpha^*)\Xi^{-1}(X_t,\beta^*)S(X_t)\mathrm{d}w_t - \frac{1}{\sqrt{T}}\int_0^T\partial_x G_{b,i}(X_t)S(X_t)\mathrm{d}w_t + o_p(1), \tag{4.7}$$

$$-\frac{\sqrt{h_n}}{\sqrt{n}}\delta_{\beta_j}u_n(\theta^*) = \frac{1}{\sqrt{T}}\int_0^T\partial_x G_{\sigma,j}(X_t)S(X_t)\mathrm{d}w_t + o_p(1). \tag{4.8}$$

For the proof of (4.7), setting $f_i(x) = (\delta_{\alpha_i}b)^\star(x,\alpha^*)\Xi^{-1}(x,\beta^*)$ and $g_{b,i}(x) = f_i(x)(B(x) - b(x,\alpha^*))$, one has

$$\begin{aligned}-\frac{1}{\sqrt{nh_n}}\delta_{\alpha_i}u_n(\theta^*) &= \frac{1}{\sqrt{nh_n}}\sum_{k=1}^n\int_{t_{k-1}^n}^{t_k^n}f_i(X_{t_{k-1}^n})S(X_t)\mathrm{d}w_t \\ &+\frac{1}{\sqrt{nh_n}}\sum_{k=1}^n\int_{t_{k-1}^n}^{t_k^n}f_i(X_{t_{k-1}^n})(B(X_t)-B(X_{t_{k-1}^n}))\mathrm{d}t \\ &+\frac{1}{\sqrt{nh_n}}\sum_{k=1}^n\int_{t_{k-1}^n}^{t_k^n}g_{b,i}(X_{t_{k-1}^n})\mathrm{d}t.\end{aligned}$$

Since it is easy to see that

$$\begin{aligned}&\frac{1}{\sqrt{nh_n}}\sum_{k=1}^n\int_{t_{k-1}^n}^{t_k^n}(f_i(X_t)-f_i(X_{t_{k-1}^n}))S(X_t)\mathrm{d}w_t = o_p(1),\\ &\frac{1}{\sqrt{nh_n}}\sum_{k=1}^n\int_{t_{k-1}^n}^{t_k^n}f_i(X_{t_{k-1}^n})(B(X_t)-B(X_{t_{k-1}^n}))\mathrm{d}t = o_p(1),\\ &\frac{1}{\sqrt{nh_n}}\sum_{k=1}^n\int_{t_{k-1}^n}^{t_k^n}(g_{b,i}(X_t)-g_{b,i}(X_{t_{k-1}^n}))\mathrm{d}t = o_p(1),\end{aligned}$$

we obtain that

$$-\frac{1}{\sqrt{nh_n}}\delta_{\alpha_i}u_n(\theta^*) = \frac{1}{\sqrt{T}}\int_0^T f_i(X_t)S(X_t)\mathrm{d}w_t + \frac{1}{\sqrt{nh_n}}\sum_{k=1}^n\int_{t_{k-1}^n}^{t_k^n}g_{b,i}(X_t)\mathrm{d}t + o_p(1).$$

Furthermore, by $A4$, $LG_{b,i}(x) = g_{b,i}(x)$. It follows from Ito's formula that

$$G_{b,i}(X_{t_k^n}) - G_{b,i}(X_{t_{k-1}^n}) = \int_{t_{k-1}}^{t_k}(\partial_x G_{b,i})(X_t)S(X_t)\mathrm{d}w_t + \int_{t_{k-1}}^{t_k}g_{b,i}(X_t)\mathrm{d}t.$$

Therefore, using $A4$, one has

$$\frac{1}{\sqrt{nh_n}}\sum_{k=1}^n\int_{t_{k-1}}^{t_k}g_{b,i}(X_t)\mathrm{d}t = -\frac{1}{\sqrt{T}}\int_0^T(\partial_x G_{b,i})(X_t)S(X_t)\mathrm{d}w_t + o_p(1).$$

This completes the proof of (4.7). □

Next we will prove (4.8). Let

$$M_{j,k}(\theta) = \frac{1}{2}\mathrm{tr}\left\{\left(\delta_{\beta_j}\Xi^{-1}(X_{t^n_{k-1}},\beta)\right)\left(\Delta_k(\alpha)\Delta_k^\star(\alpha) - h_n[SS^\star](X_{t^n_{k-1}})\right)\right\}.$$

Note that

$$\delta_{\beta_j}u_n(\theta^*) = \frac{1}{h_n}\sum_{k=1}^n M_{j,k}(\theta^*) + \sum_{k=1}^n g_{\sigma,j}(X_{t^n_{k-1}}). \tag{4.9}$$

Since it follows from Lemmas 1 and 2 that

$$\begin{aligned}
\frac{1}{\sqrt{nh_n}}\sum_{k=1}^n E\left[M_{j,k}(\theta^*)|\mathcal{G}^n_{k-1}\right] &= \frac{1}{\sqrt{nh_n}}\sum_{k=1}^n R(\theta, h_n^2, X_{t^n_{k-1}}) \to^p 0,\\
\frac{1}{nh_n}\sum_{k=1}^n E\left[M_{j,k}(\theta^*)^2|\mathcal{G}^n_{k-1}\right] &= \frac{1}{nh_n}\sum_{k=1}^n R(\theta, h_n^2, X_{t^n_{k-1}}) \to^p 0,
\end{aligned}$$

Lemma 9 in Genon-Catalot and Jacod [3] implies that $\frac{1}{\sqrt{nh_n}}\sum_{k=1}^n M_{j,k}(\theta^*) = o_p(1)$. An easy estimate yields that $\frac{1}{\sqrt{nh_n}}\sum_{k=1}^n \int_{t_{k-1}}^{t_k}\left(g_{\sigma,j}(X_t) - g_{\sigma,j}(X_{t^n_{k-1}})\right)\mathrm{d}t = o_p(1)$. Therefore,

$$-\frac{\sqrt{h_n}}{\sqrt{n}}\delta_{\beta_j}u_n(\theta^*) = -\frac{1}{\sqrt{nh_n}}\sum_{k=1}^n\int_{t_{k-1}}^{t_k} g_{\sigma,j}(X_t)\mathrm{d}t + o_p(1).$$

Moreover, it follows from *A4* that $LG_{\sigma,j}(x) = g_{\sigma,j}(x)$. By using Ito's formula,

$$G_{\sigma,j}(X_{t^n_k}) - G_{\sigma,j}(X_{t^n_{k-1}}) = \int_{t_{k-1}}^{t_k}(\partial_x G_{\sigma,j})(X_t)S(X_t)\mathrm{d}w_t + \int_{t_{k-1}}^{t_k} g_{\sigma,j}(X_t)\mathrm{d}t.$$

Thus, *A4* implies that

$$\frac{1}{\sqrt{nh_n}}\sum_{k=1}^n\int_{t_{k-1}}^{t_k} g_{\sigma,j}(X_t)\mathrm{d}t = -\frac{1}{\sqrt{T}}\int_0^T(\partial_x G_{\sigma,j})(X_t)S(X_t)\mathrm{d}w_t + o_p(1).$$

This completes the proof of (4.8). The central limit theorem for martingale yields the result. This completes the proof.

Lemma 6. *Assume A1–A2 and A4′. Then, as* $h_n \to 0$, $nh_n \to \infty$ *and* $nh_n^2 \to 0$,

$$\bar{L}_n \to^d N(0, \bar{K}).$$

Proof. Let

$$\begin{aligned}
\xi_{i,k} &= \frac{1}{\sqrt{nh_n}}\sum_{l=1}^d\left((\delta_{\alpha_i}b)^\star(X_{t^n_{k-1}},\alpha^*)\Xi^{-1}(X_{t^n_{k-1}},\beta^*) - \partial_x G_{b,i}(X_{t^n_{k-1}})\right)^l\bar{\Delta}^l_k,\\
\eta_{j,k} &= -\frac{1}{2\sqrt{nh_n}}\sum_{l_1,l_2=1}^d\left(\delta_{\beta_j}\Xi^{-1}(X_{t^n_{k-1}},\beta^*)\right)^{l_1l_2}\left(\bar{\Delta}^{l_1}_k\bar{\Delta}^{l_2}_k - h_n[SS^\star]^{l_1l_2}(X_{t^n_{k-1}})\right).
\end{aligned}$$

We first show that

$$-\frac{1}{\sqrt{nh_n}}\delta_{\alpha_i}u_n(\theta^*) = \sum_{k=1}^n \xi_{i,k} + o_p(1), \tag{4.10}$$

$$-\frac{1}{\sqrt{n}}\delta_{\beta_j}u_n(\theta^*) = \sum_{k=1}^n \eta_{j,k} + o_p(1). \tag{4.11}$$

For the proof of (4.10), we note that

$$-\delta_{\alpha_i}u_n(\theta^*) = \sum_{k=1}^n\sum_{l=1}^d \left[(\delta_{\alpha_i}b)^\star(X_{t^n_{k-1}},\alpha^*)\Xi^{-1}(X_{t^n_{k-1}},\beta^*)\right]^l \bar{\Delta}^l_k + h_n\sum_{k=1}^n g_{b,i}(X_{t^n_{k-1}}).$$

Hence,

$$\begin{aligned}
-\frac{1}{\sqrt{nh_n}}\delta_{\alpha_i}u_n(\theta^*) - \sum_{k=1}^n \xi_{i,k} &= \frac{h_n}{\sqrt{nh_n}}\sum_{k=1}^n g_{b,i}(X_{t^n_{k-1}}) + \frac{1}{\sqrt{nh_n}}\sum_{k=1}^n\sum_{l=1}^d\left(\partial_x G_{b,i}(X_{t^n_{k-1}})\right)^l\bar{\Delta}^l_k \\
&= -\frac{1}{\sqrt{nh_n}}\sum_{k=1}^n\int_{t^n_{k-1}}^{t^n_k}(g_{b,i}(X_t)-g_{b,i}(X_{t^n_{k-1}}))\mathrm{d}t \\
&\quad +\frac{1}{\sqrt{nh_n}}\sum_{k=1}^n\left(\int_{t^n_{k-1}}^{t^n_k} g_{b,i}(X_t)\mathrm{d}t + \int_{t^n_{k-1}}^{t^n_k}\partial_x G_{b,i}(X_{t^n_{k-1}})S(X_t)\mathrm{d}w_t\right) \\
&\quad +\frac{1}{\sqrt{nh_n}}\sum_{k=1}^n\int_{t^n_{k-1}}^{t^n_k}\partial_x G_{b,i}(X_{t^n_{k-1}})(B(X_t)-B(X_{t^n_{k-1}}))\mathrm{d}t.
\end{aligned}$$

It follows from $A4'$ and Ito's formula that

$$-\frac{1}{\sqrt{nh_n}}\delta_{\alpha_i}u_n(\theta^*) - \sum_{k=1}^n \xi_{i,k} = \frac{1}{\sqrt{nh_n}}\sum_{k=1}^n A_{i,k} + o_p(1),$$

where $A_{i,k} = \int_{t^n_{k-1}}^{t^n_k}\left\{\partial_x G_{b,i}(X_{t^n_{k-1}})(B(X_t)-B(X_{t^n_{k-1}})) - (g_{b,i}(X_t)-g_{b,i}(X_{t^n_{k-1}}))\right\}\mathrm{d}t$. Moreover, one has that

$$\begin{aligned}
&\frac{1}{\sqrt{nh_n}}\sum_{k=1}^n E[A_k|\mathcal{G}^n_{k-1}] = \frac{1}{\sqrt{nh_n}}\sum_{k=1}^n R(\theta,h_n^{3/2},X_{t^n_{k-1}}) \to^p 0, \\
&\frac{1}{nh_n}\sum_{k=1}^n E[A_k^2|\mathcal{G}^n_{k-1}] \le \frac{C}{n}\sum_{k=1}^n\int_{t^n_{k-1}}^{t^n_k} E\left[\left(\partial_x G_{b,i}(X_{t^n_{k-1}})(B(X_t)-B(X_{t^n_{k-1}}))\right)^2\right. \\
&\qquad\qquad \left.+(g_i(X_t)-g_i(X_{t^n_{k-1}}))^2\right|\mathcal{G}^n_{k-1}\bigg]\mathrm{d}t \le \frac{C}{n}\sum_{k=1}^n R(\theta,h_n^2,X_{t^n_{k-1}}) \to^p 0.
\end{aligned}$$

By Lemma 9 in Genon-Catalot and Jacod [3], one has that $\frac{1}{\sqrt{nh_n}}\sum_{k=1}^n A_{i,k}\to 0$, which completes the proof of (4.10).

Next we will prove (4.11). An easy calculation together with (4.4) and (4.9) yields that

$$-\frac{1}{\sqrt{n}}\delta_{\beta_j}u_n(\theta^*)-\sum_{k=1}^n\eta_{j,k}=-\frac{1}{\sqrt{n}}\sum_{k=1}^n g_{\sigma,j}(X_{t^n_{k-1}})-\frac{1}{2\sqrt{n}h_n}\sum_{k=1}^n\sum_{l_1,l_2=1}^d\Big(\delta_{\beta_j}\Xi^{-1}(X_{t^n_{k-1}},\beta^*)\Big)^{l_1l_2}$$
$$\times\Big[h_n\bar{\Delta}^{l_1}_k(B(X_{t^n_{k-1}})-b(X_{t^n_{k-1}},\alpha^*))^{l_2}$$
$$+h_n\bar{\Delta}^{l_2}_k(B(X_{t^n_{k-1}})-b(X_{t^n_{k-1}},\alpha^*))^{l_1}++h_n^2(B(X_{t^n_{k-1}})$$
$$-b(X_{t^n_{k-1}},\alpha^*))^{l_1}(B(X_{t^n_{k-1}})-b(X_{t^n_{k-1}},\alpha^*))^{l_2}\Big].$$

By noting that $g_{\sigma,j}(x)=0$, it follows from Lemma 2 and (4.1) that $-\frac{1}{\sqrt{n}}\delta_{\beta_j}u_n(\bar{\theta}^*)-\sum_{k=1}^n\eta_{j,k}\to^p 0$. This completes the proof of (4.11).

Finally, we show that

$$\sum_{k=1}^n E[\xi_{i,k}|\mathcal{G}^n_{k-1}]\to^p 0,\quad \sum_{k=1}^n E[\eta_{j,k}|\mathcal{G}^n_{k-1}]\to^p 0,\quad \sum_{k=1}^n E[\xi_{i,k}\eta_{j,k}|\mathcal{G}^n_{k-1}]\to^p 0, \tag{4.12}$$
$$\sum_{k=1}^n |E[\xi_{i,k}|\mathcal{G}^n_{k-1}]|^2\to^p 0,\quad \sum_{k=1}^n |E[\eta_{j,k}|\mathcal{G}^n_{k-1}]|^2\to^p 0, \tag{4.13}$$
$$\sum_{k=1}^n E[\xi_{i,k}\xi_{j,k}|\mathcal{G}^n_{k-1}]\to^p K^{ij}_{bb},\quad \sum_{k=1}^n E[\eta_{i,k}\eta_{j,k}|\mathcal{G}^n_{k-1}]\to^p \bar{K}^{ij}_{\sigma\sigma}, \tag{4.14}$$
$$\sum_{k=1}^n E[\xi^4_{i,k}|\mathcal{G}^n_{k-1}]\to^p 0,\quad \sum_{k=1}^n E[\eta^4_{j,k}|\mathcal{G}^n_{k-1}]\to^p 0. \tag{4.15}$$

Lemmas 1 and 2 yield that

$$\sum_{k=1}^n E[\xi_{i,k}|\mathcal{G}^n_{k-1}] = \frac{1}{\sqrt{nh_n}}\sum_{k=1}^n R(\theta,h_n^2,X_{t^n_{k-1}})\to^p 0,$$
$$\sum_{k=1}^n E[\eta_{j,k}|\mathcal{G}^n_{k-1}] = \frac{1}{2\sqrt{n}h_n}\sum_{k=1}^n R(\theta,h_n^2,X_{t^n_{k-1}})\to^p 0,$$
$$\sum_{k=1}^n E[\xi_{i,k}\eta_{j,k}|\mathcal{G}^n_{k-1}] = \frac{1}{2nh_n^{3/2}}\sum_{k=1}^n R(\theta,h_n^2,X_{t^n_{k-1}})\to^p 0,$$

which completes the proof of (4.12). In the same way,

$$\sum_{k=1}^n |E[\xi_{i,k}|\mathcal{G}^n_{k-1}]|^2 = \frac{1}{nh_n}\sum_{k=1}^n R(\theta,h_n^4,X_{t^n_{k-1}})\to^p 0,$$
$$\sum_{k=1}^n |E[\eta_{j,k}|\mathcal{G}^n_{k-1}]|^2 = \frac{1}{4nh_n^2}\sum_{k=1}^n R(\theta,h_n^4,X_{t^n_{k-1}})\to^p 0.$$

This completes the proof of (4.13). Next, setting $f_{i,k} = (\partial_{\alpha_i} b)^\star(X_{t_k^n}, \alpha^*)\Xi^{-1}(X_{t_k^n}, \beta^*) - \partial_x G_{b,i}(X_{t_k^n})$ and $N_{j,k} = \delta_{\beta_j}\Xi^{-1}(X_{t_k^n}, \beta^*)$, one has that

$$\begin{aligned}
\sum_{k=1}^n E[\xi_{i,k}\xi_{j,k}|\mathcal{G}_{k-1}^n] &= \frac{1}{n}\sum_{k=1}^n \sum_{l_1,l_2=1}^d f_{i,k-1}^{l_1} f_{i,k-1}^{l_2}[SS^\star]^{l_1 l_2}(X_{t_{k-1}^n}) + \frac{1}{n}\sum_{k=1}^n R(\theta, h_n, X_{t_{k-1}^n}) \\
&\to^p K_{bb}^{ij}, \\
\sum_{k=1}^n E[\eta_{i,k}\eta_{j,k}|\mathcal{G}_{k-1}^n] &= \frac{1}{4n}\sum_{k=1}^n \sum_{l_1,l_2=1}^d \sum_{l_3,l_4=1}^d N_{i,k-1}^{l_1 l_2} N_{j,k-1}^{l_3 l_4} \\
&\quad \times([SS^\star]^{l_1 l_3}(X_{t_{k-1}^n})[SS^\star]^{l_2 l_4}(X_{t_{k-1}^n}) + [SS^\star]^{l_1 l_4}(X_{t_{k-1}^n})[SS^\star]^{l_2 l_3}(X_{t_{k-1}^n})) \\
&\quad + \frac{1}{nh_n^2}\sum_{k=1}^n R(\theta, h_n^{5/2}, X_{t_{k-1}^n}) \\
&\to^p \bar{K}_{\sigma\sigma}^{ij}.
\end{aligned}$$

This completes the proof of (4.14). Furthermore,

$$\begin{aligned}
\sum_{k=1}^n E[\xi_{i,k}^4|\mathcal{G}_{k-1}^n] &\le \frac{C}{(nh_n)^2}\sum_{k=1}^n \sum_{l=1}^d (f_{i,k-1}^l)^4 E[(\bar{\Delta}_k^l)^4|\mathcal{G}_{k-1}^n] \\
&= \frac{1}{n^2}\sum_{k=1}^n R(\theta, 1, X_{t_{k-1}^n}) \to^p 0, \\
\sum_{k=1}^n E[\eta_{i,k}^4|\mathcal{G}_{k-1}^n] &\le \frac{C}{(nh_n^2)^2}\sum_{k=1}^n \sum_{l_1,l_2=1}^d (N_{i,k-1}^{l_1 l_2})^4 E[(\bar{\Delta}_k^{l_1}\bar{\Delta}_k^{l_2})^4 + h_n^4(\Xi^{l_1 l_2}(X_{t_{k-1}^n}))^4|\mathcal{G}_{k-1}^n] \\
&= \frac{1}{(nh_n^2)^2}\sum_{k=1}^n R(\theta, h_n^4, X_{t_{k-1}^n}) \to^p 0,
\end{aligned}$$

where we used the following estimate:

$$E[(\bar{\Delta}_k^l)^8|\mathcal{G}_{k-1}^n] \le CE[|X_{t_k^n} - X_{t_{k-1}^n}|^8|\mathcal{G}_{k-1}^n] + R(\theta, h_n^8, X_{t_{k-1}^n}) \le R(\theta, h_n^4, X_{t_{k-1}^n})$$

and this completes the proof of (4.15). By using a combination of Theorems 3.2 and 3.4 in Hall and Heyde [5], we obtain the asymptotic normality. This completes the proof. □

Proof of Theorem 2.1. Let $B(\theta^*;\rho) = \{\theta : |\theta - \theta^*| \le \rho\}$. Since $\theta^* \in \mathrm{Int}(\Theta)$, one has that $B(\theta^*;\rho) \subset \mathrm{Int}(\Theta)$ for sufficiently small $\rho > 0$. It follows from the Taylor expansion that

$$\int_0^1 C_n(\theta^* + t(\hat{\theta}_n - \theta^*))\mathrm{d}t \begin{pmatrix} \sqrt{nh_n}(\hat{\alpha}_n - \alpha^*) \\ \sqrt{nh_n}(\hat{\beta}_n - \beta^*) \end{pmatrix} 1_{\{\hat{\theta}_n \in B(\theta^*;\rho)\}} = L_n 1_{\{\hat{\theta}_n \in B(\theta^*;\rho)\}}.$$

The consistency of $\hat{\theta}_n$ yields that for sufficiently small $\rho > 0$, $1_{\{\hat{\theta}_n \in B(\theta^*;\rho)\}} \to^p 1$ as $h_n \to 0$ and $nh_n \to \infty$. Lemma 4-(i) and the continuity of $J(\theta)$ imply that as $h_n \to 0$ and $nh_n \to \infty$,

$$\begin{aligned}
C_n(\theta^*) &\to^p J(\theta^*), \\
\sup_{|\theta| \le \epsilon_n} |C_n(\theta^* + \theta) - C_n(\theta^*)| &\to^p 0
\end{aligned}$$

for any sequence ϵ_n of positive numbers tending to zero. By Lemma 5 together with the above estimates, it is easy to show the result. This completes the proof. □

Proof of Theorem 2.2. In the same way as in the proof of Theorem 2.1,

$$\int_0^1 \bar{C}_n(\theta^* + t(\hat{\theta}_n - \theta^*))\mathrm{d}t \begin{pmatrix} \sqrt{nh_n}(\hat{\alpha}_n - \alpha^*) \\ \sqrt{n}(\hat{\beta}_n - \beta^*) \end{pmatrix} 1_{\{\hat{\theta}_n \in B(\theta^*;\rho)\}} = \bar{L}_n 1_{\{\hat{\theta}_n \in B(\theta^*;\rho)\}}.$$

By using the analogous argument with the the proof of Theorem 2.1, it follows from the consistency of $\hat{\theta}_n$, Lemma 4-(ii) and Lemma 6 that we obtain the result. This completes the proof. □

5. Conclusion and discussion

This paper treated the parametric estimation for two kinds of misspecified ergodic diffusion models: the completely misspecified case in Section 2.2 and the semi-misspecified case in Section 2.3. For the estimation of the semi-misspecified case based on the continuously observed data $X = \{X_t; t \in [0,T]\}$, under some regularity conditions with the assumption that $\Xi(x,\beta^*) = [SS^\star](x)$ for all x, the maximum likelihood estimator $\hat{\alpha}_T$ has the asymptotic normality

$$\sqrt{T}(\hat{\alpha}_T - \alpha^*) \to^d N(0, J_{bb}^{-1}(\theta^*)K_{bb}J_{bb}^{-1}(\theta^*)) \tag{5.1}$$

as $T \to \infty$, see McKeague [11], Yoshida [17] and Kutoyants [9]. Meanwhile, in the case of parameter estimation with discrete observations for the correctly specified parametric case where there exists a true parameter $\theta^* = (\alpha^*,\beta^*) \in \Theta_\alpha \times \Theta_\beta$ such that $b(x,\alpha^*) = B(x)$ and $[\sigma\sigma^\star](x,\beta^*) = [SS^\star](x)$ for all x, under some regularity conditions, the minimum contrast estimator defined by (2.1) is asymptotically efficient as follows:

$$\begin{pmatrix} \sqrt{nh_n}(\hat{\alpha}_n - \alpha^*) \\ \sqrt{n}(\hat{\beta}_n - \beta^*) \end{pmatrix} \to^d N(0, \bar{J}^{-1}(\theta^*)), \tag{5.2}$$

see Yoshida [18,19] and Kessler [7]. Here we note that in the correctly specified parametric case, $\bar{J}(\theta^*)$ is the asymptotic Fisher information matrix, see Gobet [4]. By Theorem 2.2, we see that the minimum contrast estimator for the semi-misspecified case has the same rate of convergence as the correctly specified parametric case. If we take (5.1) and (5.2) into account, the rate of convergence in Theorem 2.2 seems natural.

On the other hand, as we have seen from Theorem 2.1, the rate of convergence for the estimator $\hat{\beta}_n$ of the diffusion coefficient parameter β in the completely misspecified case is different from the one given in (5.2) in the correctly specified parametric case. This fact results from the difference between the rates of convergence for $\delta_\beta u_n(\theta^*)$, see (4.8) in the proof of Lemma 5 and (4.11) in the proof of Lemma 6. Here we note that in the completely misspecified case,

$$\delta_{\beta_j} u_n(\theta^*) = \frac{1}{h_n}\sum_{k=1}^n M_{j,k}(\theta^*) + \sum_{k=1}^n g_{\sigma,j}(X_{t_{k-1}^n}),$$

while in the semi-misspecified case,

$$\delta_{\beta_j} u_n(\theta^*) = \frac{1}{h_n}\sum_{k=1}^n M_{j,k}(\theta^*)$$

since $g_{\sigma,j}(x) = 0$. If $\delta_\beta u_n(\theta^*)$ had the same rate of convergence in both cases, then the rate of convergence for $\hat{\beta}_n$ in the completely misspecified case could be the same as the one in the correctly specified parametric case. However, checking the proof of (4.8) carefully, we see that $\frac{1}{\sqrt{n}}\delta_{\beta_j} u_n(\theta^*)$ can diverge as $n \to \infty$, which follows from the estimates that $\frac{1}{\sqrt{nh_n}}\sum_{k=1}^n M_{j,k}(\theta^*) = O_p(1)$ but $\frac{1}{\sqrt{n}}\sum_{k=1}^n g_{\sigma,j}(X_{t_{k-1}^n})$ can diverge as $n \to \infty$.

Because of it, the rate of convergence for $\hat{\beta}_n$ in the completely misspecified case is worse than the one in the correctly specified parametric case. Therefore, it does not seem that the asymptotic result of the estimator in the completely misspecified case (Theorem 1) is trivial, in particular, for the diffusion coefficient estimator $\hat{\beta}_n$. Furthermore, the difference of the rates of convergence for $\hat{\beta}_n$ between the completely misspecified model and the semi-misspecified model might be available to test whether a diffusion coefficient is completely misspecified. Our results suggest that for example we should be careful when testing a hypothesis on a volatility parameter because the null hypothesis will be rejected in completely misspecified case if we take a critical region based on $\sqrt{n}(\hat{\beta}_n - \beta^*)$.

Acknowledgements. The authors wish to thank the referee and the associate editor for their valuable comments. Masayuki Uchida's research was partially supported by JST, CREST, Grant-in-Aid for Scientific Research from the JSPS, and by Cooperative Research Program of the Institute of Statistical Mathematics. Nakahiro Yoshida's research was in part supported by JST Basic Research Programs PRESTO, Grants-in-Aid for Scientific Research No. 19340021, the Global COE program "The research and training center for new development in mathematics" of Graduate School of Mathematical Sciences, University of Tokyo, and by Cooperative Research Program of the Institute of Statistical Mathematics.

References

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

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

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

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

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

[6] I.A. Ibragimov and R.Z. Has'minskii, *Statistical estimation.* Springer Verlag, New York (1981).

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

[8] S. Kusuoka and N. Yoshida, Malliavin calculus, geometric mixing, and expansion of diffusion functionals, *Probab. Theory Relat. Fields* **116** (2000) 457–484.

[9] Yu.A. Kutoyants, *Statistical inference for ergodic diffusion processes.* Springer-Verlag, London (2004).

[10] H. Masuda, Ergodicity and exponential β-mixing bound for multidimensional diffusions with jumps. *Stochastic Processes Appl.* **117** (2007) 35–56.

[11] I.W. McKeague, Estimation for diffusion processes under misspecified models. *J. Appl. Probab.* **21** (1984) 511–520.

[12] S.P. Meyn and P.L. Tweedie, Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.* **25** (1993) 518–548.

[13] E. Pardoux and A.Y. Veretennikov, On the Poisson equation and diffusion approximation 1. *Ann. Prob.* **29** (2001) 1061–1085.

[14] B.L.S. Prakasa Rao, Asymptotic theory for nonlinear least squares estimator for diffusion processes. *Math. Operationsforsch. Statist. Ser. Statist.* **14** (1983) 195–209.

[15] B.L.S. Prakasa Rao, Statistical inference from sampled data for stochastic processes. *Contemp. Math.* **80** (1988) 249–284. Amer. Math. Soc., Providence, RI.

[16] M. Uchida and N. Yoshida, Information criteria in model selection for mixing processes. *Statist. Infer. Stochast. Process.* **4** (2001) 73–98.

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

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

[19] N. Yoshida, Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations (to appear in *Ann. Inst. Statist. Math.*) (2005).