Statistical Inference for Stochastic Processes **7:** 35–67, 2004.

Information Criteria for Small Diffusions *via* the Theory of Malliavin–Watanabe

MASAYUKI UCHIDA[1,*] and NAKAHIRO YOSHIDA[2]
[1]*Faculty of Mathematics, Kyushu University, Ropponmatsu, Fukuoka 810-8560, Japan*
[2]*Graduate School of Mathematical Sciences, University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan*

Abstract. Information criteria based on the expected Kullback–Leibler information are presented by means of the asymptotic expansions derived with the Malliavin calculus. We consider the evaluation problem of statistical models for diffusion processes with small noise. The correction terms are essentially different from the ones for ergodic diffusion models presented in Uchida and Yoshida [34, 35].

Mathematics Subject Classifications (1991): 62B10, 62E20, 62M05, 60H07.

Key words: information criterion, asymptotic expansion, median unbiasedness, diffusion process with small noise, Malliavin calculus, M-estimators.

1. Introduction

Akaike's information criterion AIC [2, 3] is a model evaluation–selection tool obtained by estimating the expected Kullback–Leibler information of the fitted model with respect to the true model. It can be derived under the assumptions: (i) the data are independent random samples from an unknown distribution, (ii) estimation is done by the maximum likelihood method, and (iii) the parametric family of distributions includes the true model. Takeuchi [33] derived Takeuchi's information criterion TIC from the assumptions (i) and (ii), relaxing the assumption (iii) to the case where the model class may be misspecified. Konishi and Kitagawa [13] recently proposed a generalized information criterion GIC under the assumption (i) and for functional-type estimators instead of the assumption (ii).

In this paper, based on the Kullback–Leibler divergence, we consider information criteria, which work for (i′) diffusion processes with small noise, (ii′) various estimators including M-estimators, and (iii′) generally misspecified cases. For details of diffusion models with small noise, see [16, 18]. There are many applications of diffusion models with small noise to finance, see [12, 14, 31, 36, 41] and references therein. In order to treat sampled data, it is the first step for statisticians to consider continuous-time stochastic models, as it was the case for ergodic

*Author for correspondence: Tel.: +81-92-726-4785; Fax: +81-92-726-4785; e-mail: uchida@math.kyushu-u.ac.jp

diffusions. It was after inference for continuous-time was established when several authors studied sampled data as a modification: see [4, 7–9, 11, 25, 26, 42] for detailed history of the inference in continuous-time and discrete-time observation settings. Genon-Catalot [6] and Laredo [19] treated discretely observed diffusion processes with small noise, and recently, Sørensen [30] investigated small diffusion asymptotics for martingale estimating functions from discrete observations.

For the derivation of information criteria, the emphasis is put on the use of the asymptotic expansion of an estimator of the Kullback–Leibler divergence. From asymptotic expansion's point of view, the determination of the existing information criteria AIC, TIC, GIC and others is nothing but to find the second-order correction terms so that the estimator of the Kullback–Leibler divergence becomes expectation unbiased up to the second order in each setting. That is, in each case, the validity of the choice of the correction term has been explained by this expectation unbiasedness. On the other hand, from the decision theoretic point of view, those choices are based on the quadratic loss, and it may be a natural question how the correction procedure should change for other loss functions. Besides the traditional expectation unbiased corrections, in order to show another possibility, we will present the median unbiased correction, as it corresponds to the absolute loss, and hence the median unbiased information criterion (MUIC). Though the newly presented MUIC is just an example of information criteria not based on expectation unbiasedness, it should be noted that all those corrections are derived in a unified way from the asymptotic expansion of the estimator of the divergence. In fact, as it was shown in [34, 35], it is possible to construct f-unbiased information criteria, that is, the f-bias $E[f(\text{discrepancy})] = 0$ up to the second order, where each f is a measurable function, for example, $f(x) = x$, $1_{(-\infty,0]}(x) - 1/2$, etc.

In order to derive asymptotic expansions for diffusion processes with small noise, we will use the Malliavin calculus. Some fundamentals are summarized in Section 4. For more details of the Malliavin calculus, see [10, 21–24, 27, 28, 38, 40]. Watanabe [37] presented the concept of the generalized Wiener functional (i.e. the Schwartz distribution on the probability space), the pull-back of Schwartz distribution under Wiener mappings, and in his renowned work [39] he formulated the asymptotic expansion of the generalized Wiener functionals in some Sobolev space. To use this theory, the crucial step is to show the nondegeneracy of the Malliavin covariance of functionals. However, it is not easy to check this even for a simple statistical estimator, whose Malliavin covariance is given by an integration of some nonadaptive process. In addition, as for estimators such as maximum likelihood estimators, we cannot ensure their existence on the whole sample space in general. This difficulties has been solved by Takanobu and Watanabe [32] and Yoshida [14] in the modification of Watanabe's theory with truncation. For more details of asymptotic expansions for diffusion processes with small noise, see [5, 43, 45, 47].

In the previous papers, Uchida and Yoshida [34, 35] obtained two information criteria which work for mixing processes including ergodic diffusion processes as application, and for (ii′) and (iii′). It took advantage of the valid asymptotic expansion of the distribution of the estimators of the divergence for ϵ-Markov, geometric mixing stochastic processes with continuous-time parameter (cf. [15, 29]). Both the present paper and the previous ones are based on the Malliavin calculus. However, the asymptotics are utterly different; in fact, the former uses the expansion of generalized Wiener functionals as mentioned above and the latter used the method called 'local approach' [15]. This difference also reflects the different forms of correction terms: there is no full correspondence between the correction terms in the present paper and the previous ones in [34, 35].

The organization of the paper is as follows. In Section 2, information criteria with general estimators are heuristically derived, but for rigorous statements we will be later given in Section 5. In Section 3, as a special case, the information criteria which work for M-estimators are obtained. In Section 4, we review the fundamental results of Malliavin calculus. In Section 5, we state our main results. By using the asymptotic expansion of the distribution of the estimators for diffusion processes with small noise, a general theory is developed. Therefore, the results of previous sections are special cases of the ones in Section 5. Proofs of these results are given in Section 6.

2. Asymptotic Expansion and Information Criteria

Let $X = \{X_t; t \in [0, T]\}$ be a d-dimensional diffusion process defined by the stochastic differential equation (true model)

$$\begin{aligned} \mathrm{d}X_t &= V_0(X_t)\,\mathrm{d}t + \varepsilon V(X_t)\,\mathrm{d}w_t, \quad t \in [0, T],\ \varepsilon \in (0, 1], \\ X_0 &= x_0, \end{aligned} \tag{1}$$

where T is a fixed value, x_0 is a constant, $V = (V_1, \ldots, V_r)$ is an $\mathbf{R}^d \otimes \mathbf{R}^r$-valued smooth function defined on $\mathbf{R}^d$, V_0 is an $\mathbf{R}^d$-valued smooth function defined on $\mathbf{R}^d$ with bounded x-derivative and w is an r-dimensional standard Wiener process.

Consider d-dimensional diffusion process defined by the stochastic differential equation (statistical model)

$$\begin{aligned} \mathrm{d}X_t &= \tilde{V}_0(X_t, \theta)\,\mathrm{d}t + \varepsilon V(X_t)\,\mathrm{d}w_t, \quad t \in (0, T],\ \varepsilon \in [0, 1], \\ X_0 &= x_0, \end{aligned} \tag{2}$$

where a p-dimensional unknown parameter $\theta \in \Theta$: a bounded convex domain of $\mathbf{R}^d$, $\tilde{V}_0$ is an $\mathbf{R}^d$-valued smooth function defined on $\mathbf{R}^d \times \Theta$, T, x_0, V and w are the same as (1).

Let $X^{\varepsilon,\theta}$ be the solution of the stochastic differential equation (2) for θ. Let $P_{\varepsilon,\theta}$ be the induced measure from P on $C([0, T]; \mathbf{R}^d)$ by the mapping $w \to X^{\varepsilon,\theta}(w)$. The Radon–Nikodym derivative of $P_{\varepsilon,\theta}$ with respect to P_{ε,θ_0} is given by the formula (e.g. [20])

$$\Lambda_\varepsilon(\theta; X)\Lambda_\varepsilon(\theta_0; X)^{-1},$$

where

$$\Lambda_\varepsilon(\theta; X) = \exp\left\{\int_0^T \varepsilon^{-2}\tilde{V}_0'(VV')^+(X_t,\theta)\,\mathrm{d}X_t - \right.$$
$$\left. -\frac{1}{2}\int_0^T \varepsilon^{-2}\tilde{V}_0'(VV')^+\tilde{V}_0(X_t,\theta)\,\mathrm{d}t\right\}. \tag{3}$$

Here A^+ denotes the Moore–Penrose generalized inverse matrix of a matrix A and A' indicates the transpose of a matrix A. We assume that $\tilde{V}_0(x,\theta) - \tilde{V}_0(x,\theta_0) \in M\{V(x)\}$: the linear manifold generated by column vector of $V(x)$, for x, θ and θ_0.

From (3), the log likelihood function is given by

$$l_\varepsilon(X;\theta) = \int_0^T \varepsilon^{-2}\tilde{b}(X_t,\theta)\,\mathrm{d}X_t + \int_0^T \varepsilon^{-2}\tilde{c}(X_t,\theta)\,\mathrm{d}t, \tag{4}$$

where $\tilde{b}(x,\theta) = \tilde{V}_0'(VV')^+(x,\theta)$, $\tilde{c}(x,\theta) = -(1/2)\tilde{V}_0'(VV')^+\tilde{V}_0(x,\theta)$. From (1) and (4), the log likelihood function under the true model is given by

$$l_\varepsilon(X;\theta) = \int_0^T \varepsilon^{-1}b(X_t,\theta)\,\mathrm{d}w_t + \int_0^T \varepsilon^{-2}c(X_t,\theta)\,\mathrm{d}t, \tag{5}$$

where $b(x,\theta) = \tilde{b}(x,\theta)V(x)$, $c(x,\theta) = \tilde{c}(x,\theta) + \tilde{b}(x,\theta)V_0(x)$.

Next, we prepare several notations. Let X_t^0 be the solution of the ordinary differential equation

$$\frac{\mathrm{d}X_t^0}{\mathrm{d}t} = V_0(X_t^0), \quad t \in [0,T],\ X_0^0 = x_0.$$

Let an $\mathbf{R}^d \otimes \mathbf{R}^d$-valued process $Y_t^\varepsilon(w)$ be the solution of the stochastic differential equation

$$\mathrm{d}Y_t^\varepsilon = \partial V_0(X_t^\varepsilon)Y_t^\varepsilon\,\mathrm{d}t + \varepsilon\sum_{\alpha=1}^r \partial V_\alpha(X_t^\varepsilon)Y_t^\varepsilon\,\mathrm{d}w_t^\alpha, \quad t \in [0,T],$$
$$Y_0^\varepsilon = I_d,$$

where $[\partial V_\alpha]^{i,j} = \partial_j V_\alpha^i$, $\partial_j = \partial/\partial x^j$, $i, j = 1,\ldots,d$, $\alpha = 0,1,\ldots,r$. Then, $Y_t := Y_t^0$ is a deterministic $\mathbf{R}^d \otimes \mathbf{R}^d$-valued process. For a function $f(x,\theta)$ we abbreviate

$$f_t^\varepsilon(\theta) = f(X_t^\varepsilon,\theta), \qquad \partial_i f_t^\varepsilon(\theta) = \partial_i f(X_t^\varepsilon,\theta)$$

and similarly $\delta_j f_t^\varepsilon(\theta)$, $\delta_j\delta_l f_t^\varepsilon(\theta)$, $\partial_i\delta_j f_t^\varepsilon(\theta),\ldots$, where $\delta_j = \partial/\partial\theta^j$. It is known that $\varepsilon \to X_t^\varepsilon$ is smooth. In particular, $D_t := \partial X_t^\varepsilon/\partial\varepsilon|_{\varepsilon=0}$ satisfies the stochastic differential equation

$$\mathrm{d}D_t = \partial V_{0,t}^0 D_t\,\mathrm{d}t + V_t^0\,\mathrm{d}w_t, \quad t \in [0,T],\ D_0 = 0.$$

Then, D_t is represented by

$$D_t = \int_0^t Y_t Y_s^{-1}V_s^0\,\mathrm{d}w_s, \quad t \in [0,T].$$

$E_t := \partial^2 X_t^\varepsilon / \partial\varepsilon^2|_{\varepsilon=0}$ satisfies the stochastic differential equation

$$\begin{aligned}
\mathrm{d}E_t &= \sum_{i,j=1}^d \partial_i\partial_j V_{0,t}^0 D_t^i D_t^j \,\mathrm{d}t + \sum_{i=1}^d \partial_i V_{0,t}^0 E_t^i \,\mathrm{d}t + \\
&\quad + 2\sum_{i=1}^d \partial_i V_t^0 D_t^i \,\mathrm{d}w_t, \quad t \in [0,T], \\
E_0 &= 0.
\end{aligned}$$

Then, E_t is represented by

$$\begin{aligned}
E_t &= \int_0^t Y_t Y_s^{-1} \sum_{i,j=1}^d \partial_i\partial_j V_{0,s}^0 D_s^i D_s^j \,\mathrm{d}s + \\
&\quad + 2\int_0^t Y_t Y_s^{-1} \sum_{i=1}^d \partial_i V_s^0 D_s^i \,\mathrm{d}w_s, \quad t \in [0,T].
\end{aligned}$$

First of all, in order to explain the ideas heuristically, we assume the existence of an estimator $\hat\theta_\varepsilon$ which admits the stochastic expansion

$$\hat\theta_\varepsilon - \theta_0 = \varepsilon\zeta^{(0)} + \tfrac{1}{2}\varepsilon^2\zeta^{(1)} + o_p(\varepsilon^2)$$

for some $\theta_0 \in \Theta$ under the true model. Here $\zeta^{(0)} = \int_0^T g_t \,\mathrm{d}w_t$ for some function $g \in L^2([0,T], \mathrm{d}t)$. Let

$$\begin{aligned}
\bar S_\varepsilon^* &= \varepsilon\left\{ l_\varepsilon(X^\varepsilon(w), \hat\theta_\varepsilon(w)) - \int_W\int_W l_\varepsilon(X^\varepsilon(w'), \hat\theta_\varepsilon(w)) P(\mathrm{d}w')P(\mathrm{d}w) \right\} - \\
&\quad - \varepsilon b(\hat\theta_\varepsilon(w)),
\end{aligned}$$

where b is an $\mathbf{R}$-valued smooth function defined on $\mathbf{R}^p$. As in [34], in order to derive information criteria, we need the second-order asymptotic expansion of the distribution of $\bar S_\varepsilon^*$. It follows from the formal expansion that

$$\begin{aligned}
\bar S_\varepsilon^* &= \int_0^T b(X_t^0,\theta_0)\,\mathrm{d}w_t + \int_0^T \partial c(X_t^0,\theta_0) D_t \,\mathrm{d}t + \zeta^{(0)\prime}\int_0^T \delta c(X_t^0,\theta_0)\,\mathrm{d}t + \\
&\quad + \varepsilon\int_0^T \partial b(X_t^0,\theta_0) D_t \,\mathrm{d}w_t + \varepsilon\zeta^{(0)\prime}\int_0^T \delta b(X_t^0,\theta_0)\,\mathrm{d}w_t + \varepsilon\zeta^{(0)\prime} \times \\
&\quad \times \int_0^T \partial\delta c(X_t^0,\theta_0) D_t \,\mathrm{d}t + \frac{1}{2}\varepsilon P_{\mathbf{R}1^\perp}[\zeta^{(1)\prime}]\int_0^T \delta c(X_t^0,\theta_0)\,\mathrm{d}t + \\
&\quad + \frac{1}{2}\varepsilon\int_0^T \partial^2 c(X_t^0,\theta_0) P_{\mathbf{R}1^\perp}[(D_t, D_t)]\,\mathrm{d}t + \frac{1}{2}\varepsilon\int_0^T \partial c(X_t^0,\theta_0) \times \\
&\quad \times P_{\mathbf{R}1^\perp}[E_t]\,\mathrm{d}t + \frac{1}{2}\varepsilon P_{\mathbf{R}1^\perp}\left[(\zeta^{(0)})'\left(\int_0^T \delta^2 c(X_t^0,\theta_0)\,\mathrm{d}t\right)\zeta^{(0)}\right] - \\
&\quad - \varepsilon b(\theta_0) + o_p(\varepsilon) \\
&=: f_0 + \varepsilon(F_1 - b(\theta_0)) + o_p(\varepsilon) \quad \text{(say)},
\end{aligned}$$

where $P_{\mathbf{R}1^\perp}$ is the projection to the space orthogonal to the space $\mathbf{R}1$ in $L^2(P)$.

For $M > 0$ and $\gamma > 0$, the set $\mathcal{E}(M, \gamma)$ of measurable functions from $\mathbf{R} \to \mathbf{R}$ is defined by

$$\mathcal{E}(M, \gamma) = \{f : \mathbf{R} \to \mathbf{R},\ \text{measurable},\ |f(x)| \leqslant M(1 + |x|)^\gamma\}.$$

Let $\phi(x; \Sigma)$ be the probability density function of the one-dimensional normal distribution with mean 0 and variance $\Sigma = \mathrm{Var}[f_0]$.

From the viewpoint of the second-order statistical inference, we will use the following asymptotic expansion

$$\begin{aligned} E[f(\bar{S}_\varepsilon^*)] &= \int_{\mathbf{R}} f(x)\phi(x; \Sigma)\,\mathrm{d}x - \varepsilon \int_{\mathbf{R}} f(x)\partial_x \times \\ &\quad \times \{E[F_1 - b(\theta_0)|f_0 = x]\phi(x; \Sigma)\}\,\mathrm{d}x + o(\varepsilon) \\ &= \int_{\mathbf{R}} f(x)\phi(x; \Sigma)\,\mathrm{d}x - \varepsilon \int_{\mathbf{R}} f(x)\partial_x\{E[F_1|f_0 = x]\phi(x; \Sigma)\}\,\mathrm{d}x + \\ &\quad + \varepsilon b(\theta_0) \int_{\mathbf{R}} f(x)\partial_x\{\phi(x; \Sigma)\}\,\mathrm{d}x + o(\varepsilon) \end{aligned}$$

for any $f \in \mathcal{E}(M, \gamma)$. See Section 5 for technical details. For $M > 0$ and $\gamma > 0$, the set $\mathcal{E}'(M, \gamma)$ of measurable functions from $\mathbf{R} \to \mathbf{R}$ is defined by

$$\begin{aligned} \mathcal{E}'(M, \gamma) = \Bigg\{ f \in \mathcal{E}(M, \gamma) \Bigg| & \int_{\mathbf{R}} f(x)\phi(x; \Sigma)\,\mathrm{d}x = 0, \\ & \int_{\mathbf{R}} f(x)\partial_x\{\phi(x; \Sigma)\}\,\mathrm{d}x \neq 0 \Bigg\}. \end{aligned}$$

For any $f \in \mathcal{E}'(M, \gamma)$, let $IC_f(X; \hat{\theta}_\epsilon) = \varepsilon l_\varepsilon(X, \hat{\theta}_\varepsilon) - \varepsilon b_f(\hat{\theta}_\varepsilon)$, where

$$b_f(\theta_0) = \frac{\int_{\mathbf{R}} f(x)\partial_x\{E[F_1|f_0 = x]\phi(x; \Sigma)\}\,\mathrm{d}x}{\int_{\mathbf{R}} f(x)\partial_x\{\phi(x; \Sigma)\}\,\mathrm{d}x}.$$

Then, under certain regularity conditions, IC_f is the f-unbiased information criterion for $f \in \mathcal{E}'(M, \gamma)$, that is,

$$E\left[f(IC_f(X; \hat{\theta}_\epsilon) - \varepsilon \int_W \int_W l_\varepsilon(X^\varepsilon(w'), \hat{\theta}_\varepsilon(w))P(\mathrm{d}w')P(\mathrm{d}w))\right] = o(\varepsilon).$$

In particular, for $f(x) = x$, we obtain the asymptotically expectation unbiased information criterion (EUIC). Moreover, for $f(x) = 1_{(-\infty,0)} - 1/2$ and $f(x) = 1_{(0,\infty)} - 1/2$, we also have the second-order asymptotically MUIC. For details of the second-order asymptotically median unbiasedness, see [1].

Information criterion 1 (EUIC). Let $IC_1(X; \hat{\theta}_\epsilon) = \varepsilon l_\varepsilon(X, \hat{\theta}_\varepsilon) - \varepsilon b_1(\hat{\theta}_\varepsilon)$, where $b_1(\theta_0) = E[F_1]$. Then

$$E\left[IC_1(X; \hat{\theta}_\epsilon) - \varepsilon \int_W \int_W l_\varepsilon(X^\varepsilon(w'), \hat{\theta}_\varepsilon(w))P(\mathrm{d}w')P(\mathrm{d}w)\right] = o(\varepsilon).$$

Information criterion 2 (MUIC). Let $IC_2(X;\hat{\theta}_\epsilon) = \varepsilon l_\varepsilon(X,\hat{\theta}_\varepsilon) - \varepsilon b_2(\hat{\theta}_\varepsilon)$, where $b_2(\theta_0) = E[F_1|f_0 = 0]$. Then

$$P\left[IC_2(X;\hat{\theta}_\epsilon) - \varepsilon\int_W\int_W l_\varepsilon(X^\varepsilon(w'),\hat{\theta}_\varepsilon(w))P(\mathrm{d}w')P(\mathrm{d}w) < 0\right] = \frac{1}{2} + o(\varepsilon),$$

and

$$P\left[IC_2(X;\hat{\theta}_\epsilon) - \varepsilon\int_W\int_W l_\varepsilon(X^\varepsilon(w'),\hat{\theta}_\varepsilon(w))P(\mathrm{d}w')P(\mathrm{d}w) > 0\right] = \frac{1}{2} + o(\varepsilon).$$

Remark 1. For the deviation of IC_1 in the misspecified case, we need to estimate unknown functions $V_0(X_t^0)$ and $\partial V_0(X_t^0)$ in order to calculate the biased term b_1. Using the kernel-type estimator $\hat{V}_{0,t}^\varepsilon$ presented by Kutoyants [17], we can obtain the estimated biased term $\hat{b}_1$. It is also possible to derive IC_2 for the misspecified model in a similar way.

3. Information Criteria with M-estimators

In this section, we consider the information criteria which work for an M-estimator defined as a solution of a given estimating function.

Define a functional Ψ_ε by

$$\Psi_\varepsilon(X;\theta) = \int_0^T \varepsilon^{-2}\tilde{B}(X_t,\theta)\,\mathrm{d}X_t + \int_0^T \varepsilon^{-2}\tilde{C}(X_t,\theta)\,\mathrm{d}t, \tag{6}$$

where $\tilde{B}$, $\tilde{C}$ are given functions. From (1) and (6), the functional Ψ_ε under the true model is given by

$$\Psi_\varepsilon(X;\theta) = \int_0^T \varepsilon^{-1}B(X_t,\theta)\,\mathrm{d}w_t + \int_0^T \varepsilon^{-2}C(X_t,\theta)\,\mathrm{d}t, \tag{7}$$

where $B(x,\theta) = \tilde{B}(x,\theta)V(x)$, $C(x,\theta) = \tilde{C}(x,\theta) + \tilde{B}(x,\theta)V_0(x)$.

Let $\hat{\theta}_\varepsilon$ be the M-estimator, that is,

$$\hat{\theta}_\varepsilon := \begin{cases} \underset{\theta}{\operatorname{argmax}}\,\Psi_\varepsilon(X^\varepsilon,\theta) & \text{if maximum exists in } \Theta, \\ \text{arbitary}, & \text{otherwise}. \end{cases} \tag{8}$$

ASSUMPTION 1. V, V_0, B and C satisfy the following conditions.

(i) $V(x) \in C_b^\infty(\mathbf{R}^d \to \mathbf{R}^d \otimes \mathbf{R}^r)$, that is, $V(x)$ is an $\mathbf{R}^d \otimes \mathbf{R}^r$-valued smooth function defined on $\mathbf{R}^d$ with bounded x-derivatives.

(ii) $V_0(x) \in C_b^\infty(\mathbf{R}^d \to \mathbf{R}^d)$.

(iii) $B(x,\theta) \in C^{\infty}_{\uparrow}(\mathbf{R}^d \times \Theta \to \mathbf{R} \otimes \mathbf{R}^r (\cong \mathbf{R}^r))$, that is, $B(x,\theta)$ is an $\mathbf{R}^r$-valued smooth function defined on $\mathbf{R}^d \times \Theta$ and for any $i, j \in \mathbf{N}$ there exist $m_1, C_1 > 0$ such that
$$\sup_{\theta\in\Theta} |\delta^i \partial^j B(x,\theta)| \leqslant C_1(1+|x|)^{m_1}$$
for any $x \in \mathbf{R}^d$.

(iv) $C(x,\theta) \in C^{\infty}_{\uparrow}(\mathbf{R}^d \times \Theta \to \mathbf{R})$. Moreover, there exist $\theta_0 \in \Theta$ and $a_0 > 0$ such that
$$\int_0^T \{C(X_t^0,\theta) - C(X_t^0,\theta_0)\}\,\mathrm{d}t \leqslant -a_0|\theta-\theta_0|^2.$$

Define several functions as follows.

$$\begin{aligned}
\lambda_{t,s} &= Y_t Y_s^{-1} V(X_s^0),\\
\lambda_{t,s}^i &= [Y_t Y_s^{-1} V(X_s^0)]^{i\cdot}, \quad i=1,\dots,d,\\
\mu_{i,t,s} &= Y_t Y_s^{-1} \partial_i V(X_s^0), \quad i=1,\dots,d,\\
\mu_{i,t,s}^j &= [Y_t Y_s^{-1} \partial_i V(X_s^0)]^{j\cdot}, \quad i,j=1,\dots,d,\\
\nu_{i,j,t,s} &= Y_t Y_s^{-1} \partial_i\partial_j V_0(X_s^0), \quad i,j=1,\dots,d,\\
\nu_{i,j,t,s}^l &= [Y_t Y_s^{-1} \partial_i\partial_j V_0(X_s^0)]^{l}, \quad i,j,l=1,\dots,d,\\
I(\theta_0) &= \int_0^T \delta^2 C(X_t^0,\theta_0)\,\mathrm{d}t,\\
I(\theta_0)^{i,j} &= [I(\theta_0)^{-1}]^{i,j}, \quad i,j=1,\dots,p,\\
b_{0,t}^i &= -\sum_{j=1}^p I(\theta_0)^{i,j}\left\{\delta_j B(X_t^0,\theta_0) + \sum_{\alpha=1}^d \int_t^T \partial_\alpha \delta_j C(X_s^0,\theta_0)\lambda_{s,t}^\alpha\,\mathrm{d}s\right\},\\
b_{1,t}^{j,k} &= \delta_j\delta_k B(X_t^0,\theta_0) + \sum_{\alpha=1}^d \int_t^T \partial_\alpha\delta_j\delta_k C(X_s^0,\theta_0)\lambda_{s,t}^\alpha\,\mathrm{d}s,\\
b_{2,t}^i &= \delta_i b(X_t^0,\theta_0) + \sum_{\alpha=1}^d \int_t^T \partial_\alpha\delta_i c(X_s^0,\theta_0)\lambda_{s,t}^\alpha\,\mathrm{d}s,\\
C_0 &= E\left[\frac{1}{2}\sum_{i=1}^p \int_0^T \delta_i c(X_t^0,\theta_0)\,\mathrm{d}t\,\zeta_i^{(1)} + \right.\\
&\quad + \frac{1}{2}\sum_{\alpha,\beta=1}^d \int_0^T \partial_\alpha\partial_\beta c(X_t^0,\theta_0) D_t^\alpha D_t^\beta\,\mathrm{d}t + \frac{1}{2}\sum_{\alpha=1}^d \int_0^T \partial_\alpha c(X_t^0,\theta_0)\times\\
&\quad \left.\times E_t^\alpha\,\mathrm{d}t + \frac{1}{2}\sum_{i,j=1}^p \int_0^T \delta_i\delta_j c(X_t^0,\theta_0)\,\mathrm{d}t\,\zeta_i^{(0)}\zeta_j^{(0)}\right].\\
a_t &= b(X_t^0,\theta_0) + \sum_{\alpha=1}^d \int_t^T \partial_\alpha c(X_s^0,\theta_0)\lambda_{s,t}^\alpha\,\mathrm{d}s + \sum_{i=1}^p \int_0^T \delta_i c(X_t^0,\theta_0)\,\mathrm{d}t\, b_{0,t}^i.
\end{aligned}$$

Under Assumption 1, $\varepsilon^{-1}(\hat{\theta}_\varepsilon - \theta_0)$ has the asymptotic expansion

$$\varepsilon^{-1}(\hat{\theta}_\varepsilon - \theta_0)^i \sim \zeta_i^{(0)} + \tfrac{1}{2}\varepsilon\zeta_i^{(1)} + \cdots$$

as $\varepsilon \downarrow 0$ in some sense, where

$$\begin{aligned}
\zeta_i^{(0)} &= \int_0^T b_{0,t}^i \,\mathrm{d}w_t, && (9)\\
\zeta_i^{(1)} &= -\sum_{j=1}^p I(\theta_0)^{i,j}\Bigg[2\sum_{\alpha=1}^d \int_0^T \partial_\alpha\delta_j B(X_t^0,\theta_0)D_t^\alpha \,\mathrm{d}w_t + \\
&\quad + 2\sum_{k=1}^p \int_0^T b_{1,t}^{j,k}\,\mathrm{d}w_t \cdot \zeta_k^{(0)} + \sum_{\alpha,\beta=1}^d \int_0^T \partial_\alpha\partial_\beta\delta_j C(X_t^0,\theta_0)D_t^\alpha D_t^\beta \,\mathrm{d}t + \\
&\quad + \sum_{\alpha=1}^d \int_0^T \partial_\alpha\delta_j C(X_t^0,\theta_0)E_t^\alpha \,\mathrm{d}t + \\
&\quad + \sum_{k,l=1}^p \int_0^T \partial_j\partial_k\partial_l C(X_t^0,\theta_0)\,\mathrm{d}t \cdot \zeta_k^{(0)}\zeta_l^{(0)}\Bigg]. && (10)
\end{aligned}$$

For a rigorous statement of the asymptotic expansion of $\varepsilon^{-1}(\hat{\theta}_\varepsilon - \theta_0)$, see Lemma 4. Moreover, under Assumption 1, $\bar{S}_\varepsilon^*$ has the asymptotic expansion

$$\bar{S}_\varepsilon^* \sim f_0 + \varepsilon(F_1 - b(\theta_0)) + \cdots$$

as $\varepsilon \downarrow 0$ in some sense, where

$$\begin{aligned}
f_0 &= \int_0^T a_t \,\mathrm{d}w_t,\\
F_1 &= \sum_{\alpha=1}^d \int_0^T \partial_\alpha b(X_t^0,\theta_0)D_t^\alpha \,\mathrm{d}w_t + \sum_{i=1}^p \int_0^T b_{2,t}^i \,\mathrm{d}w_t \zeta_i^{(0)} + \\
&\quad + \frac{1}{2}\sum_{i=1}^p \int_0^T \delta_i c(X_t^0,\theta_0)\,\mathrm{d}t\zeta_i^{(1)} + \frac{1}{2}\sum_{\alpha,\beta=1}^d \int_0^T \partial_\alpha\partial_\beta c(X_t^0,\theta_0)D_t^\alpha D_t^\beta \,\mathrm{d}t + \\
&\quad + \frac{1}{2}\sum_{\alpha=1}^d \int_0^T \partial_\alpha c(X_t^0,\theta_0)E_t^\alpha \,\mathrm{d}t + \\
&\quad + \frac{1}{2}\sum_{i,j=1}^p \int_0^T \delta_i\delta_j c(X_t^0,\theta_0)\,\mathrm{d}t\zeta_i^{(0)}\zeta_j^{(0)} - C_0. && (11)
\end{aligned}$$

For a rigorous and general statement of the asymptotic expansion of $\bar{S}_\varepsilon^*$, see Lemma 5.

THEOREM 1 (EUIC). *Suppose that Assumption* 1 *holds true. Let* $IC_1(X;\hat{\theta}_\varepsilon) = \varepsilon l_\varepsilon(X,\hat{\theta}_\varepsilon) - \varepsilon b_1(\hat{\theta}_\varepsilon)$, *where*

$$\begin{aligned} b_1(\theta_0) &= \sum_{i=1}^{p} \int_0^T b_{0,t}^i (\delta_i b(X_t^0,\theta_0))' \,\mathrm{d}t + \\ &\quad + \sum_{i=1}^{p}\sum_{\alpha=1}^{d} \int_0^T b_{0,t}^i \left(\int_t^T \partial_\alpha \delta_i c(X_s^0,\theta_0)\lambda_{s,t}^\alpha \,\mathrm{d}s \right)' \mathrm{d}t. \end{aligned}$$

Then,

$$E\left[IC_1(X;\hat{\theta}_\varepsilon) - \varepsilon \int_W \int_W l_\varepsilon(X^\varepsilon(w'),\hat{\theta}_\varepsilon(w))P(\mathrm{d}w')P(\mathrm{d}w) \right] = o(\varepsilon).$$

Remark 2. In particular, in the correctly specified and MLE case, since $\delta c(X_s^0,\theta_0) = 0$,

$$b_{0,t}^i = -\sum_{j=1}^{p} I_0(\theta_0)^{i,j} \delta_j b(X_t^0,\theta_0),$$

where $I_0(\theta_0) = \int_0^T \delta^2 c(X_t^0,\theta_0)\,\mathrm{d}t$. We then have AIC:

$$\begin{aligned} b_1(\theta_0) &= -\sum_{i,j=1}^{p} I_0(\theta_0)^{i,j} \int_0^T \delta_j b(X_t^0,\theta_0)(\delta_i b(X_t^0,\theta_0))' \,\mathrm{d}t \\ &= p \text{ (dimension of parameter space)}. \end{aligned}$$

ASSUMPTION 2. There exists $t \in [0,T]$ such that $a_t \neq 0$.

For $\mathbf{R} \otimes \mathbf{R}^r$-valued function h_t,

$$C_2^i(h)_T = \int_0^T \int_0^t a_t h_t' \lambda_{t,s}^i a_s' \,\mathrm{d}s\,\mathrm{d}t, \quad i = 1,\ldots,d.$$

For $\mathbf{R} \otimes \mathbf{R}^r$-valued functions b_t and c_t, put

$$C_2(b,c)_T = \frac{1}{2}\int_0^T \int_0^T a_t\, [b_t' c_s + c_t' b_s] a_s' \,\mathrm{d}s\,\mathrm{d}t.$$

Let

$$C_2^{i,j}(t) = C_2(\lambda_{t,\cdot}^i I_{\{\cdot \leqslant t\}}, \lambda_{t,\cdot}^j I_{\{\cdot \leqslant t\}})_T.$$

For the second-order asymptotically MUIC, it is necessary to calculate $E[F_1 | f_0 = x]$ explicitly. For this purpose we prepare the following lemma. The proof is easy and is omitted.

LEMMA 1. *Let w be an r-dimensional Wiener process and let functions a_t, b_t, c_t on $[0, T]$ be deterministic. Let $\Sigma = \int_0^T a_t a_t'\, dt$.*

(1) *Let a_t, b_t and c_t be $\mathbf{R} \otimes \mathbf{R}^r$-valued functions. Then*

$$\begin{aligned} I^{11}(b_t, c_t)(x) &:= E\left[\int_0^T b_t\, dw_t \int_0^T c_t\, dw_t \,\middle|\, \int_0^T a_t\, dw_t = x\right] \\ &= \frac{1}{2}\int_0^T \int_0^T a_t\, [c_t' b_s + b_t' c_s] a_s' \Sigma^{-2}(x^2 - \Sigma)\, ds\, dt + \\ &\quad + \int_0^T b_t c_t'\, dt. \end{aligned}$$

(2) *For $\mathbf{R} \otimes \mathbf{R}^r$-valued function h_t,*

$$\begin{aligned} J_i^1(h_t)(x) &:= E\left[\int_0^T h_t D_t^i\, dw_t \,\middle|\, \int_0^T a_t\, dw_t = x\right] \\ &= [C_2^i(h)_T]\Sigma^{-2}(x^2 - \Sigma). \end{aligned}$$

(3) *For $\mathbf{R}$-valued function h_t,*

$$\begin{aligned} J_{i,j}^2(h_t)(x) &:= E\left[\int_0^T h_t D_t^i D_t^j\, dt \,\middle|\, \int_0^T a_t\, dw_t = x\right] \\ &= \int_0^T h_t\, [C_2^{i,j}(t)]\, dt\, \Sigma^{-2}(x^2 - \Sigma) + \\ &\quad + \int_0^T \int_0^t h_t \lambda_{t,s}^i (\lambda_{t,s}^j)'\, ds\, dt. \end{aligned}$$

(4) *For $\mathbf{R}$-valued function h_t,*

$$\begin{aligned} J_\alpha^3(h_t)(x) &:= E\left[\int_0^T h_t E_t^\alpha\, dt \,\middle|\, \int_0^T a_t\, dw_t = x\right] \\ &= \int_0^T h_t \int_0^t \nu_{i,j,t,s}^\alpha\, [C_2^{i,j}(t)]\, ds\, dt\, \Sigma^{-2}(x^2 - \Sigma) + \\ &\quad + \int_0^T h_t \int_0^t \int_0^s \nu_{i,j,t,s}^\alpha \lambda_{s,u}^i (\lambda_{s,u}^j)'\, du\, ds\, dt + \\ &\quad + \int_0^T h_t\, [C_2^i(2\mu_{i,t,\cdot}^\alpha)_t]\, dt\, \Sigma^{-2}(x^2 - \Sigma). \end{aligned}$$

THEOREM 2 (MUIC). *Suppose that Assumptions 1 and 2 hold true. Let $IC_2(X; \hat{\theta}_\varepsilon) = \varepsilon l_\varepsilon(X, \hat{\theta}_\varepsilon) - \varepsilon b_2(\hat{\theta}_\varepsilon)$, where*

$$\begin{aligned} b_2(\theta_0) &= \sum_{\alpha=1}^d J_\alpha^1(\partial_\alpha b(X_t^0, \theta_0))(0) + \sum_{i=1}^p I^{11}(b_{2,t}^i, b_{0,t}^i)(0) - \\ &\quad - \frac{1}{2}\sum_{i,j=1}^p I(\theta_0)^{i,j} \int_0^T \delta_i c(X_t^0, \theta_0)\, dt \left\{ 2\sum_{\alpha=1}^d J_\alpha^1(\partial_\alpha \delta_j B(X_t^0, \theta_0))(0) + \right. \end{aligned}$$

$$
\begin{aligned}
&+ 2\sum_{k=1}^{p} I^{11}(b_{1,t}^{j,k}, b_{0,t}^{k})(0) + \sum_{\alpha,\beta=1}^{d} J_{\alpha,\beta}^{2}(\partial_\alpha\partial_\beta\delta_j C(X_t^0,\theta_0))(0) + \\
&+ \sum_{\alpha=1}^{d} J_{\alpha}^{3}(\partial_\alpha\delta_j C(X_t^0,\theta_0))(0) + \\
&+ \sum_{k,l=1}^{p} \int_0^T \partial_j\partial_k\partial_l C(X_t^0,\theta_0)\,\mathrm{d}t\, I^{11}(b_{0,t}^{k}, b_{0,t}^{l})(0)\Bigg\} + \\
&+ \frac{1}{2}\sum_{\alpha,\beta=1}^{d} J_{\alpha,\beta}^{2}(\partial_\alpha\partial_\beta c(X_t^0,\theta_0))(0) + \frac{1}{2}\sum_{\alpha=1}^{d} J_{\alpha}^{3}(\partial_\alpha c(X_t^0,\theta_0))(0) + \\
&+ \frac{1}{2}\sum_{i,j=1}^{p} \int_0^T \delta_i\delta_j c(X_t^0,\theta_0)\,\mathrm{d}t\, I^{11}(b_{0,t}^{i}, b_{0,t}^{j})(0) - C_0.
\end{aligned}
$$

Then,

$$
\begin{aligned}
&P\left[IC_2(X;\hat{\theta}_\varepsilon) - \varepsilon\int_W\int_W l_\varepsilon(X^\varepsilon(w'),\hat{\theta}_\varepsilon(w))P(\mathrm{d}w')P(\mathrm{d}w) < 0\right] \\
&\quad = \frac{1}{2} + o(\varepsilon),
\end{aligned}
$$

and

$$
\begin{aligned}
&P\left[IC_2(X;\hat{\theta}_\varepsilon) - \varepsilon\int_W\int_W l_\varepsilon(X^\varepsilon(w'),\hat{\theta}_\varepsilon(w))P(\mathrm{d}w')P(\mathrm{d}w) > 0\right] \\
&\quad = \frac{1}{2} + o(\varepsilon).
\end{aligned}
$$

4. Fundamental Results: Malliavin Calculus with Truncation[1]

Let (W, P) be the r-dimensional Wiener space and let H be the Cameron–Martin subspace of W endowed with the inner product $\langle h_1, h_2\rangle_H = \int_0^T (\dot{h}_{1,t}, \dot{h}_{2,t})\,\mathrm{d}t$ for $h_1, h_2 \in H$. For a Hilbert space E, $\|\cdot\|_p$ denotes the $L^p(E)$-norm of E-valued Wiener functional, that is, for Wiener functional $f : (W, P) \to E$, $\|f\|_p^p = \int_W |f|_E^p P(\mathrm{d}w)$, where $|f|_E = \langle f, f\rangle_E^{1/2}$ and $\langle\cdot,\cdot\rangle_E$ is the inner product of E. For $s \in \mathbf{R}$ and $p \in (1,\infty)$, the norm $\|\cdot\|_{p,s}$ on the totality of E-valued Wiener functional f is defined by $\|f\|_{p,s} = \|(I-L)^{s/2} f\|_p$, where L is the Ornstein–

[1] A simple exposition of the Malliavin calculus towards statistics may be seen in [46].

Uhlenbeck operator (see [38]). The Banach space $D_p^s(E)$ is the completion of the totality $P(E)$ of E-valued polynomials on the Wiener space (W, P) with respect to $\|\cdot\|_{p,s}$. The set of Wiener test functionals of Watanabe [38] is denoted by $D^\infty(E) = \bigcap_{s>0}\bigcap_{1<p<\infty} D_p^s(E)$. Then, $D^{-\infty}(E) = \bigcup_{s>0}\bigcup_{1<p<\infty} D_p^{-s}(E)$ and $\tilde{D}^{-\infty}(E) = \bigcup_{s>0}\bigcap_{1<p<\infty} D_p^{-s}(E)$ are the spaces of generalized Wiener functionals. We suppress $\mathbf{R}$ when $E = \mathbf{R}$. The Fréchet space $S(\mathbf{R}^d)$ is the totality of rapidly decreasing smooth functions on $\mathbf{R}^d$ and $S'(\mathbf{R}^d)$ is its dual. Let $A = 1+|x|^2-(1/2)\Delta$, where $\Delta = \sum_{i=1}^d(\partial/\partial x^i)^2$, and A^{-1} is an integral operator. The space C_{2k}, $k = 0, \pm 1, \pm 2, \dots$ is the completion of $S(\mathbf{R}^d)$ with respect to the norm $\|u\|_{2k} = \sup_x|A^k u(x)|$. We owe the following theorem. For details, see [27, 32, 38, 40].

THEOREM 3 (Yoshida [40]). *Let* $F \in D^\infty(\mathbf{R}^d)$ *and* $\xi \in D^\infty$. Let $\psi : \mathbf{R} \to \mathbf{R}$ *be a smooth function such that* $0 \leqslant \psi(x) \leqslant 1$ *for* $x \in \mathbf{R}$, $\psi(x) = 1$ for $|x| \leqslant 1/2$ *and* $\psi(x) = 0$ *for* $|x| \geqslant 1$. *Suppose that for any* $p \in (1, \infty)$, *the Malliavin covariance* σ_F of F *satisfies*

$$E[1_{\{|\xi| \leqslant 1\}}(\det \sigma_F)^{-p}] < \infty.$$

Then, there exists a linear mapping $T \in S'(\mathbf{R}^d) \to \hat{T} \in \tilde{D}^{-\infty}$ *satisfying the following conditions*:

(1) *if* $T \in S(\mathbf{R}^d)$ *then* $\hat{T} = \psi(\xi)T(F) \in D^\infty$;
(2) *for* $k = 0, 1, \dots$ *and* $p \in (1, \infty)$ *there exists a constant* $C(p, k)$ *such that*
$$\|\hat{T}\|_{p,-2k} \leqslant C(p, k)\|T\|_{-2k}$$
for $T \in C_{-2k}$. *This mapping is uniquely determined.*

If F is nondegenerate in the usual sense of Malliavin,

$${}_{D^{-\infty}}\langle \hat{T}, J\rangle_{D^\infty} = {}_{D^{-\infty}}\langle T(F), \psi(\xi)J\rangle_{D^\infty}$$

for $J \in D^\infty$. Thus $\hat{T}$ is denoted by $\psi(\xi)T \circ F$ or $\psi(\xi)T(F)$ if there is no confusion.

Let us consider a family of E-valued Wiener functionals (or generalized Wiener functionals) $\{F_\varepsilon(w)\}$, $\varepsilon \in (0, 1)$. For $k > 0$ if

$$\limsup_{\varepsilon\downarrow 0} \frac{\|F_\varepsilon\|_{p,s}}{\varepsilon^k} < \infty,$$

we say $F_\varepsilon(w) = O(\varepsilon^k)$ in $D_p^s(E)$ as $\varepsilon \downarrow 0$. It is said that $F_\varepsilon(w) \in D^\infty(E)$ has the asymptotic expansion

$$F_\varepsilon(w) \sim f_0 + \varepsilon f_1 + \cdots$$

in $D^\infty(E)$ as $\varepsilon \downarrow 0$ with $f_0, f_1, \dots \in D^\infty(E)$ if for every $p > 1, s > 0$ and $k = 1, 2, \dots$

$$F_\varepsilon(w) - (f_0 + \varepsilon f_1 + \cdots + \varepsilon^{k-1} f_{k-1}) = O(\varepsilon^k)$$

in $D_p^s(E)$ as $\varepsilon \downarrow 0$. Similarly, we say that $F_\varepsilon(w) \in \tilde{D}^{-\infty}(E)$ has the asymptotic expansion

$$F_\varepsilon(w) \sim f_0 + \varepsilon f_1 + \cdots$$

in $\tilde{D}^{-\infty}(E)$ as $\varepsilon \downarrow 0$ with $f_0, f_1, \ldots \in \tilde{D}^{-\infty}(E)$ if for every $k = 1, 2, \ldots$ there exists $s > 0$ such that for every $p > 1$ $F_\varepsilon(w), f_0, f_1, \ldots \in D_p^{-s}(E)$ and

$$F_\varepsilon(w) - (f_0 + \varepsilon f_1 + \cdots + \varepsilon^{k-1} f_{k-1}) = O(\varepsilon^k)$$

in $D_p^{-s}(E)$ as $\varepsilon \downarrow 0$. The generalized means of these expansions yield the ordinary asymptotic expansions.

The following theorem will be our fundamental tool.

THEOREM 4 (Takanobu and Watanabe [32] and Yoshida [41]). *Let ψ be a function defined in Theorem 3. Let Λ be an index set. Suppose that families $\{F_\varepsilon(w); \varepsilon \in (0, 1]\} \subset D^\infty(\mathbf{R}^d)$, $\{\xi_\varepsilon(w); \varepsilon \in (0, 1]\} \subset D^\infty$ and $\{T_\lambda; \lambda \in \Lambda\} \subset S'(\mathbf{R}^d)$ satisfy the following conditions.*

(1) *For any $p \in (1, \infty)$*
$$\sup_{\varepsilon\in(0,1]} E[1_{\{|\xi_\varepsilon| \leqslant 1\}} (\det \sigma_{F_\varepsilon})^{-p}] < \infty.$$
(2) $\{F_\varepsilon(w); \varepsilon \in (0, 1]\}$ *has the asymptotic expansion*
$$F_\varepsilon(w) \sim f_0 + \varepsilon f_1 + \cdots \text{ in } D^\infty(\mathbf{R}^d) \text{ as } \varepsilon \downarrow 0$$
with $f_i \in D^\infty(\mathbf{R}^d)$.
(3) $\{\xi_\varepsilon(w); \varepsilon \in (0, 1]\}$ *is $O(1)$ in D^∞ as $\varepsilon \downarrow 0$.*
(4) *For any $n = 1, 2, \ldots,$*
$$\lim_{\varepsilon\downarrow 0} \varepsilon^{-n} P\{|\xi_\varepsilon| > \tfrac{1}{2}\} = 0.$$
(5) *For any $n = 1, 2, \ldots,$ there exists a nonnegative integer m such that $A^{-m} T_\lambda \in C_b^n(\mathbf{R}^d)$ for all $\lambda \in \Lambda$ and*
$$\sup_{\lambda\in\Lambda} \sum_{|\mathbf{n}| \leqslant n} \|\partial^{\mathbf{n}} A^{-m} T_\lambda\|_\infty < \infty,$$
where $\mathbf{n} = (n_1, \ldots, n_d)$ is a multi-index, $|\mathbf{n}| = n_1 + \cdots + n_d$, $\partial^{\mathbf{n}} = \partial_1^{n_1} \cdots \partial_d^{n_d}$, $\partial_i = \partial/\partial x^i$, $i = 1, \ldots, d$. Then the composite functional $\psi(\xi_\varepsilon) T_\lambda(F_\varepsilon) \in \tilde{D}^{-\infty}$ is well defined and has the asymptotic expansion
$$\psi(\xi_\varepsilon) T_\lambda(F_\varepsilon) \sim \Psi_{\lambda,0} + \varepsilon \Psi_{\lambda,1} + \cdots$$
in $\tilde{D}^{-\infty}$ as $\varepsilon \downarrow 0$ uniformly in $\lambda \in \Lambda$ with $\Psi_{\lambda,0}, \Psi_{\lambda,1}, \ldots \in \tilde{D}^{-\infty}$ determined by the formal Taylor expansion
$$\begin{aligned} T_\lambda(f_0 + [\varepsilon f_1 + \varepsilon^2 f_2 + \cdots]) &= \sum_{\mathbf{n}} \frac{1}{\mathbf{n}!} \partial^{\mathbf{n}} T_\lambda(f_0)\, [\varepsilon f_1 + \varepsilon^2 f_2 + \cdots]^{\mathbf{n}} \\ &= \Psi_{\lambda,0} + \varepsilon \Psi_{\lambda,1} + \varepsilon^2 \Psi_{\lambda,2} + \cdots, \end{aligned}$$

where $\mathbf{n}! = n_1! \cdots n_d!$, $a^{\mathbf{n}} = a_1^{n_1} \cdots a_d^{n_d}$ *for* $a \in \mathbf{R}^d$. *In particular,*

$$\Psi_{\lambda,0} = T_\lambda(f_0), \qquad \Psi_{\lambda,1} = \sum_{i=1}^{d} f_1^i \partial_i T_\lambda(f_0),$$

$$\Psi_{\lambda,2} = \sum_{i=1}^{d} f_2^i \partial_i T_\lambda(f_0) + \frac{1}{2} \sum_{i,j=1}^{d} f_1^i f_1^j \partial_i \partial_j T_\lambda(f_0).$$

Remark 3. If we consider the asymptotic expansion of $E[f(F_\varepsilon)]$ for a particular measurable function f, there is no need for Theorem 4 to assume condition (5).

LEMMA 2 (Yoshida [40]). *Let* $M, \gamma > 0$. *For* $n = 1, 2, \ldots$, *there exists a positive integer m such that*

$$\sup_{f \in \tilde{\mathcal{E}}(M,\gamma)} \sum_{|\mathbf{n}| \leqslant n} \|\partial^{\mathbf{n}} A^{-m} f\|_\infty < \infty,$$

where $\tilde{\mathcal{E}}(M, \gamma) = \{f : \mathbf{R}^d \to \mathbf{R}$, *measurable*, $|f(x)| \leqslant M(1 + |x|)^\gamma \ (x \in \mathbf{R}^d)\}$.

The composite function of a measurable function and a Wiener functional has a usual meaning.

LEMMA 3 (Yoshida [40]). *Let* $M, \gamma > 0$. *For* ψ, ξ, F *given in Theorem* 3 *and any* $f \in \tilde{\mathcal{E}}(M, \gamma)$,

$$\psi(\xi) f \circ F = \psi(\xi) f(F)$$

in $\tilde{D}^{-\infty}$.

5. Information Criteria with General Estimators

In this and the next sections, we will return to the information criteria for general estimators given in Section 2, and provide a rigorous mathematical basis for the heuristic argument there. The proofs will be put in Section 6. The results in Section 3 are merely corollaries of the ones in this section, and proved in Section 6. In order to treat our problem in a rigorous way, we need the so-called truncation technique explained in Section 4. We will suppose a weak Assumption 3 for the estimator $\hat{\theta}_\varepsilon$. It may seem slightly technical, involving the existence of a truncation functional, however, the following lemma shows that Assumption 3 is satisfied for the M-estimators.

LEMMA 4. *Suppose that Assumption* 1 *holds true. Then there exists* $R_\varepsilon = O(1)$ in D^∞ *such that for every* $K > 0$ *and* $c > 0$,

$$\lim_{\varepsilon \downarrow 0} \varepsilon^{-K} P[|R_\varepsilon| > c] = 0,$$

and for $\psi_\varepsilon = \psi(3R_\varepsilon)$, *the functional* $\psi_\varepsilon \varepsilon^{-1}(\hat{\theta}_\varepsilon - \theta_0)$ *is in* $D^\infty(\mathbf{R}^p)$ *and has the asymptotic expansion*

$$\psi_\varepsilon \varepsilon^{-1}(\hat{\theta}_\varepsilon - \theta_0) \sim \zeta^{(0)} + \tfrac{1}{2}\varepsilon\zeta^{(1)} + \cdots$$

in $D^\infty(\mathbf{R}^p)$ *as* $\varepsilon \downarrow 0$ *with* $\zeta^{(0)}, \zeta^{(1)}, \ldots \in D^\infty(\mathbf{R}^p)$. *In particular,*

$$\begin{aligned}
\zeta^{(0)} &= -I(\theta_0)^{-1}\left[\int_0^T \delta B(X_t^0, \theta_0)\,\mathrm{d}w_t + \int_0^T \partial\delta C(X_t^0, \theta_0) D_t\,\mathrm{d}t\right],\\
\zeta^{(1)} &= -I(\theta_0)^{-1}\left[2\int_0^T \partial\delta B(X_t^0, \theta_0) D_t\,\mathrm{d}w_t + 2\int_0^T \delta^2 B(X_t^0, \theta_0)\,\mathrm{d}w_t \cdot \zeta^{(0)} + \right.\\
&\quad + \int_0^T \partial^2\delta C(X_t^0, \theta_0)(D_t, D_t)\,\mathrm{d}t + 2\int_0^T \partial\delta^2 C(X_t^0, \theta_0) D_t\,\mathrm{d}t \cdot \zeta^{(0)} + \\
&\quad \left. + \int_0^T \partial\delta C(X_t^0, \theta_0) E_t\,\mathrm{d}t + \int_0^T \delta^3 C(X_t^0, \theta_0)\,\mathrm{d}t \cdot (\zeta^{(0)})^{\otimes 2}\right].
\end{aligned}$$

Hence, in general, we may suppose the following assumption for the estimator $\hat{\theta}_\varepsilon$.

ASSUMPTION 3. There exists $R_\varepsilon = O(1)$ in D^∞ such that for every $K > 0$ and $c > 0$,

$$\lim_{\varepsilon \downarrow 0} \varepsilon^{-K} P[|R_\varepsilon| > c] = 0,$$

and for $\psi_\varepsilon = \psi(3R_\varepsilon)$,

$$\psi_\varepsilon \varepsilon^{-1}(\hat{\theta}_\varepsilon - \theta_0) \sim \zeta^{(0)} + \tfrac{1}{2}\varepsilon\zeta^{(1)} + \cdots$$

in $D^\infty(\mathbf{R}^p)$ as $\varepsilon \downarrow 0$ with $\zeta^{(0)}, \zeta^{(1)}, \ldots$ in $D^\infty(\mathbf{R}^p)$, where $\zeta^{(0)} = \int_0^T g_t\,\mathrm{d}w_t$ for some function $g \in L^2([0, T], \mathrm{d}t)$.

We then have the following lemma.

LEMMA 5. *Let* $\bar{\psi}_\varepsilon = \psi(9R_\varepsilon)$. *Suppose that Assumption* 3 *holds true. Then* $\bar{\psi}_\varepsilon \bar{S}_\varepsilon^* \in D^\infty$ *has the asymptotic expansion*

$$\bar{\psi}_\varepsilon \bar{S}_\varepsilon^* \sim f_0 + \varepsilon f_1 + \cdots$$

in D^∞ *as* $\varepsilon \downarrow 0$ *with* $f_0, f_1, \ldots \in D^\infty$. *In particular*,

$$\begin{aligned}
f_0 &= \int_0^T b(X_t^0, \theta_0)\,\mathrm{d}w_t + \int_0^T \partial c(X_t^0, \theta_0) D_t\,\mathrm{d}t + \zeta^{(0)\prime} \int_0^T \delta c(X_t^0, \theta_0)\,\mathrm{d}t,\\
f_1 &= F_1 - b(\theta_0),\\
F_1 &= \int_0^T \partial b(X_t^0, \theta_0) D_t\,\mathrm{d}w_t + \zeta^{(0)\prime} \int_0^T \delta b(X_t^0, \theta_0)\,\mathrm{d}w_t +\\
&\quad + \zeta^{(0)\prime} \int_0^T \partial\delta c(X_t^0, \theta_0) D_t\,\mathrm{d}t + \frac{1}{2} P_{\mathbf{R}1^\perp}[\zeta^{(1)\prime}] \int_0^T \delta c(X_t^0, \theta_0)\,\mathrm{d}t +\\
&\quad + \frac{1}{2}\int_0^T \partial^2 c(X_t^0, \theta_0) P_{\mathbf{R}1^\perp}[(D_t, D_t)]\,\mathrm{d}t +\\
&\quad + \frac{1}{2}\int_0^T \partial c(X_t^0, \theta_0) P_{\mathbf{R}1^\perp}[E_t]\,\mathrm{d}t +\\
&\quad + \frac{1}{2} P_{\mathbf{R}1^\perp}\left[(\zeta^{(0)})' \left(\int_0^T \delta^2 c(X_t^0, \theta_0)\,\mathrm{d}t\right) \zeta^{(0)}\right].
\end{aligned}$$

Let $a_t^* = b(X_t^0, \theta_0) + \sum_{\alpha=1}^d \int_t^T \partial_\alpha c(X_s^0, \theta_0) \lambda_{s,t}^\alpha\,\mathrm{d}s + \sum_{i=1}^p \int_0^T \delta_i c(X_t^0, \theta_0)\,\mathrm{d}t\, g_t^i$. We then have $f_0 = \int_0^T a_t^*\,\mathrm{d}w_t$ and $\sigma_{f_0} = \int_0^T a_t^* (a_t^*)'\,\mathrm{d}t$.

ASSUMPTION 4. There exists $t \in [0, T]$ such that $a_t^* \neq 0$.

Under the nondegeneracy of f_0, we obtain a stochastic expansion of the pull-back of a Schwartz distribution under $\bar{S}_\varepsilon^*$ in a space of generalized Wiener functionals.

LEMMA 6. *Let* $\xi_\varepsilon = 2(\sigma_{\bar{\psi}_\varepsilon \bar{S}_\varepsilon^*} - \sigma_{f_0})/\sigma_{f_0}$ *and* $\psi_\varepsilon^* = \psi(\xi_\varepsilon)$. *Suppose that Assumptions* 3 *and* 4 *hold true. Then for any* $T(x) \in S'(\mathbf{R})$, $\psi_\varepsilon^* T(\bar{\psi}_\varepsilon \bar{S}_\varepsilon^*) \in \tilde{D}^{-\infty}$ *has the asymptotic expansion*

$$\psi_\varepsilon^* T(\bar{\psi}_\varepsilon \bar{S}_\varepsilon^*) \sim \Phi_0 + \varepsilon \Phi_1 + \cdots$$

in $\tilde{D}^{-\infty}$ as $\varepsilon \downarrow 0$ *uniformly in every class* $\{T\}$ *satisfying the condition* (5) *of Theorem* 4, *and* $\Phi_0, \Phi_1, \ldots$ in $\tilde{D}^{-\infty}$ *are determined by the formal Taylor expansion. In particular*,

$$\Phi_0 = T(f_0), \qquad \Phi_1 = f_1 \cdot \partial T(f_0).$$

Finally, we obtain the asymptotic expansion of $\bar{S}_\varepsilon^*$.

THEOREM 5. *Let* $M, \gamma > 0$. *Suppose that Assumptions* 3 *and* 4 *hold true. Then for any* $f \in \mathcal{E}(M, \gamma)$,

$$E[f(\bar{S}_\varepsilon^*)] \sim \int_{\mathbf{R}} f(x) p_0(x)\,\mathrm{d}x + \varepsilon \int_{\mathbf{R}} f(x) p_1(x)\,\mathrm{d}x + \cdots$$

as $\varepsilon \downarrow 0$ *uniformly in* $f \in \mathcal{E}(M, \gamma)$. *In particular,*

$$p_0(x) \equiv p_{f_0}(x) = \phi(x; \Sigma), \qquad p_1(x) = -\partial_x\{E[f_1|f_0 = x]p_{f_0}(x)\}.$$

Remark 4. In Theorem 5, if f is a smooth function, there is no need to suppose Assumption 4.

From Theorem 5, we have the following three theorems.

THEOREM 6 (f-unbiased information criterion). *Let* $M, \gamma > 0$. *Suppose that Assumptions* 3 *and* 4 *hold true. For any* $f \in \mathcal{E}'(M, \gamma)$, let $IC_f(X; \hat{\theta}_\epsilon) = \varepsilon l_\varepsilon(X, \hat{\theta}_\varepsilon) - \varepsilon b_f(\hat{\theta}_\varepsilon)$, *where*

$$b_f(\theta_0) \;=\; \frac{\int_{\mathbf{R}} f(x)\partial_x\{E[F_1|f_0 = x]\phi(x; \Sigma)\}\,\mathrm{d}x}{\int_{\mathbf{R}} f(x)\partial_x\{\phi(x; \Sigma)\}\,\mathrm{d}x}.$$

Then, IC_f *is the* f*-unbiased information criterion for* $f \in \mathcal{E}'(M, \gamma)$, *that is,*

$$E\left[f(IC_f(X; \hat{\theta}_\epsilon) - \varepsilon \int_W \int_W l_\varepsilon(X^\varepsilon(w'), \hat{\theta}_\varepsilon(w))P(\mathrm{d}w')P(\mathrm{d}w))\right] = o(\varepsilon).$$

THEOREM 7 (EUIC). *Suppose that Assumption* 3 *holds true. Let* $IC_1(X; \hat{\theta}_\epsilon) = \varepsilon l_\varepsilon(X, \hat{\theta}_\varepsilon) - \varepsilon b_1(\hat{\theta}_\varepsilon)$, *where*

$$b_1(\theta_0) = E\left[(\zeta^{(0)})'\left(\int_0^T \delta b(X_t^0, \theta_0)\,\mathrm{d}w_t + \int_0^T \partial\delta c(X_t^0, \theta_0)D_t\,\mathrm{d}t\right)\right].$$

Then

$$E\left[IC_1(X; \hat{\theta}_\epsilon) - \varepsilon \int_W \int_W l_\varepsilon(X^\varepsilon(w'), \hat{\theta}_\varepsilon(w))P(\mathrm{d}w')P(\mathrm{d}w)\right] = o(\varepsilon).$$

THEOREM 8 (MUIC). *Suppose that Assumptions* 3 *and* 4 *hold true. Let* $IC_2(X; \hat{\theta}_\epsilon) = \varepsilon l_\varepsilon(X, \hat{\theta}_\varepsilon) - \varepsilon b_2(\hat{\theta}_\varepsilon)$, *where* $b_2(\theta_0) = E[F_1|f_0 = 0]$. *Then*

$$P\left[IC_2(X; \hat{\theta}_\epsilon) - \varepsilon \int_W \int_W l_\varepsilon(X^\varepsilon(w'), \hat{\theta}_\varepsilon(w))P(\mathrm{d}w')P(\mathrm{d}w) < 0\right]$$
$$= \frac{1}{2} + o(\varepsilon),$$

and

$$P\left[IC_2(X; \hat{\theta}_\epsilon) - \varepsilon \int_W \int_W l_\varepsilon(X^\varepsilon(w'), \hat{\theta}_\varepsilon(w))P(\mathrm{d}w')P(\mathrm{d}w) > 0\right]$$
$$= \frac{1}{2} + o(\varepsilon).$$

6. Proofs

To begin with, we will prove the general results of Section 5 because Theorems 1 and 2 of Section 3 are special cases of Section 5.

Proof of Lemma 4. In Section 3, in order to consider the M-estimator, the functional $G(X;\varepsilon,\theta) = \varepsilon^2\{\Psi_\varepsilon(X;\theta) - \Psi_\varepsilon(X;\theta_0)\}$ is given explicitly. From the smoothness of $G(X;\varepsilon,\theta)$ and coefficients in the stochastic differential equation (1), we can show that $G(X;\varepsilon,\theta)$ satisfies the following conditions:

(C1) The functional $G(\cdot;\cdot,\cdot) : W \to C([0,1]\times\Theta \to \mathbf{R})$ is smooth, where $G(X;0,\theta) = \lim_{\varepsilon\downarrow 0} G(X;\varepsilon,\theta)$.

(C2) For each $\theta \in \Theta$, $G(X;0,\theta)$ is deterministic ($G(0,\theta)$, say), and there exists $a_0 > 0$ such that $-G(0,\theta) \geqslant a_0|\theta-\theta_0|^2$ for any $\theta \in \Theta$.

(C3) There exist $h^{(i)} \in H$, $i = 1,2,\ldots,p$, such that

$$\delta_\varepsilon\delta_i G(X;0,\theta_0) = \int_0^\infty \dot{h}_s^{(i)}\cdot \mathrm{d}w_s,$$

for $i = 1,\ldots,p$, where $\delta_\varepsilon = \partial/\partial\varepsilon$.

See [44] for smoothness in (C1). For M-estimator $\hat{\theta}_\varepsilon$, in the same way as [44, 45], it follows from the conditions (C1)–(C3) that there exists $R_\varepsilon = O(1)$ in D^∞ such that for every $K > 0$ and $c > 0$,

$$\psi(3R_\varepsilon)\varepsilon^{-1}(\hat{\theta}_\varepsilon - \theta_0) \sim \zeta^{(0)} + \tfrac{1}{2}\varepsilon\zeta^{(1)} + \cdots$$

in $D^\infty(\mathbf{R}^p)$ as $\varepsilon \downarrow 0$ with $\zeta^{(0)},\zeta^{(1)},\ldots$ in $D^\infty(\mathbf{R}^p)$, where $\lim_{\varepsilon\downarrow 0}\varepsilon^{-K}P[|R_\varepsilon| > c] = 0$.

From (7), we have

$$\begin{aligned} G(X,\varepsilon,\theta) \;=\; & \varepsilon\int_0^T \{B(X_t,\theta) - B(X_t,\theta_0)\}\,\mathrm{d}w_t + \\ & + \int_0^T \{C(X_t,\theta) - C(X_t,\theta_0)\}\,\mathrm{d}t. \end{aligned}$$

From the results in [44, 45], it is enough for the proof of the first statement to show that it follows from Assumption 1 hat $G(X;\varepsilon,\theta)$ satisfies the conditions (C1)–(C3).

From Assumption 1, as in the last part of Section 3 of Yoshida [44], we see that $G(X;\cdot,\cdot)$ is smooth in the sense of Malliavin calculus. $G(0,\theta) = \int_0^T\{C(X_t^0,\theta) - C(X_t^0,\theta_0)\}\,\mathrm{d}t$ is deterministic. From (iv) in Assumption 1, there exists $a_0 > 0$ such that $-G(0,\theta) \geqslant a_0|\theta-\theta_0|^2$. It follows from Itô's formula that

$$\delta_\varepsilon\delta_i G(X;0,\theta_0) = \int_0^T \dot{h}_s^{(i)}\,\mathrm{d}w_s,$$

where

$$\dot{h}_s^{(i)} = \delta_i B(X_s^0,\theta_0) + \int_s^T \delta_i\partial C(X_u^0,\theta_0)Y_uY_s^{-1}V(X_s^0)\,\mathrm{d}u.$$

We then obtain $h^{(i)} \in H$. This completes the proof of the first statement.

Next, we compute $\zeta^{(0)}$, $\zeta^{(1)}$ in the same way as Lemma 7.3 in [45]. The following is formal computation but for details of rigorous proof, see Lemma 7.3 in [45]. Let

$$\begin{aligned} F(\theta, \varepsilon) &:= \varepsilon^2 \delta \Psi_\varepsilon(X, \theta) \\ &= \varepsilon \int_0^T \delta B(X_t, \theta)\, \mathrm{d}w_t + \int_0^T \delta C(X_t, \theta)\, \mathrm{d}t. \end{aligned}$$

Since it follows from the definition of $\hat{\theta}_\varepsilon$ that $F(\hat{\theta}_\varepsilon, \varepsilon) = 0$,

$$\begin{aligned} 0 &= \delta_\varepsilon F(\hat{\theta}_\varepsilon, \varepsilon) \\ &= \delta F(\theta, \varepsilon)|_{\theta=\hat{\theta}_\varepsilon} \delta_\varepsilon \hat{\theta}_\varepsilon + \delta_\varepsilon F(\theta, \varepsilon)|_{\theta=\hat{\theta}_\varepsilon}, \end{aligned}$$

where

$$\begin{aligned} \delta F(\theta, \varepsilon) &= \varepsilon \int_0^T \delta^2 B(X_t, \theta)\, \mathrm{d}w_t + \int_0^T \delta^2 C(X_t, \theta)\, \mathrm{d}t, \\ \delta_\varepsilon F(\theta, \varepsilon) &= \int_0^T \delta B(X_t, \theta)\, \mathrm{d}w_t + \varepsilon \int_0^T \partial \delta B(X_t, \theta) \delta_\varepsilon X_t\, \mathrm{d}w_t + \\ &\quad + \int_0^T \partial \delta C(X_t, \theta) \delta_\varepsilon X_t\, \mathrm{d}t. \end{aligned}$$

Since $\zeta^{(0)} = \lim_{\varepsilon \downarrow 0} \delta_\varepsilon \hat{\theta}_\varepsilon$, we obtain the result.

In the same way as above,

$$\begin{aligned} 0 &= \delta_\varepsilon^2 F(\hat{\theta}_\varepsilon, \varepsilon) \\ &= \delta_\varepsilon \left(\delta F(\theta, \varepsilon)|_{\theta=\hat{\theta}_\varepsilon} \delta_\varepsilon \hat{\theta}_\varepsilon + \delta_\varepsilon F(\theta, \varepsilon)|_{\theta=\hat{\theta}_\varepsilon} \right) \\ &= \delta^2 F(\theta, \varepsilon)|_{\theta=\hat{\theta}_\varepsilon} (\delta_\varepsilon \hat{\theta}_\varepsilon)^{\otimes 2} + 2\delta_\varepsilon \delta F(\theta, \varepsilon)|_{\theta=\hat{\theta}_\varepsilon} \delta_\varepsilon \hat{\theta}_\varepsilon + \\ &\quad + \delta F(\theta, \varepsilon)|_{\theta=\hat{\theta}_\varepsilon} \delta_\varepsilon^2 \hat{\theta}_\varepsilon + \delta_\varepsilon^2 F(\theta, \varepsilon)|_{\theta=\hat{\theta}_\varepsilon}, \end{aligned}$$

where

$$\begin{aligned} \delta^2 F(\theta, \varepsilon) &= \varepsilon \int_0^T \delta^3 B(X_t, \theta)\, \mathrm{d}w_t + \int_0^T \delta^3 C(X_t, \theta)\, \mathrm{d}t, \\ \delta_\varepsilon \delta F(\theta, \varepsilon) &= \int_0^T \delta^2 B(X_t, \theta)\, \mathrm{d}w_t + \varepsilon \int_0^T \partial \delta^2 B(X_t, \theta) \delta_\varepsilon X_t\, \mathrm{d}w_t + \\ &\quad + \int_0^T \partial \delta^2 C(X_t, \theta) \delta_\varepsilon X_t\, \mathrm{d}t, \end{aligned}$$

$$\begin{aligned}\delta_\varepsilon^2 F(\theta,\varepsilon) \;=\; & 2\int_0^T \partial\delta B(X_t,\theta)\delta_\varepsilon X_t \,\mathrm{d}w_t \\ & +\varepsilon\int_0^T \partial^2\delta B(X_t,\theta)(\delta_\varepsilon X_t)^{\otimes 2}\,\mathrm{d}w_t + \\ & +\varepsilon\int_0^T \partial\delta B(X_t,\theta)\delta_\varepsilon^2 X_t\,\mathrm{d}w_t + \\ & +\int_0^T \partial^2\delta C(X_t,\theta)(\delta_\varepsilon X_t)^{\otimes 2}\,\mathrm{d}t + \int_0^T \partial\delta C(X_t,\theta)\delta_\varepsilon^2 X_t\,\mathrm{d}t.\end{aligned}$$

Thus, we obtain $\zeta^{(1)} = \lim_{\varepsilon\downarrow 0}\delta_\varepsilon^2\hat{\theta}_\varepsilon$. This completes the proof. □

Proof of Lemma 5. From Assumption 3, we see that

$$\psi_\varepsilon\varepsilon^{-1}(\hat{\theta}_\varepsilon - \theta_0) = \zeta^{(0)} + \frac{1}{2}\varepsilon\zeta^{(1)} + \cdots + \frac{1}{k!}\varepsilon^{k-1}\zeta^{(k-1)} + \varepsilon^k\zeta_\varepsilon^{(k)},$$

where $\zeta_\varepsilon^{(k)} = O(1)$ in $D^\infty(\mathbf{R}^p)$. Moreover, in case that $|R_\varepsilon| < 1/9$, it follows from the definition of ψ_ε that $\psi_\varepsilon = \psi(3R_\varepsilon) = 1$ and

$$\begin{aligned}1\cdot\varepsilon^{-1}(\hat{\theta}_\varepsilon - \theta_0) \;=\; & \psi_\varepsilon\varepsilon^{-1}(\hat{\theta}_\varepsilon - \theta_0) \\ =\; & \zeta^{(0)} + \frac{1}{2}\varepsilon\zeta^{(1)} + \cdots + \frac{1}{k!}\varepsilon^{k-1}\zeta^{(k-1)} + \varepsilon^k\zeta_\varepsilon^{(k)}. \qquad (12)\end{aligned}$$

First, expanding $\bar{\psi}_\varepsilon\varepsilon l_\varepsilon(X^\varepsilon(w),\hat{\theta}_\varepsilon(w))$ in a Taylor series around θ_0 and substituting (12) in the resulting expansion, we obtain the second-order stochastic expansion as follows:

$$\begin{aligned}&\bar{\psi}_\varepsilon\varepsilon l_\varepsilon(X^\varepsilon(w),\hat{\theta}_\varepsilon(w)) \\ &\quad = \bar{\psi}_\varepsilon\Big[\varepsilon l_\varepsilon(X^\varepsilon(w),\theta_0) + \varepsilon(\delta l_\varepsilon(X^\varepsilon(w),\theta_0))'(\hat{\theta}_\varepsilon(w)-\theta_0) + \\ &\qquad + \frac{1}{2}\varepsilon(\hat{\theta}_\varepsilon(w)-\theta_0)'(\delta^2 l_\varepsilon(X^\varepsilon(w),\theta_0))(\hat{\theta}_\varepsilon(w)-\theta_0) + \\ &\qquad + \frac{1}{2}\varepsilon\int_0^1 \delta^3 l_\varepsilon(X^\varepsilon,\theta_0+u(\hat{\theta}_\varepsilon(w)-\theta_0))(1-u)^2\,\mathrm{d}u(\hat{\theta}_\varepsilon(w)-\theta_0)^{\otimes 3}\Big] \\ &\quad = \bar{\psi}_\varepsilon\Big[\int_0^T b(X_t^\varepsilon,\theta_0)\,\mathrm{d}w_t + \varepsilon^{-1}\int_0^T c(X_t^\varepsilon,\theta_0)\,\mathrm{d}t + \\ &\qquad + \Big(\int_0^T \delta b(X_t^\varepsilon,\theta_0)\,\mathrm{d}w_t + \varepsilon^{-1}\int_0^T \delta c(X_t^\varepsilon,\theta_0)\,\mathrm{d}t\Big)'\times\end{aligned}$$

$$
\begin{aligned}
&\times\left(\varepsilon\zeta^{(0)}(w)+\frac{1}{2}\varepsilon^2\zeta^{(1)}(w)+\varepsilon^3\zeta_\varepsilon^{(2)}(w)\right)+\\
&+\frac{1}{2}\left(\int_0^T\delta^2 b(X_t^\varepsilon,\theta_0)\,\mathrm{d}w_t+\varepsilon^{-1}\int_0^T\delta^2 c(X_t^\varepsilon,\theta_0)\,\mathrm{d}t\right)\times\\
&\times\left(\varepsilon\zeta^{(0)}(w)+\frac{1}{2}\varepsilon^2\zeta^{(1)}(w)+\varepsilon^3\zeta_\varepsilon^{(2)}(w)\right)^{\otimes 2}+\\
&+\frac{1}{2}\int_0^1\left(\int_0^T\delta^3 b(X_t^\varepsilon,\theta)\,\mathrm{d}w_t\bigg|_{\theta=\theta_0+u(\hat{\theta}_\varepsilon(w)-\theta_0)}+\right.\\
&\left.+\varepsilon^{-1}\int_0^T\delta^3 c(X_t^\varepsilon,\theta_0+u(\hat{\theta}_\varepsilon(w)-\theta_0))\,\mathrm{d}t\right)\times\\
&\left.\times(1-u)^2\,\mathrm{d}u\left(\varepsilon\zeta^{(0)}(w)+\frac{1}{2}\varepsilon^2\zeta^{(1)}(w)+\varepsilon^3\zeta_\varepsilon^{(2)}(w)\right)^{\otimes 3}\right]\\
={}&\bar{\psi}_\varepsilon\left[\varepsilon^{-1}\int_0^T c(X_t^\varepsilon,\theta_0)\,\mathrm{d}t+\int_0^T b(X_t^\varepsilon,\theta_0)\,\mathrm{d}w_t+\right.\\
&+\left(\int_0^T\delta c(X_t^\varepsilon,\theta_0)\,\mathrm{d}t\right)'\zeta^{(0)}(w)+\varepsilon\left(\int_0^T\delta b(X_t^\varepsilon,\theta_0)\,\mathrm{d}w_t\right)'\zeta^{(0)}(w)+\\
&+\frac{1}{2}\varepsilon\left(\int_0^T\delta c(X_t^\varepsilon,\theta_0)\,\mathrm{d}t\right)'\zeta^{(1)}(w)+\\
&\left.+\frac{1}{2}\varepsilon\left(\int_0^T\delta^2 c(X_t^\varepsilon,\theta_0)\,\mathrm{d}t\right)(\zeta^{(0)}(w))^{\otimes 2}+\varepsilon^2 f_1^{(2)}(X_t^\varepsilon,\theta_0)\right],
\end{aligned}
$$

where

$$
\begin{aligned}
&f_1^{(2)}(X_t^\varepsilon,\theta_0)\\
&\quad=\frac{1}{2}\left(\int_0^T\delta b(X_t^\varepsilon,\theta_0)\,\mathrm{d}w_t\right)'\zeta^{(1)}(w)+\varepsilon\left(\int_0^T\delta b(X_t^\varepsilon,\theta_0)\,\mathrm{d}w_t\right)'\zeta_\varepsilon^{(2)}(w)+\\
&\quad+\left(\int_0^T\delta c(X_t^\varepsilon,\theta_0)\,\mathrm{d}t\right)'\zeta_\varepsilon^{(2)}(w)+\frac{1}{2}\left(\int_0^T\delta^2 b(X_t^\varepsilon,\theta_0)\,\mathrm{d}w_t\right)\times\\
&\quad\times(\zeta^{(0)}(w))^{\otimes 2}+\frac{1}{2}\left(\varepsilon\int_0^T\delta^2 b(X_t^\varepsilon,\theta_0)\,\mathrm{d}w_t+\int_0^T\delta^2 c(X_t^\varepsilon,\theta_0)\,\mathrm{d}t\right)\times\\
&\quad\times\varepsilon\left(\frac{1}{2}\zeta^{(1)}(w)+\varepsilon\zeta_\varepsilon^{(2)}(w)\right)^{\otimes 2}+(\zeta^{(0)}(w))'\times\\
&\quad\times\left(\varepsilon\int_0^T\delta^2 b(X_t^\varepsilon,\theta_0)\,\mathrm{d}w_t+\int_0^T\delta^2 c(X_t^\varepsilon,\theta_0)\,\mathrm{d}t\right)\times\\
&\quad\times\left(\frac{1}{2}\zeta^{(1)}(w)+\varepsilon\zeta_\varepsilon^{(2)}(w)\right)+
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{1}{2}\int_0^1\left(\varepsilon\int_0^T\delta^3 b(X_t^\varepsilon,\theta)\,\mathrm{d}w_t\Big|_{\theta=\theta_0+u(\hat{\theta}_\varepsilon(w)-\theta_0)}+\right.\\
&+\left.\int_0^T\delta^3 c(X_t^\varepsilon,\theta_0+u(\hat{\theta}_\varepsilon(w)-\theta_0))\,\mathrm{d}t\right)\times\\
&\times(1-u)^2\,\mathrm{d}u\left(\zeta^{(0)}(w)+\frac{1}{2}\varepsilon\zeta^{(1)}(w)+\varepsilon^2\zeta_\varepsilon^{(2)}(w)\right)^{\otimes 3}.
\end{aligned}
$$

Since $\bar{\psi}_\varepsilon=1-O_M(\varepsilon^K)$ for any $K>0$, where $O_M(\varepsilon^K)$ means $O(\varepsilon^K)$ in D^∞, we have

$$
\begin{aligned}
&\bar{\psi}_\varepsilon\int_0^1\int_0^T\delta^3 b(X_t^\varepsilon,\theta)\,\mathrm{d}w_t\Big|_{\theta=\theta_0+u(\hat{\theta}_\varepsilon(w)-\theta_0)}(1-u)^2\,\mathrm{d}u=O_M(1),\\
&\bar{\psi}_\varepsilon\int_0^1\int_0^T\delta^3 c(X_t^\varepsilon,\theta_0+u(\hat{\theta}_\varepsilon(w)-\theta_0))\,\mathrm{d}t\,(1-u)^2\,\mathrm{d}u=O_M(1).
\end{aligned}
$$

Therefore, we obtain $\bar{\psi}_\varepsilon f_1^{(2)}(X_t^\varepsilon,\theta_0)=O_M(1)$ and

$$
\begin{aligned}
&\bar{\psi}_\varepsilon\varepsilon l_\varepsilon(X^\varepsilon(w),\hat{\theta}_\varepsilon(w))\\
&\quad=\bar{\psi}_\varepsilon\left[\varepsilon^{-1}\int_0^T c(X_t^\varepsilon,\theta_0)\,\mathrm{d}t+\int_0^T b(X_t^\varepsilon,\theta_0)\,\mathrm{d}w_t+\right.\\
&\quad+\left(\int_0^T\delta c(X_t^\varepsilon,\theta_0)\,\mathrm{d}t\right)'\zeta^{(0)}(w)+\varepsilon\left(\int_0^T\delta b(X_t^\varepsilon,\theta_0)\,\mathrm{d}w_t\right)'\zeta^{(0)}(w)+\\
&\quad+\frac{1}{2}\varepsilon\left(\int_0^T\delta c(X_t^\varepsilon,\theta_0)\,\mathrm{d}t\right)'\zeta^{(1)}(w)+\frac{1}{2}\varepsilon\left(\int_0^T\delta^2 c(X_t^\varepsilon,\theta_0)\,\mathrm{d}t\right)\times\\
&\quad\times\left.(\zeta^{(0)}(w))^{\otimes 2}\right]+O_M(\varepsilon^2).
\end{aligned}
$$

Next, expanding X_t^ε in a Taylor series around X_t^0, we obtain

$$
\begin{aligned}
&\bar{\psi}_\varepsilon\varepsilon l_\varepsilon(X^\varepsilon(w),\hat{\theta}_\varepsilon(w))\\
&\quad=\bar{\psi}_\varepsilon\left[\varepsilon^{-1}\left(\int_0^T c(X_t^0,\theta_0)\,\mathrm{d}t+\varepsilon\int_0^T\partial c(X_t^0,\theta_0)D_t\,\mathrm{d}t+\right.\right.\\
&\quad+\frac{1}{2}\varepsilon^2\int_0^T\partial^2 c(X_t^0,\theta_0)(D_t,D_t)\,\mathrm{d}t+\frac{1}{2}\varepsilon^2\int_0^T\partial c(X_t^0,\theta_0)E_t\,\mathrm{d}t+\\
&\quad+\left.\frac{1}{2}\varepsilon^3\int_0^1\int_0^T\delta_\varepsilon^3 c(X_t^\varepsilon,\theta_0)\,\mathrm{d}t\Big|_{\varepsilon\leftarrow u\varepsilon}(1-u)^2\,\mathrm{d}u\right)+\\
&\quad+\left(\int_0^T b(X_t^0,\theta_0)\,\mathrm{d}w_t+\varepsilon\int_0^T\partial b(X_t^0,\theta_0)D_t\,\mathrm{d}w_t+\right.
\end{aligned}
$$

$$
\begin{aligned}
& + \varepsilon^2 \int_0^1 \int_0^T \delta_\varepsilon^2 b(X_t^\varepsilon, \theta_0)\, \mathrm{d}w_t \Big|_{\varepsilon \leftarrow u\varepsilon} (1-u)\, \mathrm{d}u \bigg) + \\
& + \bigg(\int_0^T \delta c(X_t^0, \theta_0)\, \mathrm{d}t + \varepsilon \int_0^T \partial \delta c(X_t^0, \theta_0) D_t\, \mathrm{d}t + \\
& + \varepsilon^2 \int_0^1 \int_0^T \delta_\varepsilon^2 \delta c(X_t^\varepsilon, \theta_0)\, \mathrm{d}t \Big|_{\varepsilon \leftarrow u\varepsilon} (1-u)\, \mathrm{d}u \bigg)' \zeta^{(0)}(w) + \\
& + \varepsilon \bigg(\int_0^T \delta b(X_t^0, \theta_0)\, \mathrm{d}w_t + \varepsilon \int_0^1 \int_0^T \delta_\varepsilon \delta b(X_t^\varepsilon, \theta_0)\, \mathrm{d}w_t \Big|_{\varepsilon \leftarrow u\varepsilon} \mathrm{d}u \bigg)' \zeta^{(0)}(w) + \\
& + \frac{1}{2} \varepsilon \bigg(\int_0^T \delta c(X_t^0, \theta_0)\, \mathrm{d}t + \varepsilon \int_0^1 \int_0^T \delta_\varepsilon \delta c(X_t^\varepsilon, \theta_0)\, \mathrm{d}t \Big|_{\varepsilon \leftarrow u\varepsilon} \mathrm{d}u \bigg)' \zeta^{(1)}(w) + \\
& + \frac{1}{2} \varepsilon (\zeta^{(0)})' \bigg(\int_0^T \delta^2 c(X_t^0, \theta_0)\, \mathrm{d}t + \\
& + \varepsilon \int_0^1 \int_0^T \delta_\varepsilon \delta^2 c(X_t^\varepsilon, \theta_0)\, \mathrm{d}t \Big|_{\varepsilon \leftarrow u\varepsilon} \mathrm{d}u \bigg) \zeta^{(0)}(w) \bigg] + O_M(\varepsilon^2) \\
= \; & \bar{\psi}_\varepsilon \bigg[\varepsilon^{-1} \int_0^T c(X_t^0, \theta_0)\, \mathrm{d}t + \\
& + \int_0^T b(X_t^0, \theta_0)\, \mathrm{d}w_t + \int_0^T \partial c(X_t^0, \theta_0) D_t\, \mathrm{d}t + \\
& + \bigg(\int_0^T \delta c(X_t^0, \theta_0)\, \mathrm{d}t \bigg)' \zeta^{(0)}(w) + \varepsilon \int_0^T \partial b(X_t^0, \theta_0) D_t\, \mathrm{d}w_t + \\
& + \varepsilon \bigg(\int_0^T \delta b(X_t^0, \theta_0)\, \mathrm{d}w_t \bigg)' \zeta^{(0)}(w) + \varepsilon \bigg(\int_0^T \partial \delta c(X_t^0, \theta_0) D_t\, \mathrm{d}t \bigg)' \zeta^{(0)}(w) + \\
& + \frac{1}{2} \varepsilon \bigg(\int_0^T \delta c(X_t^0, \theta_0)\, \mathrm{d}t \bigg)' \zeta^{(1)}(w) + \frac{1}{2} \varepsilon \int_0^T \partial^2 c(X_t^0, \theta_0)(D_t, D_t)\, \mathrm{d}t + \\
& + \frac{1}{2} \varepsilon \int_0^T \partial c(X_t^0, \theta_0) E_t\, \mathrm{d}t + \frac{1}{2} \varepsilon (\zeta^{(0)}(w))' \times \\
& \times \bigg(\int_0^T \delta^2 c(X_t^0, \theta_0)\, \mathrm{d}t \bigg) \zeta^{(0)}(w) + \varepsilon^2 f_2^{(2)}(X_t^\varepsilon, \theta_0) \bigg] + \\
& + O_M(\varepsilon^2),
\end{aligned}
$$

where

$$
\begin{aligned}
f_2^{(2)}(X_t^\varepsilon, \theta_0) \; = \; & \frac{1}{2} \int_0^1 \int_0^T \delta_\varepsilon^3 c(X_t^\varepsilon, \theta_0)\, \mathrm{d}t \Big|_{\varepsilon \leftarrow u\varepsilon} (1-u)^2\, \mathrm{d}u + \\
& + \int_0^1 \int_0^T \delta_\varepsilon^2 b(X_t^\varepsilon, \theta_0)\, \mathrm{d}w_t \Big|_{\varepsilon \leftarrow u\varepsilon} (1-u)\, \mathrm{d}u + \\
& + \bigg(\int_0^1 \int_0^T \delta_\varepsilon^2 \delta c(X_t^\varepsilon, \theta_0)\, \mathrm{d}t \Big|_{\varepsilon \leftarrow u\varepsilon} (1-u)\, \mathrm{d}u \bigg)' \zeta^{(0)}(w) +
\end{aligned}
$$

$$+ \left(\int_0^1 \int_0^T \delta_\varepsilon \delta b(X_t^\varepsilon, \theta_0) \, \mathrm{d}w_t \bigg|_{\varepsilon \leftarrow u\varepsilon} \mathrm{d}u \right)' \zeta^{(0)}(w) +$$
$$+ \frac{1}{2} \left(\int_0^1 \int_0^T \delta_\varepsilon \delta c(X_t^\varepsilon, \theta_0) \, \mathrm{d}t \bigg|_{\varepsilon \leftarrow u\varepsilon} \mathrm{d}u \right)' \zeta^{(1)}(w) +$$
$$+ \frac{1}{2} (\zeta^{(0)})' \left(\int_0^1 \int_0^T \delta_\varepsilon \delta^2 c(X_t^\varepsilon, \theta_0) \, \mathrm{d}t \bigg|_{\varepsilon \leftarrow u\varepsilon} \mathrm{d}u \right) \zeta^{(0)}(w).$$

Since $\bar{\psi}_\varepsilon = 1 - O_M(\varepsilon^K)$ for any $K > 0$, we have

$$\bar{\psi}_\varepsilon \int_0^1 \int_0^T \delta_\varepsilon \delta b(X_t^\varepsilon, \theta_0) \, \mathrm{d}w_t \bigg|_{\varepsilon \leftarrow u\varepsilon} \mathrm{d}u = O_M(1),$$
$$\bar{\psi}_\varepsilon \int_0^1 \int_0^T \delta_\varepsilon^2 b(X_t^\varepsilon, \theta_0) \, \mathrm{d}w_t \bigg|_{\varepsilon \leftarrow u\varepsilon} (1-u) \, \mathrm{d}u = O_M(1),$$
$$\bar{\psi}_\varepsilon \int_0^1 \int_0^T \delta_\varepsilon \delta c(X_t^\varepsilon, \theta_0) \, \mathrm{d}t \bigg|_{\varepsilon \leftarrow u\varepsilon} \mathrm{d}u = O_M(1),$$
$$\bar{\psi}_\varepsilon \int_0^1 \int_0^T \delta_\varepsilon \delta^2 c(X_t^\varepsilon, \theta_0) \, \mathrm{d}t \bigg|_{\varepsilon \leftarrow u\varepsilon} \mathrm{d}u = O_M(1),$$
$$\bar{\psi}_\varepsilon \int_0^1 \int_0^T \delta_\varepsilon^2 \delta c(X_t^\varepsilon, \theta_0) \, \mathrm{d}t \bigg|_{\varepsilon \leftarrow u\varepsilon} (1-u) \, \mathrm{d}u = O_M(1),$$
$$\bar{\psi}_\varepsilon \int_0^1 \int_0^T \delta_\varepsilon^3 c(X_t^\varepsilon, \theta_0) \, \mathrm{d}t \bigg|_{\varepsilon \leftarrow u\varepsilon} (1-u)^2 \, \mathrm{d}u = O_M(1).$$

It then follows that $\bar{\psi}_\varepsilon f_2^{(2)}(X_t^\varepsilon, \theta_0) = O_M(1)$ and

$$\begin{aligned}
&\bar{\psi}_\varepsilon \varepsilon l_\varepsilon(X^\varepsilon(w), \hat{\theta}_\varepsilon(w)) \\
&= \varepsilon^{-1} \int_0^T c(X_t^0, \theta_0) \, \mathrm{d}t + \int_0^T b(X_t^0, \theta_0) \, \mathrm{d}w_t + \int_0^T \partial c(X_t^0, \theta_0) D_t \, \mathrm{d}t + \\
&\quad + \left(\int_0^T \delta c(X_t^0, \theta_0) \, \mathrm{d}t \right)' \zeta^{(0)}(w) + \varepsilon \int_0^T \partial b(X_t^0, \theta_0) D_t \, \mathrm{d}w_t + \\
&\quad + \varepsilon \left(\int_0^T \delta b(X_t^0, \theta_0) \, \mathrm{d}w_t \right)' \zeta^{(0)}(w) + \varepsilon \left(\int_0^T \partial \delta c(X_t^0, \theta_0) D_t \, \mathrm{d}t \right)' \zeta^{(0)}(w) + \\
&\quad + \frac{1}{2} \varepsilon \left(\int_0^T \delta c(X_t^0, \theta_0) \, \mathrm{d}t \right)' \zeta^{(1)}(w) + \frac{1}{2} \varepsilon \int_0^T \partial^2 c(X_t^0, \theta_0)(D_t, D_t) \, \mathrm{d}t + \\
&\quad + \frac{1}{2} \varepsilon \int_0^T \partial c(X_t^0, \theta_0) E_t \, \mathrm{d}t + \frac{1}{2} \varepsilon (\zeta^{(0)}(w))' \times \\
&\quad \times \left(\int_0^T \delta^2 c(X_t^0, \theta_0) \, \mathrm{d}t \right) \zeta^{(0)}(w) + O_M(\varepsilon^2).
\end{aligned}$$

Similarly, we obtain

$$\begin{aligned}
&\bar{\psi}_\varepsilon \int_W \int_W \varepsilon l_\varepsilon(X^\varepsilon(w'), \hat{\theta}_\varepsilon(w))\, \mathrm{d}P(w')\, \mathrm{d}P(w) \\
&\quad = \varepsilon^{-1} \int_0^T c(X_t^0, \theta_0)\, \mathrm{d}t + \\
&\qquad + \frac{1}{2}\varepsilon \left(\int_0^T \delta c(X_t^0, \theta_0)\, \mathrm{d}t \right)' \int_W \zeta^{(1)}(w)\, \mathrm{d}P(w) + \\
&\qquad + \frac{1}{2}\varepsilon \int_W \int_0^T \partial^2 c(X_t^0, \theta_0)(D_t, D_t)\, \mathrm{d}t\, \mathrm{d}P(w') + \\
&\qquad + \frac{1}{2}\varepsilon \int_W \int_0^T \partial c(X_t^0, \theta_0) E_t\, \mathrm{d}t\, \mathrm{d}P(w') + \\
&\qquad + \frac{1}{2}\varepsilon \int_W (\zeta^{(0)}(w))' \left(\int_0^T \delta^2 c(X_t^0, \theta_0)\, \mathrm{d}t \right) \zeta^{(0)}(w)\, \mathrm{d}P(w) + O_M(\varepsilon^2).
\end{aligned}$$

Since $\bar{\psi}_\varepsilon = 1 - O_M(\varepsilon^K)$ for any $K > 0$ and $\bar{\psi}_\varepsilon \int_0^1 \delta b(\theta_0 + u(\hat{\theta}_\varepsilon - \theta_0))\, \mathrm{d}u = O_M(1)$, we see that

$$\begin{aligned}
\bar{\psi}_\varepsilon \varepsilon b(\hat{\theta}_\varepsilon) &= \bar{\psi}_\varepsilon \varepsilon b(\theta_0) + \bar{\psi}_\varepsilon \varepsilon^2 \int_0^1 \delta b(\theta_0 + u(\hat{\theta}_\varepsilon - \theta_0))\, \mathrm{d}u \times \\
&\qquad \times \left(\zeta^{(0)} + \frac{1}{2}\varepsilon \zeta^{(1)} + \varepsilon^2 \zeta_\varepsilon^{(2)} \right) \\
&= \varepsilon b(\theta_0) + O_M(\varepsilon^2).
\end{aligned}$$

We then have

$$\begin{aligned}
\bar{\psi}_\varepsilon \bar{S}_\varepsilon^* &= \bar{\psi}_\varepsilon \varepsilon l_\varepsilon(X^\varepsilon(w), \hat{\theta}_\varepsilon(w)) - \\
&\quad - \bar{\psi}_\varepsilon \int_W \int_W \varepsilon l_\varepsilon(X^\varepsilon(w'), \hat{\theta}_\varepsilon(w))\, \mathrm{d}P(w')\, \mathrm{d}P(w) - \bar{\psi}_\varepsilon \varepsilon b(\hat{\theta}_\varepsilon) \\
&= \int_0^T b(X_t^0, \theta_0)\, \mathrm{d}w_t + \int_0^T \partial c(X_t^0, \theta_0) D_t\, \mathrm{d}t + \\
&\quad + \left(\int_0^T \delta c(X_t^0, \theta_0)\, \mathrm{d}t \right)' \zeta^{(0)}(w) + \varepsilon \int_0^T \partial b(X_t^0, \theta_0) D_t\, \mathrm{d}w_t + \\
&\quad + \varepsilon \left(\int_0^T \delta b(X_t^0, \theta_0)\, \mathrm{d}w_t \right)' \zeta^{(0)}(w) + \\
&\quad + \varepsilon \left(\int_0^T \partial \delta c(X_t^0, \theta_0) D_t\, \mathrm{d}t \right)' \zeta^{(0)}(w) + \\
&\quad + \frac{1}{2}\varepsilon \left(\int_0^T \delta c(X_t^0, \theta_0)\, \mathrm{d}t \right)' \zeta^{(1)}(w) -
\end{aligned}$$

$$
\begin{aligned}
&- \frac{1}{2}\varepsilon\biggl(\int_0^T \delta c(X_t^0, \theta_0)\,\mathrm{d}t\biggr)'\int_W \zeta^{(1)}(w)\,\mathrm{d}P(w) + \\
&+ \frac{1}{2}\varepsilon\int_0^T \partial^2 c(X_t^0, \theta_0)(D_t, D_t)\,\mathrm{d}t - \\
&- \frac{1}{2}\varepsilon\int_W\int_0^T \partial^2 c(X_t^0, \theta_0)(D_t, D_t)\,\mathrm{d}t\,\mathrm{d}P(w') + \\
&+ \frac{1}{2}\varepsilon\int_0^T \partial c(X_t^0, \theta_0)E_t\,\mathrm{d}t - \frac{1}{2}\varepsilon\int_W\int_0^T \partial c(X_t^0, \theta_0)E_t\,\mathrm{d}t\,\mathrm{d}P(w') + \\
&+ \frac{1}{2}\varepsilon(\zeta^{(0)}(w))'\biggl(\int_0^T \delta^2 c(X_t^0, \theta_0)\,\mathrm{d}t\biggr)\zeta^{(0)}(w) - \\
&- \frac{1}{2}\varepsilon\int_W (\zeta^{(0)}(w))'\biggl(\int_0^T \delta^2 c(X_t^0, \theta_0)\,\mathrm{d}t\biggr)\zeta^{(0)}(w)\,\mathrm{d}P(w) - \\
&- \varepsilon b(\theta_0) + O_M(\varepsilon^2).
\end{aligned}
$$

From the second-order asymptotic expansion of $\bar{S}_\varepsilon^*$ as above, f_0 and f_1 are determined.

In the same way as the second-order asymptotic expansion, we can see that for any $k \in \mathbf{N}$,

$$\bar{\psi}_\varepsilon \bar{S}_\varepsilon^* = \bar{\psi}_\varepsilon(f_0 + \varepsilon f_1 + \cdots + \varepsilon^{k-1} f_{k-1} + \varepsilon^k f_\varepsilon^{(k)}),$$

where $f_\varepsilon^{(k)} = O_M(1)$. Moreover, since $\bar{\psi}_\varepsilon = 1 - O_M(\varepsilon^K)$ for any $K > 0$, we obtain

$$\bar{\psi}_\varepsilon \bar{S}_\varepsilon^* \sim f_0 + \varepsilon f_1 + \cdots$$

in D^∞ as $\varepsilon \downarrow 0$. This completes the proof. □

Proof of Lemma 6. From definition, $|\xi_\varepsilon| > 1$ if $\sigma_{\bar{\psi}_\varepsilon \bar{S}_\varepsilon^*} < (1/2)\sigma_{f_0}$ or $\sigma_{\bar{\psi}_\varepsilon \bar{S}_\varepsilon^*} > (3/2)\sigma_{f_0}$. In case that $(3/2)\sigma_{f_0} \geqslant \sigma_{\bar{\psi}_\varepsilon \bar{S}_\varepsilon^*} \geqslant (1/2)\sigma_{f_0}$, we obtain $(2/3)\sigma_{f_0}^{-1} \leqslant \sigma_{\bar{\psi}_\varepsilon \bar{S}_\varepsilon^*}^{-1} \leqslant 2\sigma_{f_0}^{-1}$. Therefore, it follows from Assumption 4 that for any $p \in (1, \infty)$,

$$\sup_{\varepsilon \in (0,1]} E[1_{\{|\xi_\varepsilon| \leqslant 1\}}(\sigma_{\bar{\psi}_\varepsilon \bar{S}_\varepsilon^*})^{-p}] < \infty.$$

From Lemma 5, $\bar{\psi}_\varepsilon \bar{S}_\varepsilon^* \in D^\infty$ has the asymptotic expansion

$$\bar{\psi}_\varepsilon \bar{S}_\varepsilon^* \sim f_0 + \varepsilon f_1 + \cdots$$

in D^∞ as $\varepsilon \downarrow 0$ with $f_0, f_1, \ldots \in D^\infty$. We then have $\bar{\psi}_\varepsilon \bar{S}_\varepsilon^* = f_0 + \varepsilon f_\varepsilon^{(1)}$, where $f_0 \in D^\infty$, $f_\varepsilon^{(1)} = O(1) \in D^\infty$. Since

$$|\sigma_{\bar{\psi}_\varepsilon \bar{S}_\varepsilon^*}^{1/2} - \sigma_{f_0}^{1/2}| \leqslant \sigma_{\bar{\psi}_\varepsilon \bar{S}_\varepsilon^* - f_0}^{1/2} = \sigma_{\varepsilon f_\varepsilon^{(1)}}^{1/2} = \varepsilon\sigma_{f_\varepsilon^{(1)}}^{1/2},$$

we have

$$|\sigma_{\bar{\psi}_\varepsilon \bar{S}^*_\varepsilon} - \sigma_{f_0}| = |\sigma^{1/2}_{\bar{\psi}_\varepsilon \bar{S}^*_\varepsilon} - \sigma^{1/2}_{f_0}||\sigma^{1/2}_{\bar{\psi}_\varepsilon \bar{S}^*_\varepsilon} + \sigma^{1/2}_{f_0}| \leqslant \varepsilon \sigma^{1/2}_{f^{(1)}_\varepsilon}(\sigma^{1/2}_{\bar{\psi}_\varepsilon \bar{S}^*_\varepsilon} + \sigma^{1/2}_{f_0}).$$

Hence, it follows that there exists $C > 0$ such that for any $p > 1$,

$$\|\sigma_{\bar{\psi}_\varepsilon \bar{S}^*_\varepsilon} - \sigma_{f_0}\|_p \leqslant C\varepsilon. \tag{13}$$

From definition, $\xi_\varepsilon = O(1)$ in D^∞ as $\varepsilon \downarrow 0$. It follows from Chebyshev's inequality that for any $a > 0$, $K > 0$,

$$\begin{aligned} P[\xi_\varepsilon > a] &\leqslant P\left[\left|\frac{2(\sigma_{\bar{\psi}_\varepsilon \bar{S}^*_\varepsilon} - \sigma_{f_0})}{\sigma_{f_0}}\right| > a\right] \\ &\leqslant \frac{1}{a^K}\left\|\frac{2(\sigma_{\bar{\psi}_\varepsilon \bar{S}^*_\varepsilon} - \sigma_{f_0})}{\sigma_{f_0}}\right\|_K^K. \end{aligned}$$

It follows from (13), Hölder's inequality and Assumption 4 that for any $K > 0$,

$$\left\|\frac{2(\sigma_{\bar{\psi}_\varepsilon \bar{S}^*_\varepsilon} - \sigma_{f_0})}{\sigma_{f_0}}\right\|_K^K = O(\varepsilon^K).$$

We then see that for any $a > 0$, $K > 0$,

$$P[\xi_\varepsilon > a] = O(\varepsilon^K). \tag{14}$$

From (14), we see that for any $p > 1$,

$$\|1 - \psi^*_\varepsilon\|_p = \|1 - \psi(\xi_\varepsilon)\|_p \leqslant \|1_{\{|\xi_\varepsilon| > 1/2\}}\|_p = O(\varepsilon^K).$$

In view of the chain rule for H-derivatives,

$$D(1 - \psi^*_\varepsilon) = -D\psi^*_\varepsilon = -\psi'(\xi_\varepsilon)D\xi_\varepsilon.$$

Since $\|\psi'(\xi_\varepsilon)\|_p = O(\varepsilon^K)$, we see that for any $p > 1$,

$$\|D(1 - \psi^*_\varepsilon)\|_p = O(\varepsilon^K).$$

Similarly, it follows that for any $p > 1$ and $j > 0$,

$$\|D^j(1 - \psi^*_\varepsilon)\|_p = O(\varepsilon^K).$$

Therefore, we see that for any $K > 0$, $\psi^*_\varepsilon = 1 - O(\varepsilon^K)$ in D^∞. By using Theorem 4,

$$\begin{aligned} \psi^*_\varepsilon T(\bar{\psi}_\varepsilon \bar{S}^*_\varepsilon) &\sim \psi^*_\varepsilon T(f_0 + \varepsilon f_1 + \cdots) \\ &\sim \psi^*_\varepsilon(\Phi_0 + \varepsilon \Phi_1 + \cdots) \\ &\sim \Phi_0 + \varepsilon \Phi_1 + \cdots \end{aligned}$$

in $\tilde{D}^{-\infty}$ as $\varepsilon \downarrow 0$ with $\Phi_0, \Phi_1, \ldots$ in $\tilde{D}^{-\infty}$. This completes the proof. □

Proof of Theorem 5. From Lemmas 2, 3 and 6, we have

$$\begin{aligned} E[f(\bar{S}^*_\varepsilon)] &\sim E[\psi^*_\varepsilon f(\bar{\psi}_\varepsilon \bar{S}^*_\varepsilon)] \\ &= E[\psi^*_\varepsilon f \circ (\bar{\psi}_\varepsilon \bar{S}^*_\varepsilon)] \\ &\sim E[\Xi_0] + \varepsilon E[\Xi_1] + \cdots \end{aligned}$$

as $\varepsilon \downarrow 0$ uniformly in any $f \in \mathcal{E}(M, \gamma)$. Therefore, the rest is to calculate $E[\Xi_i]$, $i = 0, 1, \ldots$. From the regularity of f_0 and integration by part formula,

$$\begin{aligned} E[\Xi_i] &= E[G_i(w) f(f_0)] \\ &= \int_{\mathbf{R}} f(x) E[G_i(w) | f_0 = x] p_{f_0}(x) \, \mathrm{d}x \end{aligned}$$

for some smooth functional $G_i(w)$. Therefore, each term is represented by an integration of a smooth function. We will only determine p_0 and p_1. It is easy to show that $p_0(x) \equiv p_{f_0}(x) = \phi(x; \Sigma)$. Moreover, we obtain

$$\begin{aligned} E[\Xi_1] &= E[f_1 \partial f(f_0)] \\ &= E[\partial f(f_0) E[f_1 | f_0]] \\ &= -\int_{\mathbf{R}} f(x) \partial_x \{E[f_1 | f_0 = x] p_{f_0}(x)\} \, \mathrm{d}x. \end{aligned}$$

We then have $p_1(x) = -\partial_x \{E[f_1 | f_0 = x] p_{f_0}(x)\}$. This completes the proof. □

Proof of Theorem 6. In Theorem 5, for any $f \in \mathcal{E}'(M, \gamma)$, let $b(\cdot) = b_f(\cdot)$. From the point of view of the second-order statistical inference, it then follows that

$$\begin{aligned} E[f(\bar{S}^*_\varepsilon)] &= \int_{\mathbf{R}} f(x) \phi(x; \Sigma) \, \mathrm{d}x - \\ &\quad - \varepsilon \int_{\mathbf{R}} f(x) \partial_x \{E[F_1 - b_f(\theta_0) | f_0 = x] \phi(x; \Sigma)\} \, \mathrm{d}x + o(\varepsilon) \\ &= \int_{\mathbf{R}} f(x) \phi(x; \Sigma) \, \mathrm{d}x - \varepsilon \int_{\mathbf{R}} f(x) \partial_x \{E[F_1 | f_0 = x] \phi(x; \Sigma)\} \, \mathrm{d}x + \\ &\quad + \varepsilon b_f(\theta_0) \int_{\mathbf{R}} f(x) \partial_x \{\phi(x; \Sigma)\} \, \mathrm{d}x + o(\varepsilon) \\ &= o(\varepsilon). \end{aligned}$$

This completes the proof. □

Proof of Theorem 7. In Theorem 5, putting $f(x) = x$ and $b(\cdot) = b_1(\cdot)$, we obtain

$$\begin{aligned} E[\bar{S}^*_T] &= E[f_0] + \varepsilon E[f_1] + o(\varepsilon) \\ &= \varepsilon E\left[(\zeta^{(0)})' \int_0^T \delta b(X^0_t, \theta_0) \, \mathrm{d}w_t + (\zeta^{(0)})' \int_0^T \partial \delta c(X^0_t, \theta_0) D_t \, \mathrm{d}t \right] - \\ &\quad - \varepsilon b_1(\theta_0) + o(\varepsilon) \\ &= o(\varepsilon). \end{aligned}$$

This completes the proof. □

Proof of Theorem 8. In Theorem 5, putting $f(x) = 1_{\{(0,\infty)\}}(x) - 1/2$ and $b(\cdot) = b_2(\cdot)$, we obtain

$$\begin{aligned} P[\bar{S}_T^* > 0] - \frac{1}{2} &= \int_0^\infty \phi(x, \Sigma)\,\mathrm{d}x - \frac{1}{2} - \varepsilon \int_0^\infty \partial_x \{E[f_1 | f_0 = x] \times \\ &\quad \times \phi(x, \Sigma)\}\,\mathrm{d}x + o(\varepsilon) \\ &= \varepsilon E[F_1 | f_0 = 0]\phi(0, \Sigma) - \varepsilon b_2(\theta_0)\phi(0, \Sigma) + o(\varepsilon) \\ &= o(\varepsilon). \end{aligned}$$

Similarly, putting $f(x) = 1_{\{(-\infty,0)\}}(x) - 1/2$ and $b(\cdot) = b_2(\cdot)$, we have

$$\begin{aligned} P[\bar{S}_T^* < 0] - \tfrac{1}{2} &= -\varepsilon E[F_1 | f_0 = 0]\phi(0, \Sigma) + \varepsilon b_2(\theta_0)\phi(0, \Sigma) + o(\varepsilon) \\ &= o(\varepsilon). \end{aligned}$$

This completes the proof. □

Proof of Theorem 1. It follows from Theorem 7 and Itô's formula that we obtain

$$\begin{aligned} b_1(\theta_0) &= \sum_{i=1}^p E\Bigg[\bigg(\int_0^T b_{0,t}^i\,\mathrm{d}w_t\bigg)\bigg(\int_0^T \delta_i b(X_t^0, \theta_0)\,\mathrm{d}w_t + \\ &\quad + \sum_{\alpha=1}^d \int_0^T \partial_\alpha \delta_i c(X_t^0, \theta_0) D_t^\alpha\,\mathrm{d}t\bigg)\Bigg] \\ &= \sum_{i=1}^p \Bigg[\int_0^T b_{0,t}^i \Big(\delta_i b(X_t^0, \theta_0)\Big)'\,\mathrm{d}t + \\ &\quad + \sum_{\alpha=1}^d E\Bigg[\bigg(\int_0^T b_{0,t}^i\,\mathrm{d}w_t\bigg)\bigg(\int_0^T \int_t^T \partial_\alpha \delta_i c(X_s^0, \theta_0)\lambda_{s,t}^\alpha\,\mathrm{d}s\,\mathrm{d}w_t\bigg)\Bigg]\Bigg] \\ &= \sum_{i=1}^p \Bigg[\int_0^T b_{0,t}^i \Big(\delta_i b(X_t^0, \theta_0)\Big)'\,\mathrm{d}t + \\ &\quad + \sum_{\alpha=1}^d \int_0^T b_{0,t}^i \bigg(\int_t^T \partial_\alpha \delta_i c(X_s^0, \theta_0)\lambda_{s,t}^\alpha\,\mathrm{d}s\bigg)'\,\mathrm{d}t\Bigg], \end{aligned}$$

which completes the proof. □

Proof of Theorem 2. From Theorem 8, we compute $b_2(\theta_0) = E[F_1|f_0 = 0]$. From (9)–(11) and Lemma 1, we have

$$\begin{aligned}E[F_1|f_0 = x] \;=\; & \sum_{\alpha=1}^{d} J_\alpha^1(\partial_\alpha b(X_t^0, \theta_0))(x) + \sum_{i=1}^{p} I^{11}(b_{2,t}^i, b_{0,t}^i)(x) + \\ & + \frac{1}{2}\sum_{i=1}^{p}\int_0^T \delta_i c(X_t^0, \theta_0)\,\mathrm{d}t\; E[\zeta_i^{(1)}|f_0 = x] + \\ & + \frac{1}{2}\sum_{\alpha,\beta=1}^{d} J_{\alpha,\beta}^2(\partial_\alpha\partial_\beta c(X_t^0, \theta_0))(x) + \\ & + \frac{1}{2}\sum_{\alpha=1}^{d} J_\alpha^3(\partial_\alpha c(X_t^0, \theta_0))(x) + \\ & + \frac{1}{2}\sum_{i,j=1}^{p}\int_0^T \delta_i\delta_j c(X_t^0, \theta_0)\,\mathrm{d}t\; I^{11}(b_{0,t}^i, b_{0,t}^j)(x) - C_0,\end{aligned}$$

where

$$\begin{aligned}E[\zeta_i^{(1)}|f_0 = x] \;=\; & -\sum_{j=1}^{p} I(\theta_0)^{i,j}\left\{2\sum_{\alpha=1}^{d} J_\alpha^1(\partial_\alpha\delta_j B(X_t^0, \theta_0))(x) + \right. \\ & + 2\sum_{k=1}^{p} I^{11}(b_{1,t}^{j,k}, b_{0,t}^k)(x) + \\ & + \sum_{\alpha,\beta=1}^{d} J_{\alpha,\beta}^2(\partial_\alpha\partial_\beta\delta_j C(X_t^0, \theta_0))(x) + \\ & + \sum_{\alpha=1}^{d} J_\alpha^3(\partial_\alpha\delta_j C(X_t^0, \theta_0))(x) + \\ & \left. + \sum_{k,l=1}^{p}\int_0^T \partial_j\partial_k\partial_l C(X_t^0, \theta_0)\,\mathrm{d}t\; I^{11}(b_{0,t}^k, b_{0,t}^l)(x)\right\}.\end{aligned}$$

This completes the proof. □

Acknowledgements

The authors are grateful to the referee for valuable comments. This work was in part supported by the Research Fund for Scientists of the Ministry of Science, Education and Culture, and by Cooperative Research Program of the Institute of Statistical Mathematics.

References

1. Akahira, M. and Takeuchi, K.: Asymptotic efficiency of statistical estimators: concepts and higher order asymptotic efficiency. *Lecture Notes in Statistics*, Vol. 7, Springer, Berlin, 1981.
2. Akaike, H.: Information theory and an extension of the maximum likelihood principle. In: B. N. Petrov and F. Csaki (eds), *Proceedings of the 2nd International Symposium on Information Theory*, Akademiai Kaido, Budapest, 1973, pp. 267–281.
3. Akaike, H.: A new look at the statistical model identification, *IEEE Trans. Auto. Control* **AC-19** (1974), 716–723.
4. Bibby, B. M. and Sørensen, M.: Martingale estimation functions for discretely observed diffusion processes, *Bernoulli* **1** (1995), 17–39.
5. Dermoune, A. and Kutoyants, Yu. A.: Expansion of distribution function of maximum likelihood estimate for misspecified diffusion type observations, *Stochast. Stochast. Rep.* **52** (1995), 121–145.
6. Genon-Catalot, V.: Maximum contrast estimation for diffusion processes from discrete observations, *Statistics* **21** (1990), 99–116.
7. Genon-Catalot, V. and Jacod, J.: On the estimation of the diffusion coefficient for multidimensional diffusion processes, *Ann. Inst. Henri Poincaré Probab. Statist.* **29** (1993), 119–151.
8. Genon-Catalot, V., Jeantheau, T. and Laredo, C.: Limit theorems for discretely observed stochastic volatility models, *Bernoulli* **4** (1998), 283–303.
9. Genon-Catalot, V., Jeantheau, T. and Laredo, C.: Parameter estimation for discretely observed stochastic volatility models, *Bernoulli* **5** (1999), 855–872.
10. Ikeda, N. and Watanabe, S.: *Stochastic Differential Equations and Diffusion Processes*, 2nd edn, North-Holland/Kodansha, Tokyo, 1989.
11. Kessler, M.: Estimation of an ergodic diffusion from discrete observations, *Scand. J. Statist.* **24** (1997), 211–229.
12. Kim, Y. J. and Kunitomo, N.: Pricing options under stochastic interest rates, *Asia Pacific Financial Markets* **6** (1999), 49–70.
13. Konishi, S. and Kitagawa, G.: Generalised information criteria in model selection, *Biometrika* **83** (1996), 875–890.
14. Kunitomo, N. and Takahashi, A.: The asymptotic expansion approach to the valuation of interest rate contingent claims, *Math. Fin.* **11** (2001), 117–151.
15. Kusuoka, S. and Yoshida, N.: Malliavin calculus, geometric mixing, and expansion of diffusion functionals, *Probab. Theory Relat. Fields* **116** (2000), 457–484.
16. Kutoyants, Yu. A.: In: B. L. S. Prakasa Rao (ed.), *Parameter Estimation for Stochastic Processes*, Heldermann, Berlin, 1984.
17. Kutoyants, Yu. A.: On nonparametric estimation of trend coefficients in a diffusion process. In: N. V. Krylov et al. (eds), *Statistics and Control of Stochastic Processes*, Optimization Software, New York, 1985, pp. 230–250.
18. Kutoyants, Yu. A.: *Identification of Dynamical Systems with Small Noise*, Kluwer Academic Publishers, Dordrecht, 1994.
19. Laredo, C. F.: A sufficient condition for asymptotic sufficiency of incomplete observations of a diffusion process, *Ann. Statist.* **18** (1990), 1158–1171.
20. Liptser, R. S. and Shiryayev, A. N.: *Statistics of Random Processes I*, Springer, Berlin, 1977.
21. Malliavin, P.: Stochastic calculus of variation and hypoelliptic operators. In: K. Itô (ed.), *Proceedings of the International Symposium on SDE Kyoto*, Kinokuniya, Tokyo, 1976, pp. 195–263.
22. Malliavin, P.: C^k-hypoellipticity with degeneracy. In: A. Friedman and M. Oinsky (eds), *Stochastic Analysis*, Academic Press, New York, 1978, pp. 199–214, 327–340.
23. Malliavin, P.: *Stochastic Analysis*, Springer, Berlin, 1997.
24. Nualart, D.: *The Malliavin Calculus and Related Topics*, Springer, Berlin, 1995.

25. Prakasa Rao, B. L. S.: Statistical inference from sampled data for stochastic processes. In: N. U. Prabhu (ed.), *Statistical Inference from Stochastic Processes, Contemporary Mathematics*, Vol. 80, American Mathematical Society, Providence, RI, 1988, pp. 249–284.
26. Prakasa Rao, B. L. S.: *Statistical Inference for Diffusion Type Processes*, Edward Arnold, London; Oxford University Press, New York, 1999.
27. Sakamoto, Y. and Yoshida, N.: Expansion of perturbed random variables based on generalized Wiener functionals, *J. Multivariate Anal.* **59** (1996), 34–59.
28. Sakamoto, Y. and Yoshida, N.: Asymptotic expansions of M-estimator over Wiener space, Statist. Infer. Stochast. Process. **1** (1998), 85–103.
29. Sakamoto, Y. and Yoshida, N.: Higher order asymptotic expansion for a functional of a mixing process with applications to diffusion processes (1999) (unpublished).
30. Sørensen, M.: Small dispersion asymptotics for diffusion martingale estimating functions, Preprint No. 2000-2, Department of Statistics and Operations Research, University of Copenhagen, 2000.
31. Takahashi, A.: An asymptotic expansion approach to pricing contingent claims, *Asia Pacific Financial Markets* **6** (1999), 115–151.
32. Takanobu, S. and Watanabe, S.: Asymptotic expansion formulas of the Schilder type for a class of conditional Wiener functional integrations. In: K. D. Elworthy and N. Ikeda (eds), *Asymptotic Problems in Probability Theory: Wiener Functionals and Asymptotics*, Proceedings of the Taniguchi International Symposium, Sanda and Kyoto, 1990, Longman Scientific & Technical, Harlow, UK, 1993, pp. 194–241.
33. Takeuchi, K.: Distribution of information statistics and criteria for adequacy of models, *Math. Sci.* **153** (1976), 12–18 (in Japanese).
34. Uchida, M. and Yoshida, N.: Information criteria in model selection for stochastic processes (II), *Research Memorandum*, Vol. 730, The Institute of the Statistical Mathematics, Tokyo, 1999.
35. Uchida, M. and Yoshida, N.: Information criteria in model selection for mixing processes, *Statist. Infer. Stochast. Process.* **4** (2001), 73–98.
36. Uchida, M. and Yoshida, N.: Asymptotic expansion for small diffusions applied to option pricing, (2001) (revised).
37. Watanabe, S.: Malliavin's calculus in terms of generalized Wiener functionals, *Lect. Notes in Control Inf. Sci.*, Vol. 49, Springer, Berlin, 1983, pp. 284–290.
38. Watanabe, S.: *Lectures on Stochastic Differential Equations and Malliavin Calculus*, Tata Institute of Fundamental Research. Springer, Berlin, 1984.
39. Watanabe, S.: Analysis of Wiener functionals (Malliavin calculus) and its applications to heat kernels, *Ann. Probab*. **15** (1987), 1–39.
40. Yoshida, N.: Asymptotic expansion of maximum likelihood estimators for small diffusions via the theory of Malliavin–Watanabe, *Probab. Theory Relat. Fields* **92** (1992), 275–311.
41. Yoshida, N.: Asymptotic expansion for statistics related to small diffusions, *J. Jpn. Statist. Soc*. **22** (1992), 139–159.
42. Yoshida, N.: Estimation for diffusion processes from discrete observation, *J. Multivariate Anal.* **41** (1992), 220–242.
43. Yoshida, N.: Asymptotic expansion of Bayes estimators for small diffusions, *Probab. Theory Relat. Fields* **95** (1993), 429–450.
44. Yoshida, N.: Banach space valued functionals and smoothness of M-estimators, *Research Memorandum* Vol. 494, The Institute of the Statistical Mathematics, Tokyo, 1993.
45. Yoshida, N.: Asymptotic expansions for perturbed systems on Wiener space: maximum likelihood estimators, *J. Multivariate Anal*. **57** (1996), 1–36.
46. Yoshida, N.: Malliavin calculus and statistics. In: Johnson et al. (eds), *Encyclopedia in Statistics*, Vol. 3, 1998.
47. Yoshida, N.: Conditional expansions and their applications (2001) (to appear in *Stoch. Proc. Appl.*).