Adaptive inference for small diffusion processes based on sampled data

Tetsuya Kawai¹ · Masayuki Uchida²

Received: 2 February 2022 / Accepted: 29 November 2022 / Published online: 16 December 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
We consider parametric estimation and tests for multi-dimensional diffusion processes with a small dispersion parameter $\varepsilon$ from discrete observations. For parametric estimation of diffusion processes, the main target is to estimate the drift parameter and the diffusion parameter. In this paper, we propose two types of adaptive estimators for both parameters and show their asymptotic properties under $\varepsilon \rightarrow 0$, $n \rightarrow \infty$ and the balance condition that $(\varepsilon n^\rho)^{-1} = O(1)$ for some $\rho > 0$. Using these adaptive estimators, we also introduce consistent adaptive testing methods and prove that test statistics for adaptive tests have asymptotic distributions under null hypothesis. In simulation studies, we examine and compare asymptotic behaviors of the two kinds of adaptive estimators and test statistics. Moreover, we treat the SIR model which describes a simple epidemic spread for a biological application.

Keywords Adaptive test · Asymptotic theory · Discrete time observation · Minimum contrast estimation · Stochastic differential equation · SIR model

1 Introduction

We consider a $d$-dimensional small diffusion process satisfying the following stochastic differential equation (SDE):

$$\begin{cases}
    dX_t = b(X_t, \alpha)dt + \varepsilon \sigma(X_t, \beta)dW_t, & t \in [0, T], \\
    X_0 = x_0,
\end{cases}$$

(1)
where \( W_t \) is the \( r \)-dimensional standard Wiener process, \( \alpha \in \Theta_\alpha \subset \mathbb{R}^p, \beta \in \Theta_\beta \subset \mathbb{R}^q, \theta = (\alpha, \beta), \Theta := \Theta_\alpha \times \Theta_\beta, \Theta_\alpha \) and \( \Theta_\beta \) are compact and convex parameter spaces, \( b : \mathbb{R}^d \times \Theta_\alpha \rightarrow \mathbb{R}^d \) and \( \sigma : \mathbb{R}^d \times \Theta_\beta \rightarrow \mathbb{R}^d \otimes \mathbb{R}^r \) are known except for the parameter \( \theta \), and the initial value \( x_0 \in \mathbb{R} \) and the small coefficient \( \varepsilon > 0 \) are known. We assume the true parameter \( \theta_0 = (\alpha_0, \beta_0) \) belongs to \( \text{Int}(\Theta) \), and the data are discrete observations \( (X_{t_n^k})_{k=0, \ldots, n} \), where \( t_n^k = kh_n \) and \( h_n = T/n \).

A family of small diffusion processes defined by (1) is an important class and called dynamical systems with small perturbations, see Azencott (1982), Freidlin and Wentzell (1998). Yoshida (1992b) proved asymptotic expansions of maximum likelihood estimators for small diffusion processes by means of the Malliavin calculus. The asymptotic expansion scheme to compute the values of options was provided by Yoshida (1992c) as an application of small diffusion processes to mathematical finance. See also Uchida and Yoshida (2004). For an application of small diffusion processes to mathematical biology, see Guy et al. (2014, 2015) and references therein.

Asymptotic theory of parametric inference for small diffusion processes has been well-developed. For continuous-time observations, see Kutoyants (1984, 1994) and Yoshida (1992b, 2003). As for discrete observations, Genon-Catalot (1990) studied minimum contrast estimation for the drift parameter and proved that this estimator has asymptotic efficiency under the assumption \( \varepsilon \sqrt{n} = O(1) \). Laredo (1990) investigated the asymptotically efficient estimator by using interpolated process under the assumption \( (\varepsilon n^2)^{-1} \rightarrow 0 \). Sørensen and Uchida (2003) studied the joint estimation for both drift and diffusion parameters based on minimum contrast estimators. They proved that the estimator for drift parameter is asymptotically efficient and the estimator for diffusion parameter is asymptotically normal under \( (\varepsilon \sqrt{n})^{-1} = O(1) \). Uchida (2004) investigated the asymptotically efficient estimator for drift parameter by using the approximate martingale estimating function under \( (\varepsilon n^l)^{-1} \rightarrow 0 \), where \( l \) is a positive integer. For the asymptotically efficient estimator of drift parameter based on an approximate martingale estimating function of a one-dimensional small diffusion process under \( \varepsilon \rightarrow 0 \) and \( n \rightarrow \infty \), see Uchida (2008). Gloter and Sørensen (2009) generalized the results of Sørensen and Uchida (2003) and Uchida (2004). They proposed the minimum contrast estimators for both drift and diffusion parameters whose asymptotic covariance matrix equals to that of the estimators in Sørensen and Uchida (2003) under \( (\varepsilon n^\rho)^{-1} = O(1) \), where \( \rho > 0 \).

The adaptive parametric inference for diffusion processes has been studied by many researchers. Since the adaptive parametric method can divide the inference for \( (\alpha, \beta) \) into that for the drift parameter \( \alpha \) and that for diffusion parameter \( \beta \), we expect that the adaptive parametric inference for diffusion processes is dealt with more accurately and quickly from the viewpoint of numerical analysis. Note that the adaptive parametric inference differs from adaptive nonparametric inference in the sense of e.g. Tsybakov (2009). As we know very well, the adaptive parametric estimator is related to the one-step estimator in the sense of e.g. van der Vaart (1998). For adaptive parametric estimation for ergodic diffusion processes, many researchers studied and obtained the asymptotic results, see Prakasa Rao (1983, 1988), Yoshida (1992a), Kessler (1995), and Uchida and Yoshida (Aug 2012). Nomura and Uchida (2016) and Kaino and Uchida (2018) proposed the adaptive Bayes type estimators and the hybrid estimators for both drift and diffusion parameters in small diffusion processes. They proved
the asymptotic efficiency for the estimator of the drift parameter and the asymptotic normality for the estimator of diffusion parameter whose asymptotic variance equals to that of the estimator in Sørensen and Uchida (2003). Moreover their estimators have convergence of moments under $(\varepsilon \sqrt{n})^{-1} = O(1)$. For adaptive tests, Kawai and Uchida (2022) proposed adaptive testing method in ergodic diffusion processes, which can test drift parameters and diffusion parameters separately. To construct test statistics of these adaptive tests, adaptive estimators in ergodic diffusions were used and asymptotic properties of these test statistics were proved. Nakakita and Uchida (2019) investigated the adaptive test for noisy ergodic diffusion processes. They derived the asymptotic null distribution of the adaptive test statistics based on the local means and consistency of the tests under alternatives.

In this paper, we utilize the results of Gloter and Sørensen (2009) and propose the two types of adaptive estimators for both drift and diffusion parameters in small diffusion processes under the assumption $(\varepsilon n^\rho)^{-1} = O(1)$, where $\rho > 0$. In small diffusions, the convergence rates of estimators for the drift parameter and the diffusion parameter are $\varepsilon^{-1}$ and $\sqrt{n}$, respectively. We first estimate drift parameter $\alpha$ and next estimate diffusion parameter $\beta$ in our adaptive estimation methods for arbitrary $\rho > 0$. Moreover, if the balance coefficient $\rho$ satisfies $0 < \rho < 1/2$, we can also utilize simpler adaptive estimation methods than the above methods. In these simpler methods, we first estimate diffusion parameter since it holds that $\varepsilon^{-1}/\sqrt{n} \to 0$.

The main result of this paper is that our proposed two kinds of adaptive estimators have asymptotic efficiency for $\alpha$ and asymptotic normality for $\beta$ whose asymptotic covariance matrix is the same as that of the estimators proposed in Sørensen and Uchida (2003) under the milder assumption than $(\varepsilon \sqrt{n})^{-1} = O(1)$. We also give the estimator for drift parameter $\alpha$ which can be estimated independently from diffusion parameter $\beta$ and show that the estimator has asymptotic normality under the assumption $(\varepsilon n^\rho)^{-1} = O(1)$. For adaptive tests, we introduce the two kinds of likelihood ratio type test statistics, which are constructed by the above proposed adaptive estimators. These test statistics have asymptotic distribution under null hypothesis, and these tests are consistent under alternatives.

The paper is organized as follows. In Sect. 2, notation and assumptions are introduced. The infinitesimal generator of the small diffusion process and its approximation used in constructing the contrast functions for adaptive estimators are defined. In Sect. 3, we propose the two kinds of adaptive estimators and state the asymptotic properties of the proposed estimators. Note that the drift parameter can be estimated independently of the diffusion parameter, and adaptive estimators can also treat the case of having the same parameters in the drift and diffusion coefficients. In Sect. 4, adaptive testing method for small diffusion processes are introduced and two kinds of test statistics are constructed. We show the asymptotic properties of these test statistics. In Sect. 5, we give some examples and simulation results of the asymptotic performance for two types of adaptive estimators and test statistics for multi-dimensional small diffusion processes. In model 1, we compare the adaptive estimators with the joint estimator proposed in Gloter and Sørensen (2009), and an adaptive test is also conducted. In model 2, we treat the numerical simulation of the SIR model. In model 3, the difference of the asymptotic performances between the two adaptive estimators
is examined. Section 6 is devoted to the proofs of the results presented in Sects. 3 and 4.

2 Notation and assumptions

In this paper, we set $\partial_{\alpha_i} := \partial/\partial \alpha_i$, $\partial_{\beta_i} := \partial/\partial \beta_i$, $\partial\alpha := (\partial_{\alpha_1}, \ldots, \partial_{\alpha_p})^\top$, $\partial\beta := (\partial_{\beta_1}, \ldots, \partial_{\beta_q})^\top$, $\partial_{\alpha}^2 := \partial\alpha\partial\alpha^\top$, $\partial_{\beta}^2 := \partial\beta\partial\beta^\top$, $\partial_{\alpha\beta}^2 := \partial\alpha\partial\beta^\top$, where $\top$ is the transpose of a matrix. The symbols $\rightarrow_p$ and $\xrightarrow{d}$ indicate convergence in probability and convergence in distribution, respectively.

Let $(X^0_t)$ be the solution of the following ordinary differential equation (ODE) which is the case of $\varepsilon = 0$ and $\alpha = \alpha_0$ in the SDE (1):

$$
\begin{cases}
    dX^0_t = b(X^0_t, \alpha_0)dt, & t \in [0, T], \\
    X^0_0 = x_0,
\end{cases}
$$

(2)

and $I(\theta_0)$ be the $(p + q) \times (p + q)$-matrix defined as

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

where

$$
I_b^{i,j}(\theta_0) = \int_0^T \left( \partial_{\alpha_i} b(X^0_s, \alpha_0) \right)^\top \left[ \sigma \sigma^\top \right]^{-1} \left( X^0_s, \beta_0 \right) \left( \partial_{\alpha_j} b(X^0_s, \alpha_0) \right) ds,
$$

$$
I_{\sigma}^{i,j}(\theta_0) = \frac{1}{2T} \int_0^T \text{tr} \left[ \left( \partial_{\beta_i} [\sigma \sigma^\top] \right) \left[ \sigma \sigma^\top \right]^{-1} \left( \partial_{\beta_j} [\sigma \sigma^\top] \right) \left[ \sigma \sigma^\top \right]^{-1} \right] \left( X^0_s, \beta_0 \right) ds.
$$

Moreover, we define $p \times p$-matrices $J_b(\alpha_0) = \left( J_b^{i,j}(\alpha_0) \right)_{1 \leq i, j \leq p}$ and $K_b(\theta_0) = \left( K_b^{i,j}(\theta_0) \right)_{1 \leq i, j \leq p}$ as

$$
J_b^{i,j}(\alpha_0) = \int_0^T \left( \partial_{\alpha_i} b(X^0_s, \alpha_0) \right)^\top \left( \partial_{\alpha_j} b(X^0_s, \alpha_0) \right) ds,
$$

$$
K_b^{i,j}(\theta_0) = \int_0^T \left( \partial_{\alpha_i} b(X^0_s, \alpha_0) \right)^\top \left[ \sigma \sigma^\top \right] \left( X^0_s, \beta_0 \right) \left( \partial_{\alpha_j} b(X^0_s, \alpha_0) \right) ds.
$$

Let $N_p(\mu, \Sigma)$ be the $p$-dimensional normal random variable with $p$-dimensional mean vector $\mu$ and $(p \times p)$-covariance matrix $\Sigma$. $\chi_n$ denotes the chi-square random variable with $n$ degree of freedom.

We make the following assumptions.
For all $\varepsilon > 0$, SDE (1) with the true value of the parameter has a unique strong solution on some probability space $(\Omega, \mathcal{F}, P)$, and ODE (2) has a unique solution.

For all $\varepsilon > 0$, SDE (1) with the true value of the parameter has a unique strong solution on some probability space $(\Omega_1, \mathcal{F}, P)$, and ODE (2) has a unique solution.

$b \in C^\infty(\mathbb{R}^d \times \Theta_\alpha)$, $\sigma \in C(\mathbb{R}^d \times \Theta_\beta)$, and there exists an open convex subset $\mathcal{U} \subset \mathbb{R}^d$ such that $X^0_t \in \mathcal{U}$ for all $t \in [0,1]$, and $\sigma \in C^\infty(\mathcal{U} \times \Theta_\beta)$. Moreover $[\sigma \sigma^\top](x, \beta)$ is invertible on $\mathcal{U} \times \Theta_\beta$.

(i) $I_b(\theta_0), J_b(\alpha_0)$ and $K_b(\theta_0)$ are non-singular, 
(ii) $I_\sigma(\theta_0)$ is non-singular.

$\varepsilon = \varepsilon_n \to 0$ as $n \to \infty$ and there exists $\rho > 0$ such that $\lim_{n \to \infty}(\varepsilon_n^{-\rho})^{-1} < +\infty$.

Remark 1

(i) Assumption [A2] is derived from a localization argument and hence [A2] is the mild condition. The strict assumption of [A2] and the relationship between the two assumptions are introduced in Sect. 6.

(ii) For [A4], if we assume a condition for $\sigma$, then we can clarify the relationship among the regularity for $I_b(\theta_0), J_b(\alpha_0)$ and $K_b(\theta_0)$. Under [A2'] in Sect. 6, in particular, it holds true that if one is regular, then the others are also regular.

For [B], we set the approximation degree $v$ as the integer such that $v = \lceil \rho + \frac{1}{2} \rceil$, where $\lceil y \rceil := \min\{z \in \mathbb{Z} | y \leq z\}$. Note that the smallest $\rho$ should be selected among the values of $\rho$ satisfying [B]. Here the two kinds of operators $L_{\alpha_0}^\varepsilon$ and $L_\alpha^0$ are introduced. We denote by $L_{\alpha_0}^\varepsilon$ the infinitesimal generator of the diffusion process $X$: for any smooth function $f$,

$$L_{\alpha_0}^\varepsilon(f)(x) := \sum_{i=1}^d b_i(x, \alpha_0) \frac{\partial}{\partial x_i} f(x) + \frac{\varepsilon^2}{2} \sum_{i,j=1}^d [\sigma \sigma^\top]^{i,j}(x, \beta_0) \frac{\partial^2}{\partial x_i \partial x_j} f(x),$$

and define the simple approximation of the generator $L_\alpha^0$ as

$$L_\alpha^0(f)(x) := \sum_{i=1}^d b_i(x, \alpha) \frac{\partial}{\partial x_i} f(x).$$

Using the operator $L_\alpha^0$, we set that for $2 \leq l \leq v$,

$$P_{1,k}(\alpha) := X^n_{k} - X^n_{k-1} - h_n b(X^n_{k-1}, \alpha),$$

$$P_{l,k}(\alpha) := P_{1,k}(\alpha) - Q_{l,k}(\alpha),$$

$$Q_{l,k}(\alpha) := \sum_{j=1}^{l-1} \frac{h_n^{j+1}}{(j + 1)!}(L_\alpha^0)^j b(X^n_{k-1}, \alpha).$$
In particular,

\[ P_{2,k}(\alpha) = X_{t_k} - X_{t_{k-1}} - h_n b(X_{t_{k-1}}, \alpha) - \frac{h_n^2}{2} \sum_{i=1}^{d} b'(X_{t_{k-1}}, \alpha) \frac{\partial}{\partial x_i} b(X_{t_{k-1}}, \alpha), \]

\[ Q_{2,k}(\alpha) = \frac{h_n^2}{2} \sum_{i=1}^{d} b'(X_{t_{k-1}}, \alpha) \frac{\partial}{\partial x_i} b(X_{t_{k-1}}, \alpha). \]

### 3 Adaptive estimation

In this section, we propose the following two types of adaptive estimators. The adaptive methods proposed in this paper are particular instances of an approximation of the joint estimation proposed in Gloter and Sørensen (2009). Note that the estimators in this section exist since the parameter spaces \( \Theta_1^{\alpha} \) and \( \Theta_1^{\beta} \) are compact sets and contrast functions are continuous on the parameter spaces. Each estimator may or may not be unique. For example, in Sect. 3.1 below, if there are two estimators \( \tilde{\alpha}_{\epsilon,n}^{(1)} \) and \( \tilde{\alpha}_{\epsilon,n}^{(1)} \) such that

\[ U_{\epsilon,n,v}(\tilde{\alpha}_{\epsilon,n}^{(1)}) = U_{\epsilon,n,v}(\tilde{\alpha}_{\epsilon,n}^{(1)}) = \inf_{\alpha \in \Theta_\alpha} U_{\epsilon,n,v}(\alpha), \tag{3} \]

then a uniform law of large numbers for \( U_{\epsilon,n,v}(\alpha) \) and [A3] together with (3) yield the consistency of \( \tilde{\alpha}_{\epsilon,n}^{(1)} \) and \( \tilde{\alpha}_{\epsilon,n}^{(1)} \). For details, see the proof of Lemma 1. This implies that \( \tilde{\alpha}_{\epsilon,n}^{(1)} - \tilde{\alpha}_{\epsilon,n}^{(1)} \xrightarrow{P} 0. \)

#### 3.1 Type I estimator

First, we introduce the adaptive estimators which divide the parameter optimization for \( \theta = (\alpha, \beta) \) into the optimization of \( \alpha \) and that of \( \beta \).

**Step 1.** Set

\[ U_{\epsilon,n,v}(\tilde{\alpha}_{\epsilon,n}^{(1)}) := \epsilon^{-2} h_n^{-1} \sum_{k=1}^{n} P_{v,k}(\alpha)^\top P_{v,k}(\alpha), \]

and \( \tilde{\alpha}_{\epsilon,n}^{(1)} \) is defined as

\[ U_{\epsilon,n,v}(\tilde{\alpha}_{\epsilon,n}^{(1)}) = \inf_{\alpha \in \Theta_\alpha} U_{\epsilon,n,v}(\alpha). \]

**Step 2.** Set

\[ U_{\epsilon,n,v}(\beta | \tilde{\alpha}) := \sum_{k=1}^{n} \left\{ \log \text{det}[\sigma \sigma^\top](X_{t_{k-1}}, \beta) + \epsilon^{-2} h_n^{-1} P_{v,k}(\tilde{\alpha})^\top [\sigma \sigma^\top]^{-1}(X_{t_{k-1}}, \beta) P_{v,k}(\tilde{\alpha}) \right\}, \]

Springer
and $\tilde{\theta}_{\varepsilon,n} = \tilde{\theta}_{\varepsilon,n}^{(1)}$ is defined as

$$U_{\varepsilon,n,v}^{(2)}(\tilde{\theta}_{\varepsilon,n}^{(1)}|\tilde{\theta}_{\varepsilon,n}^{(1)}) = \inf_{\beta \in \Theta_{\beta}} U_{\varepsilon,n,v}^{(2)}(\beta|\tilde{\theta}_{\varepsilon,n}^{(1)}).$$

**Step 3.** Set

$$U_{\varepsilon,n,v}^{(3)}(\alpha | \tilde{\beta}) := \varepsilon^{-2} h_{n}^{-1} \sum_{k=1}^{n} P_{v,k}(\alpha) \top [\sigma \sigma^{\top}]^{-1} (X_{t_{k}}^{n} - 1, \tilde{\beta}) P_{v,k}(\alpha),$$

and $\tilde{\alpha}_{\varepsilon,n} = \tilde{\alpha}_{\varepsilon,n}^{(2)}$ is defined as

$$U_{\varepsilon,n,v}^{(3)}(\tilde{\alpha}_{\varepsilon,n}^{(2)}|\tilde{\alpha}_{\varepsilon,n}^{(2)}) = \inf_{\alpha \in \Theta_{\alpha}} U_{\varepsilon,n,v}^{(3)}(\alpha|\tilde{\alpha}_{\varepsilon,n}^{(2)}).$$

In Step 1, we have the following asymptotic properties.

**Lemma 1** Assume [A1]–[A3], [A4]-(i) and [B]. Then it holds that

$$\varepsilon^{-1}(\tilde{\alpha}_{\varepsilon,n}^{(1)} - \alpha_{0}) = O_{p}(1).$$

In particular,

$$\varepsilon^{-1}(\tilde{\alpha}_{\varepsilon,n}^{(1)} - \alpha_{0}) \xrightarrow{d} N_{p}(0, J_{b}(\alpha_{0})^{-1} K_{b}(\theta_{0}) J_{b}(\alpha_{0})^{-1})$$

as $\varepsilon \to 0$ and $n \to \infty$.

**Remark 2** In order to prove the asymptotic normality of the estimator in Theorem 1 below, we need to show that $\varepsilon^{-1}(\tilde{\alpha}_{\varepsilon,n}^{(1)} - \alpha_{0}) = O_{p}(1)$. Using $\tilde{\alpha}_{\varepsilon,n}^{(1)}$, we can estimate the drift parameter $\alpha$ independent of the diffusion parameter $\beta$ and it follows from Lemma 1 that this estimator has asymptotic normality. Note that this estimator $\tilde{\alpha}_{\varepsilon,n}^{(1)}$ is not asymptotic efficient in general.

The main result for Type I method is as follows.

**Theorem 1** Assume [A1]–[A4] and [B]. Then it follows that

$$\tilde{\theta}_{\varepsilon,n} \overset{p}{\to} \theta_{0}.$$

Moreover,

$$\left( \frac{\varepsilon^{-1}(\tilde{\alpha}_{\varepsilon,n}^{(1)} - \alpha_{0})}{\sqrt{n}(\tilde{\beta}_{\varepsilon,n}^{(1)} - \beta_{0})} \right) \overset{d}{\to} N_{p+q}(0, I(\theta_{0})^{-1})$$

as $\varepsilon \to 0$ and $n \to \infty$. 

 Springer
Remark 3 From the viewpoint of asymptotic theory, the \( k \)-step estimators \( \beta_{\epsilon, n}^{(k+2)} \) and \( \alpha_{\epsilon, n}^{(k+1)} \) obtained by \( k \)-times repeating Steps 2 and 3 of Type I method are asymptotically equivalent to \( \alpha_{\epsilon, n}^{(2)} \) and \( \beta_{\epsilon, n}^{(1)} \) obtained by applying Steps 2 and 3 only once. The algorithm of Type I method will lead to a dramatic saving in computation time compared to the \( k \)-step estimators \( \beta_{\epsilon, n}^{(k+2)} \) and \( \alpha_{\epsilon, n}^{(k+1)} \) for a large \( k \). However, from the viewpoint of nonasymptotic quality of the estimator, we expect that the \( k \)-step estimator \( \beta_{\epsilon, n}^{(k+2)} \) and \( \alpha_{\epsilon, n}^{(k+1)} \) improve Type I estimators \( \alpha_{\epsilon, n}^{(2)} \) and \( \beta_{\epsilon, n}^{(1)} \). It seems that there is a tradeoff between numerical cost and statistical quality (i.e. few iterations vs. higher precision). This is future work.

3.2 Type II estimator

In the Type II method, we divide the Step 1 of Type I method into several steps. By this split, we expect that the estimators in Type II method are computed more quickly than those in the Type I method from the viewpoint of computation time. For example, see Table 8 in Sect. 5.

Step 1. Set

\[
V_{\epsilon, n}^{(1)}(\alpha) := \epsilon^{-2} h_{\epsilon}^{-1} \sum_{k=1}^{n} P_{1,k}(\alpha)^{\top} P_{1,k}(\alpha),
\]

and \( \hat{\alpha}_{\epsilon, n}^{(1)} \) is defined as

\[
V_{\epsilon, n}^{(1)}(\hat{\alpha}_{\epsilon, n}^{(1)}) = \inf_{\alpha \in \Theta} V_{\epsilon, n}^{(1)}(\alpha).
\]

Step l=2 to v. Set

\[
V_{\epsilon, n}^{(l)}(\alpha|\bar{\alpha}) := \epsilon^{-2} h_{\epsilon}^{-1} \sum_{k=1}^{n} (P_{1,k}(\alpha) - Q_{l,k}(\bar{\alpha}))^{\top} (P_{1,k}(\alpha) - Q_{l,k}(\bar{\alpha})),
\]

and \( \hat{\alpha}_{\epsilon, n}^{(l)} \) is defined as

\[
V_{\epsilon, n}^{(l)}(\hat{\alpha}_{\epsilon, n}^{(l)}|\hat{\alpha}_{\epsilon, n}^{(l-1)}) = \inf_{\alpha \in \Theta} V_{\epsilon, n}^{(l)}(\alpha|\hat{\alpha}_{\epsilon, n}^{(l-1)}).
\]

Step v+1. Set

\[
V_{\epsilon, n}^{(v+1)}(\beta|\bar{\alpha}) := \sum_{k=1}^{n} \left\{ \log \text{det}\{\sigma \sigma^{\top}\}(X_{\epsilon, n}^{-1}, \beta) + \epsilon^{-2} h_{\epsilon}^{-1} P_{v,k}(\bar{\alpha})^{\top} \left[ \sigma \sigma^{\top}\right]^{-1}(X_{\epsilon, n}^{-1}, \beta) P_{v,k}(\bar{\alpha}) \right\},
\]
and $\hat{\beta}_{n} = \hat{\beta}_{n}^{(1)}$ is defined as
\[
V_{\varepsilon,n}^{(v+1)}(\hat{\beta}_{n}, \hat{\alpha}_{n}) = \inf_{\beta \in \Theta_{\varepsilon}} V_{\varepsilon,n}^{(v+1)}(\beta | \hat{\alpha}_{n}).
\]

**Step v+2.** Set
\[
V_{\varepsilon,n}^{(v+2)}(\alpha | \hat{\alpha}_{n}, \hat{\beta}_{n}) := \varepsilon^{-2} h_{n}^{-1} \sum_{k=1}^{n} \left( P_{1,k}(\alpha) - Q_{v,k}(\hat{\alpha}) \right)^{\top} \left( \sigma \sigma^{\top} \right)^{-1} \left( X_{t_{v},k} - \hat{\beta} \right) \left( P_{1,k}(\alpha) - Q_{v,k}(\hat{\alpha}) \right),
\]

and $\hat{\alpha}_{n} = \hat{\alpha}_{n}^{(v+1)}$ is defined as
\[
V_{\varepsilon,n}^{(v+2)}(\hat{\alpha}_{n}, \hat{\beta}_{n}) = \inf_{\alpha \in \Theta_{\varepsilon}} V_{\varepsilon,n}^{(v+2)}(\alpha | \hat{\alpha}_{n}).
\]

In Step v, we have the following asymptotic properties.

**Lemma 2** Assume [A1]–[A3], [A4]-(i) and [B]. Then it holds that
\[
\varepsilon^{-1} (\hat{\alpha}_{n} - \alpha_{0}) = O_{p}(1).
\]

In particular,
\[
\varepsilon^{-1} (\hat{\alpha}_{n} - \alpha_{0}) \overset{d}{\to} N_p(0, J_{b}(\alpha_{0})^{-1} K_{b}(\theta_{0}) J_{b}(\alpha_{0})^{-1})
\]
as $\varepsilon \to 0$ and $n \to \infty$.

The main result for the Type II method is as follows.

**Theorem 2** Assume [A1]–[A4] and [B]. Then it follows that
\[
\hat{\beta}_{n} \overset{p}{\to} \theta_{0}.
\]

Moreover,
\[
\left( \varepsilon^{-1} (\hat{\alpha}_{n} - \alpha_{0}) \right) \overset{d}{\to} N_{p+q}(0, I(\theta_{0})^{-1})
\]
as $\varepsilon \to 0$ and $n \to \infty$.

### 3.3 Remarks for adaptive estimation

The Type I and Type II estimators work well for any $\rho > 0$. However, if we consider the case $\rho < \frac{1}{2}$, that is, the case $\varepsilon^{-1} / \sqrt{n} \to 0$, it is natural that we first estimate diffusion parameter instead of drift parameter. Therefore, we introduce another adaptive method which estimates diffusion parameter first:
Step 1. Set

\[ W^{(1)}_{\varepsilon,n}(\beta) := \sum_{k=1}^{n} \left\{ \log \det[\sigma \sigma^\top](X_{t_{k-1}}^{n}, \beta) + \varepsilon^{-2} h_n^{-1}(X_{t_{k}}^{n} - X_{t_{k-1}}^{n})^\top [\sigma \sigma^\top]^{-1}(X_{t_{k}}^{n}, \beta)(X_{t_{k}}^{n} - X_{t_{k-1}}^{n}) \right\}, \]

and \( \hat{\beta}_{\varepsilon,n} \) is defined as

\[ W^{(1)}_{\varepsilon,n}(\hat{\beta}_{\varepsilon,n}) = \inf_{\beta \in \Theta_1} W^{(1)}_{\varepsilon,n}(\beta). \]

Step 2. Set

\[ W^{(2)}_{\varepsilon,n}(\alpha|\hat{\beta}) := \varepsilon^{-2} h_n^{-1} \sum_{k=1}^{n} P_{1,k}(\alpha)^\top [\sigma \sigma^\top]^{-1}(X_{t_{k-1}}^{n}, \hat{\beta}) P_{1,k}(\alpha), \]

and \( \hat{\alpha}_{\varepsilon,n} \) is defined as

\[ W^{(2)}_{\varepsilon,n}(\hat{\alpha}_{\varepsilon,n}|\hat{\beta}_{\varepsilon,n}) = \inf_{\alpha \in \Theta_\alpha} W^{(2)}_{\varepsilon,n}(\alpha|\hat{\beta}_{\varepsilon,n}). \]

These adaptive estimators also have asymptotic normality with the same asymptotic variance in Theorems 1 and 2.

**Proposition 1** Assume \([A1]–[A4]\) and \([B]\) with \( \rho < \frac{1}{2} \). Then it follows that

\[ \hat{\theta}_{\varepsilon,n} := (\hat{\alpha}_{\varepsilon,n}, \hat{\beta}_{\varepsilon,n}) \xrightarrow{p} \theta_0. \]

Moreover,

\[ \left( \varepsilon^{-1}(\hat{\alpha}_{\varepsilon,n} - \alpha_0), \sqrt{n}(\hat{\beta}_{\varepsilon,n} - \beta_0) \right) \xrightarrow{d} N_{p+q}(0, I(\theta_0)^{-1}) \]

as \( \varepsilon \to 0 \) and \( n \to \infty \).

This proof is similar to that of Theorem 1. We omit the proof.

Next, we consider the following special cases and show that we can also apply our adaptive methods.

(i) Case of \( \sigma(x, \beta) = \sigma(x) \) (see, for example, Model 3 in Sect. 5.3 below) : We fix some algorithm and estimate only the drift parameter \( \alpha \).

Type I : Calculate only the estimator in Step 3, and output \( \hat{\alpha}_{\varepsilon,n} \).

Type II : Compute the estimators in Step 1 to \( v \), and skip Step \( v + 1 \). Next, run Step \( v + 2 \). Finally, output \( \hat{\alpha}_{\varepsilon,n} \).
From these algorithms, estimators $\tilde{a}_{\epsilon,n}$ and $\tilde{a}_{\epsilon,n}$ are asymptotically efficient, respectively.

(ii) Case of $\alpha = \beta$ (see, for example, Model 2 in Sect. 5.2 below): We estimate only drift or diffusion parameter. This choice depends on the speed of convergence for both parameters.

(a) Case of $\epsilon^{-1}/\sqrt{n} \to \infty$: We estimate only drift parameter.

Type I: Calculate the estimator in Step 1, and skip Step 2. Next, run Step 3 with $\tilde{\beta}_{\epsilon,n} = (1)$. Finally, output $\tilde{a}_{\epsilon,n}$.

Type II: Compute the estimators in Step 1 to $v$, and skip Step $v + 1$. Next, calculate Step $v + 2$ with $\tilde{\beta}_{\epsilon,n} = (v)$. Lastly, output $\tilde{a}_{\epsilon,n}$.

By these methods, estimators $\tilde{a}_{\epsilon,n}$ and $\tilde{a}_{\epsilon,n}$ are asymptotically efficient, respectively.

(b) Case of $\epsilon^{-1}/\sqrt{n} \to 0$: We utilize the adaptive method for $\rho < \frac{1}{2}$ and estimate only diffusion parameter: Calculate only Step 1 in Sect. 3.3, and output $\tilde{\beta}_{\epsilon,n}$.

(c) Case of $\epsilon^{-1}/\sqrt{n} \to M \neq \{0, \infty\}$: In this case, we need to consider the contrast functions for both drift and diffusion simultaneously. Therefore, proposed adaptive estimators cannot be used. For estimating method in this case, see Uchida (2003).

4 Adaptive tests

In this section, we consider the following set of parametric test problems:

\[
\begin{align*}
H_0^{(1)} : \alpha_1 = \ldots = \alpha_r &= 0, \\
H_1^{(1)} : \text{not } H_0^{(1)}, \\
H_0^{(2)} : \beta_1 = \ldots = \beta_s &= 0, \\
H_1^{(2)} : \text{not } H_0^{(2)}. \\
\end{align*}
\]  

(4)

where $1 \leq r \leq p$, $1 \leq s \leq q$. These set of tests give more information about parameters than the joint test and can provide the following four interpretations: (i) $H_0^{(1)}$ is rejected and $H_0^{(2)}$ is rejected; (ii) $H_0^{(1)}$ is rejected and $H_0^{(2)}$ is not rejected; (iii) $H_0^{(1)}$ is not rejected and $H_0^{(2)}$ is rejected; (iv) $H_0^{(1)}$ is not rejected and $H_0^{(2)}$ is not rejected. In order to construct test statistics for the above tests, we define restricted parameter spaces $\Theta_{\alpha}^{H_0}$ and $\Theta_{\beta}^{H_0}$ as $\Theta_{\alpha}^{H_0} := \{\alpha \in \Theta_{\alpha} | \alpha \text{ satisfies } H_0^{(1)}\}$, $\Theta_{\beta}^{H_0} := \{\beta \in \Theta_{\beta} | \beta \text{ satisfies } H_0^{(2)}\}$ and $\Theta_{H_0} := \Theta_{\alpha}^{H_0} \times \Theta_{\beta}^{H_0}$, respectively. Under [A4], we set

\[
J_{b}^{-1}(\alpha_0) = \begin{pmatrix} J_{b,1}^{-1}(i,j) (\beta_0) & J_{b,2}^{-1}(i,j) (\beta_0) \\ J_{b,2}^{-1}(i,j) (\beta_0) & J_{b,3}^{-1}(i,j) (\beta_0) \end{pmatrix}, \\
J_{\sigma}^{-1}(\beta_0) = \begin{pmatrix} I_{\sigma,1}^{-1}(i,j) (\beta_0) & I_{\sigma,2}^{-1}(i,j) (\beta_0) \\ I_{\sigma,2}^{-1}(i,j) (\beta_0) & I_{\sigma,3}^{-1}(i,j) (\beta_0) \end{pmatrix},
\]
and define \((p \times p)\)-matrix \(G^r_1(\alpha_0)\) and \((q \times q)\)-matrix \(G^s_2(\beta_0)\) as

\[
G^r_1(\alpha_0) = \begin{pmatrix} 0 & 0 \\ J_{b,3}(\beta_0) & 0 \end{pmatrix}, \quad G^s_2(\beta_0) = \begin{pmatrix} 0 & 0 \\ 0 & I_{\sigma,3}(\beta_0) \end{pmatrix},
\]

respectively. Let \(Z\) be a \(p\)-dimensional random vector which has the normal distribution \(N_p(0, K_b(\theta_0))\) and \(\pi_r\) be the distribution of the random variable \(Z^\top \left( J_{b}^{-1}(\alpha_0) - G^r_1(\alpha_0) \right) Z\). Moreover, we define an optimal parameter under null hypothesis \((\alpha_0^{H_0}, \beta_0^{H_0})\) as

\[
U_1(\alpha_0^{H_0}; \alpha_0) := \inf_{\alpha \in \Theta_{\alpha}^{H_0}} U_1(\alpha; \alpha_0), \quad U_2(\beta_0^{H_0}; \beta_0) := \inf_{\beta \in \Theta_{\beta}^{H_0}} U_2(\beta; \beta_0),
\]

where,

\[
U_1(\alpha, \alpha_0) := \int_0^1 \left( b(X_s^0, \alpha_0) - b(X_s^0, \alpha) \right)^\top \left( b(X_s^0, \alpha_0) - b(X_s^0, \alpha) \right) ds, \quad (5)
\]

\[
U_2(\beta, \beta_0) := \int_0^1 \left\{ \log \det \left[ (\sigma \sigma^\top) (X_s^0, \beta) (\sigma \sigma^\top)^{-1} (X_s^0, \beta_0) \right] + \text{tr} \left[ (\sigma \sigma^\top)^{-1} (X_s^0, \beta) (\sigma \sigma^\top)^{-1} (X_s^0, \beta_0) \right] \right\} ds. \quad (6)
\]

It is remarked that \(\alpha_0^{H_0} \neq \alpha_0\) and \(\beta_0^{H_0} \neq \beta_0\) under alternatives by the definition of \((\alpha_0^{H_0}, \beta_0^{H_0})\). Using restricted optimal parameters \((\alpha_0^{H_0}, \beta_0^{H_0})\), we make the following condition.

[C] (i) For any \(\delta_1 > 0\),

\[
\inf_{\alpha \in \Theta_{\alpha}^{H_0} : |\alpha - \alpha_0^{H_0}| \geq \delta_1} \left( U_1(\alpha, \alpha_0) - U_1(\alpha_0^{H_0}, \alpha_0) \right) > 0.
\]

(ii) For any \(\delta_2 > 0\),

\[
\inf_{\beta \in \Theta_{\beta}^{H_0} : |\beta - \beta_0^{H_0}| \geq \delta_2} \left( U_2(\beta, \beta_0) - U_2(\beta_0^{H_0}, \beta_0) \right) > 0.
\]

Note that \((\alpha_0^{H_0}, \beta_0^{H_0})\) exists since \(\Theta_{\alpha}^{H_0}\) and \(\Theta_{\beta}^{H_0}\) are compact sets and \(U_1\) and \(U_2\) are continuous on \(\Theta_{\alpha}^{H_0}\) and \(\Theta_{\beta}^{H_0}\), respectively. Moreover, it follows from [C] that \((\alpha_0^{H_0}, \beta_0^{H_0})\) is unique.

In this paper, we propose two kinds of likelihood ratio type test statistics (Type I, Type II) for each test using adaptive estimators.

### 4.1 Type I tests

First, we construct the test statistics utilizing the Type I estimator. The restricted Type I estimators \((\tilde{\alpha}_{\varepsilon,n}^{(1)}; H_0), \tilde{\beta}_{\varepsilon,n}^{H_0})\) is defined as

\[
\tilde{\alpha}_{\varepsilon,n}^{(1)} := \arg \min_{\alpha \in \Theta_{\alpha}^{H_0}} U_{\varepsilon,n,1}(\alpha), \quad \tilde{\beta}_{\varepsilon,n}^{H_0} := \arg \min_{\beta \in \Theta_{\beta}^{H_0}} U_{\varepsilon,n,1}(\beta|\tilde{\alpha}_{\varepsilon,n}^{(1)}),
\]

\(\otimes\) Springer
and likelihood ratio type test statistics \((\tilde{\Lambda}_n^{(1)}, \tilde{\Lambda}_n^{(2)})\) as

\[
\tilde{\Lambda}_n^{(1)} := U_{\hat{\epsilon},n,v}^{(1)} \left( \tilde{\alpha}_{\hat{\epsilon},n}^{(1)} \right) - U_{\hat{\epsilon},n,v}^{(1)} \left( \hat{\alpha}_{\hat{\epsilon},n}^{(1)} \right), \\
\tilde{\Lambda}_n^{(2)} := U_{\hat{\epsilon},n,v}^{(2)} \left( \tilde{\beta}_{\hat{\epsilon},n} \right) - U_{\hat{\epsilon},n,v}^{(2)} \left( \hat{\beta}_{\hat{\epsilon},n} \right).
\]

Note that \(\tilde{\Lambda}_n^{(1)}\) enable us to test for drift parameter \(\alpha\) without estimating the diffusion parameter \(\beta\). The following theorem gives asymptotic distributions of these test statistics under null hypothesis. From this theorem, we can calculate rejection regions for each test and conduct tests (4).

**Theorem 3** Assume \([A1]–[A4]\) and \([B]\). Then it follows that

\[
\tilde{\Lambda}_n^{(1)} \overset{d}{\to} \pi_r \quad (\text{under } H_0^{(1)}), \quad \tilde{\Lambda}_n^{(2)} \overset{d}{\to} \chi^2_s \quad (\text{under } H_0^{(2)})
\]

as \(\epsilon \to 0\) and \(n \to \infty\).

The followings ensure consistency of Type I adaptive tests. For a distribution \(\nu\) and \(\delta \in (0, 1)\), we denote \(\nu(\delta)\) as the upper \(\delta\) point of \(\nu\).

**Theorem 4** Assume \([A1]–[A4], [B]\) and \([C]\). Then it follows that for any \(\delta \in (0, 1)\),

\[
P(\tilde{\Lambda}_n^{(1)} \geq \pi_r(\delta)) \to 1 \quad (\text{under } H_1^{(1)}), \quad P(\tilde{\Lambda}_n^{(2)} \geq \chi^2_s(\delta)) \to 1 \quad (\text{under } H_1^{(2)})
\]

as \(\epsilon \to 0\) and \(n \to \infty\).

### 4.2 Type II tests

Next, we consider test statistics utilizing the Type II estimator. In a similar way to Type I tests, we define restricted Type II estimators \((\hat{\alpha}_{\hat{\epsilon},n}^{(v)}, \hat{\beta}_{\hat{\epsilon},n}^{(v)})\) as

\[
\hat{\alpha}_{\hat{\epsilon},n}^{(v)} := \arg\min_{\alpha \in \Theta_0} V_{\hat{\epsilon},n}^{(v)} (\alpha | \hat{\alpha}_{\hat{\epsilon},n}^{(v-1)}) , \quad \hat{\beta}_{\hat{\epsilon},n}^{(v)} := \arg\min_{\beta \in \Theta_0} V_{\hat{\epsilon},n}^{(v+1)} (\beta | \hat{\alpha}_{\hat{\epsilon},n}^{(v)}),
\]

and likelihood ratio type test statistics \((\hat{\Lambda}_n^{(1)}, \hat{\Lambda}_n^{(2)})\) as

\[
\hat{\Lambda}_n^{(1)} := V_{\hat{\epsilon},n}^{(v)} \left( \hat{\alpha}_{\hat{\epsilon},n}^{(v)} | \hat{\alpha}_{\hat{\epsilon},n}^{(v-1)} \right) - V_{\hat{\epsilon},n}^{(v)} \left( \hat{\alpha}_{\hat{\epsilon},n}^{(v)} | \hat{\alpha}_{\hat{\epsilon},n}^{(v-1)} \right), \\
\hat{\Lambda}_n^{(2)} := V_{\hat{\epsilon},n}^{(v+1)} \left( \hat{\beta}_{\hat{\epsilon},n}^{(v)} | \hat{\alpha}_{\hat{\epsilon},n}^{(v)} \right) - V_{\hat{\epsilon},n}^{(v+1)} \left( \hat{\beta}_{\hat{\epsilon},n}^{(v)} | \hat{\alpha}_{\hat{\epsilon},n}^{(v)} \right).
\]

Asymptotic distributions of these test statistics under null hypothesis and consistency of tests can be proved in an analogous manner to theorem 3 and 4. We omit the proofs of the following theorems.

**Theorem 5** Assume \([A1]–[A4]\) and \([B]\). Then it follows that

\[
\hat{\Lambda}_n^{(1)} \overset{d}{\to} \pi_r \quad (\text{under } H_0^{(1)}), \quad \hat{\Lambda}_n^{(2)} \overset{d}{\to} \chi^2_s \quad (\text{under } H_0^{(2)})
\]
as $\varepsilon \to 0$ and $n \to \infty$.

**Theorem 6** Assume [A1]–[A4], [B] and [C]. Then it follows that for any $\delta \in (0, 1)$,

$$P(\hat{\Lambda}_n^{(1)} \geq \pi_r(\delta)) \to 1 \quad \text{(under } H_1^{(1)}) , \quad P(\hat{\Lambda}_n^{(2)} \geq \chi_r^2(\delta)) \to 1 \quad \text{(under } H_1^{(2)})$$

as $\varepsilon \to 0$ and $n \to \infty$.

### 4.3 Remarks for adaptive tests

We introduce the adaptive test statistics for $\rho < \frac{1}{2}$. Using adaptive estimators proposed in Sect. 3.3 $(\tilde{\alpha}_{\varepsilon,n}, \tilde{\beta}_{\varepsilon,n})$, we define restricted estimators $(\tilde{\alpha}_{\varepsilon,n}^{H_0}, \tilde{\beta}_{\varepsilon,n}^{H_0})$ as

$$\tilde{\alpha}_{\varepsilon,n}^{H_0} := \arg\min_{\alpha \in \Theta_\alpha^{H_0}} W_{\varepsilon,n}^{(2)}(\alpha|\tilde{\beta}_{\varepsilon,n}), \quad \tilde{\beta}_{\varepsilon,n}^{H_0} := \arg\min_{\beta \in \Theta_\beta^{H_0}} W_{\varepsilon,n}^{(1)}(\beta),$$

and likelihood ratio type test statistics $\tilde{\Lambda}_n^{(\alpha)}$ and $\tilde{\Lambda}_n^{(\beta)}$ are defined as

$$\tilde{\Lambda}_n^{(\alpha)} := W_{\varepsilon,n}^{(2)}(\tilde{\alpha}_{\varepsilon,n}^{H_0}|\tilde{\beta}_{\varepsilon,n}) - W_{\varepsilon,n}^{(2)}(\tilde{\alpha}_{\varepsilon,n}|\tilde{\beta}_{\varepsilon,n}), \quad \tilde{\Lambda}_n^{(\beta)} := W_{\varepsilon,n}^{(1)}(\tilde{\beta}_{\varepsilon,n}^{H_0}) - W_{\varepsilon,n}^{(1)}(\tilde{\beta}_{\varepsilon,n}).$$

In this case, we first test the diffusion parameter $\beta$. Asymptotic distributions of these test statistics under null hypothesis and consistency are shown. Since $\tilde{\alpha}_{\varepsilon,n}$ is asymptotically efficient, the test statistic $\tilde{\Lambda}_n^{(\alpha)}$ converges to the chi-squared distribution under null hypothesis.

**Proposition 2** Assume [A1]–[A4] and [B] with $\rho < \frac{1}{2}$. Then it follows that

$$\tilde{\Lambda}_n^{(\alpha)} \overset{d}{\to} \chi_r^2 \quad \text{(under } H_0^{(1)}), \quad \tilde{\Lambda}_n^{(\beta)} \overset{d}{\to} \chi_s^2 \quad \text{(under } H_0^{(2)})$$

as $\varepsilon \to 0$ and $n \to \infty$.

**Proposition 3** Assume [A1]–[A4], [B] with $\rho < \frac{1}{2}$ and [C]. Then it follows that for any $\delta \in (0, 1)$,

$$P(\hat{\Lambda}_n^{(\alpha)} \geq \chi_r^2(\delta)) \to 1 \quad \text{(under } H_1^{(1)}) , \quad P(\hat{\Lambda}_n^{(\beta)} \geq \chi_s^2(\delta)) \to 1 \quad \text{(under } H_1^{(2)})$$

as $\varepsilon \to 0$ and $n \to \infty$.

We omit the proofs of these propositions.

In the testing method for $\alpha$ with the chi-squared distribution in general $\rho > 0$, we can use the asymptotic efficient estimator $\tilde{\alpha}_{\varepsilon,n}$ in Type I or $\hat{\alpha}_{\varepsilon,n}$ in Type II. If the restricted estimators $\tilde{\alpha}_{\varepsilon,n}^{H_0}$ and $\hat{\beta}_{\varepsilon,n}^{H_0}$ are defined as

$$\tilde{\alpha}_{\varepsilon,n}^{H_0} := \arg\min_{\alpha \in \Theta_\alpha^{H_0}} U_{\varepsilon,n,\gamma}(\alpha|\tilde{\beta}_{\varepsilon,n}), \quad \hat{\alpha}_{\varepsilon,n}^{H_0} := \arg\min_{\alpha \in \Theta_\alpha^{H_0}} V_{\varepsilon,n}^{(v+2)}(\alpha|\tilde{\alpha}_{\varepsilon,n}^{(v)}, \hat{\beta}_{\varepsilon,n}).$$
and the likelihood ratio type test statistics $\tilde{\Lambda}_n^{(\alpha)}$ and $\hat{\Lambda}_n^{(\alpha)}$ are defined as

$$\tilde{\Lambda}_n^{(\alpha)} := U_{e,n,v}^{(3)} \left( \tilde{\alpha}_{e,n}^{H_0} | \tilde{\beta}_{e,n}^{H_0} \right) - U_{e,n,v}^{(3)} \left( \tilde{\alpha}_{e,n} | \tilde{\beta}_{e,n} \right),$$

$$\hat{\Lambda}_n^{(\alpha)} := V_{e,n}^{(v+2)} \left( \hat{\alpha}_{e,n}^{H_0} | \hat{\beta}_{e,n}^{H_0} \right) - V_{e,n}^{(v+2)} \left( \hat{\alpha}_{e,n} | \hat{\beta}_{e,n} \right),$$

then $\tilde{\Lambda}_n^{(\alpha)}$ and $\hat{\Lambda}_n^{(\alpha)}$ converge to the chi-squared distribution with $r$ degree of freedom under null hypothesis. In the case of $\alpha = \beta$ except for (c) in Sect. 3.3, we can also conduct the chi-squared test for drift or diffusion parameter using the estimators introduced in Sect. 3.3a and b.

5 Examples and simulations

5.1 Model 1 (Case of estimating both drift and diffusion parameters)

First, we examine the asymptotic performance of Theorems 1 and 2. Consider the following two-dimensional model:

$$dX_t = \begin{cases} -\alpha_1 X_{t,1} + 2 \cos(1 + \alpha_2 X_{t,2}) \\ 2 \sin(1 + \alpha_3 X_{t,1}) - \alpha_4 X_{t,2} \end{cases} dt$$

$$+ \varepsilon \begin{pmatrix} \beta_1(1 + X_{t,1}^2)^{-1} \\ 0.1 \beta_2(1 + X_{t,2}^2)^{-1} \end{pmatrix} dW_t, \quad t \in [0, 1] \tag{7}$$

where $\theta = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta_1, \beta_2)$ are unknown parameters. The true parameter values are $\theta_0 = (3, 6, 5, 4, 1, 0.5)$, and the parameter space is assumed to be $\Theta = [0.01, 50]^6$. We estimate these parameters by the joint estimation method in Gloter and Sørensen (2009), the Type I method, and the Type II method. We choose the initial parameters $\theta = \theta_0$ or $\theta = (6, 4, 6, 8, 2, 2)$ and treat the case of $(\varepsilon, n) = (0.05, 100), (0.01, 100), (0.01, 1000)$. For the balance condition, we set $\rho = 1$, that is, the approximation degree $v = 2$. In the simulation, optim() is used with the "L-BFGS-B" method in R Language, and 10000 independent sample paths are generated by the Euler-Maruyama scheme with discretization step $1/1000$. The datasets generated during the current study are available from the corresponding author upon request.

Tables 1 and 2 show the simulation results of parameter estimation with the two choice of the initial parameter values $\theta_{\text{init}} = \theta_0 = (3, 6, 5, 4, 1, 0.5)$ or $\theta_{\text{init}} = (6, 4, 6, 8, 2, 1)$. In Table 1, we see that all of the estimation methods have good performances and there is no notable difference among the three types of methods. In Table 2, however, the joint estimation method has considerable biases while the Type I and Type II methods have good performances. There seems to be various
Table 1  Mean (S.D.) of the simulated values with the true initial parameters $\theta_{\text{init}} = \theta_0$

| $\varepsilon$ | n  | Method | $\alpha_1$ (3) | $\alpha_2$ (6) | $\alpha_3$ (5) | $\alpha_4$ (4) | $\beta_1$ (1) | $\beta_2$ (0.5) |
|--------------|----|--------|----------------|----------------|----------------|----------------|----------------|----------------|
| 0.05         | 100| Joint  | 3.0104         | 6.0071         | 5.0005         | 4.0003         | 0.9731         | 0.4863         |
|              |    |        | (0.0862)       | (0.0892)       | (0.0360)       | (0.0689)       | (0.0713)       | (0.0360)       |
|              |    | Type I | 3.0104         | 6.0071         | 5.0005         | 4.0003         | 0.9755         | 0.4868         |
|              |    |        | (0.0862)       | (0.0892)       | (0.0360)       | (0.0689)       | (0.0715)       | (0.0360)       |
|              |    | Type II| 3.0092         | 6.0080         | 5.0004         | 4.0007         | 0.9755         | 0.4868         |
|              |    |        | (0.0862)       | (0.0895)       | (0.0361)       | (0.0690)       | (0.0715)       | (0.0360)       |
| 0.01         | 100| Joint  | 3.0110         | 6.0085         | 5.0008         | 3.9993         | 0.9777         | 0.4877         |
|              |    |        | (0.0172)       | (0.0178)       | (0.0072)       | (0.0138)       | (0.0176)       | (0.0360)       |
|              |    | Type I | 3.0110         | 6.0085         | 5.0008         | 3.9993         | 0.9803         | 0.4881         |
|              |    |        | (0.0172)       | (0.0178)       | (0.0072)       | (0.0138)       | (0.0720)       | (0.0360)       |
|              |    | Type II| 3.0099         | 6.0095         | 5.0006         | 3.9997         | 0.9846         | 0.4884         |
|              |    |        | (0.0172)       | (0.0179)       | (0.0072)       | (0.0138)       | (0.0720)       | (0.0360)       |
| 0.01         | 1000| Joint | 3.0002         | 6.0015         | 5.0003         | 4.0008         | 0.9976         | 0.4986         |
|              |    |        | (0.0171)       | (0.0176)       | (0.0072)       | (0.0137)       | (0.0226)       | (0.0114)       |
|              |    | Type I | 3.0002         | 6.0015         | 5.0003         | 4.0008         | 0.9978         | 0.4986         |
|              |    |        | (0.0171)       | (0.0176)       | (0.0072)       | (0.0137)       | (0.0227)       | (0.0114)       |
|              |    | Type II| 3.0002         | 6.0015         | 5.0003         | 4.0009         | 0.9978         | 0.4986         |
|              |    |        | (0.0171)       | (0.0176)       | (0.0072)       | (0.0137)       | (0.0227)       | (0.0114)       |

Table 2  Mean (S.D.) of the simulated values with the initial parameters $\theta_{\text{init}} = (6, 4, 6, 8, 2, 1)$

| $\varepsilon$ | n  | Method | $\alpha_1$ (3) | $\alpha_2$ (6) | $\alpha_3$ (5) | $\alpha_4$ (4) | $\beta_1$ (1) | $\beta_2$ (0.5) |
|--------------|----|--------|----------------|----------------|----------------|----------------|----------------|----------------|
| 0.05         | 100| Joint  | 2.5696         | 48.172         | 42.142         | 3.8555         | 5.0627         | 3.8834         |
|              |    |        | (0.1447)       | (2.3443)       | (2.7301)       | (0.4780)       | (0.2111)       | (0.1693)       |
|              |    | Type I | 3.0104         | 6.0071         | 5.0005         | 4.0003         | 0.9755         | 0.4868         |
|              |    |        | (0.0862)       | (0.0892)       | (0.0360)       | (0.0689)       | (0.0715)       | (0.0360)       |
|              |    | Type II| 3.0070         | 6.1873         | 5.0003         | 4.0002         | 1.0008         | 0.4870         |
|              |    |        | (0.0927)       | (2.1529)       | (0.0361)       | (0.0694)       | (0.3134)       | (0.0361)       |
| 0.01         | 100| Joint  | 2.4918         | 49.999         | 49.933         | 3.9127         | 24.812         | 20.631         |
|              |    |        | (0.0148)       | (0.0530)       | (0.1038)       | (0.0424)       | (0.1539)       | (0.1132)       |
|              |    | Type I | 3.0110         | 6.0085         | 5.0008         | 3.9993         | 0.9803         | 0.4881         |
|              |    |        | (0.0172)       | (0.0178)       | (0.0072)       | (0.0138)       | (0.0720)       | (0.0360)       |
|              |    | Type II| 3.0099         | 6.0095         | 5.0006         | 3.9997         | 0.9846         | 0.4884         |
|              |    |        | (0.0172)       | (0.0179)       | (0.0072)       | (0.0138)       | (0.0720)       | (0.0360)       |
| 0.01         | 1000| Joint | 2.5730         | 30.681         | 19.359         | 3.5553         | 6.6510         | 6.4561         |
|              |    |        | (0.0275)       | (1.9874)       | (3.7729)       | (0.5760)       | (0.0907)       | (0.5269)       |
|              |    | Type I | 3.0002         | 6.0015         | 5.0003         | 4.0008         | 0.9978         | 0.4986         |
|              |    |        | (0.0171)       | (0.0176)       | (0.0072)       | (0.0137)       | (0.0227)       | (0.0114)       |
|              |    | Type II| 3.0002         | 6.0015         | 5.0003         | 4.0009         | 0.9978         | 0.4986         |
|              |    |        | (0.0171)       | (0.0176)       | (0.0072)       | (0.0137)       | (0.0227)       | (0.0114)       |
causes of the biases of the joint estimation method in this simulation. For example, the asymptotic approximation may not applicable or the numerical solver may not approximate the solution well. It seems that the adaptive estimation methods are useful for decreasing the dimensions of the parameter for optimization from the viewpoint of computational cost. As stated in Remark 3, however, we expect that the $k$-step estimator $\hat{\beta}_{k,n}$ and $\hat{\alpha}_{n}^{(k+1)}$ improve the adaptive estimators from the viewpoint of nonasymptotic quality of the estimator. This is future work. Figures 1 and 2 calculate $(\varepsilon^{-1}(\tilde{\alpha}_{e,n} - \alpha_0), \sqrt{n}(\tilde{\beta}_{e,n} - \beta_0))$ for 10000 times and create histogram, empirical distribution, and Q–Q plot for each parameter. These figures show that each parameter’s estimator has asymptotic normality and its asymptotic variance equals $I(\theta_0)^{-1}$.

Next, we consider the following adaptive tests:

\[
\begin{align*}
H_0^{(1)} &: (\alpha_1, \alpha_4) = (3.0, 4.0), \\
H_1^{(1)} &: \text{not } H_0^{(1)}, \\
H_0^{(2)} &: (\beta_1, \beta_2) = (1.0, 0.5), \\
H_1^{(2)} &: \text{not } H_0^{(2)},
\end{align*}
\]

These tests derive the four kinds of results as follows:
Case 1. Neither $\alpha$ nor $\beta$ is rejected;
Case 2. $\alpha$ is not rejected, but $\beta$ is rejected;
Case 3. $\alpha$ is rejected, but $\beta$ is not rejected;
Case 4. Both $\alpha$ and $\beta$ are rejected.

We set true parameters $(\alpha_2^*, \alpha_4^*) = (6.0, 5.0)$ and choose true parameters $= (\alpha_1^*, \alpha_4^*, \beta_1^*, \beta_2^*)$ from $\{(3.0, 4.0, 1.0, 0.5), (3.0, 4.0, 1.1, 0.6), (3.1, 4.1, 1.0, 0.5), (3.1, 4.1, 1.1, 0.6)\}$, which corresponds to the true parameters of Cases 1-4. In this adaptive test simulation, we consider the cases of $(\varepsilon, n) = (0.05, 100), (0.01, 100), (0.01, 1000)$ and treat only Type I method. The rest of the settings are the same as in the simulation of the estimation above. Let the significance level denote $\delta = 0.05$ and each test is rejected when the realization of test statistic $\tilde{\Lambda}_n^{(1)}$ or $\tilde{\Lambda}_n^{(2)}$ is greater than $\pi_2(0.05)$ or $\chi_2^2(0.05)$, respectively. The simulation is repeated 10000 times.

Table 3 shows the number of counts of Cases 1-4 selected by the Type I adaptive test statistics for the statistical testing problems (8), where the true parameters $(\alpha_1^*, \alpha_4^*, \beta_1^*, \beta_2^*)$ correspond to each case. In all cases, the adaptive tests can judge the true case most often as $\varepsilon$ decreases and $n$ increases. Table 4 shows the empirical sizes and powers.
Table 3  Results of Type I adaptive tests

| True case | $\varepsilon$ | $n$ | Judgement | Case 1 | Case 2 | Case 3 | Case 4 |
|-----------|---------------|-----|-----------|--------|--------|--------|--------|
| Case 1    | 0.05          | 100 |           | 8879   | 690    | 401    | 30     |
|           | 0.01          | 100 |           | 8808   | 640    | 520    | 32     |
|           | 0.01          | 1000|           | 9052   | 500    | 422    | 26     |
| Case 2    | 0.05          | 100 |           | 1703   | 8866   | 81     | 350    |
|           | 0.01          | 100 |           | 1838   | 7610   | 110    | 442    |
|           | 0.01          | 1000|           | 0      | 9552   | 0      | 448    |
| Case 3    | 0.05          | 100 |           | 7075   | 556    | 2205   | 164    |
|           | 0.01          | 100 |           | 0      | 0      | 9328   | 672    |
|           | 0.01          | 1000|           | 0      | 0      | 9474   | 526    |
| Case 4    | 0.05          | 100 |           | 1364   | 6267   | 420    | 1949   |
|           | 0.01          | 100 |           | 0      | 0      | 1948   | 8052   |
|           | 0.01          | 1000|           | 0      | 0      | 10000  | 10000  |

Table 4  Empirical sizes and powers of Type I test statistics

| True case | $\varepsilon$ | $n$ | Testing parameter | $\alpha$ | $\beta$ |
|-----------|---------------|-----|-------------------|----------|---------|
| Case 1    | 0.05          | 100 |                   | 0.0431   | 0.0720  |
|           | 0.01          | 100 |                   | 0.0552   | 0.0672  |
|           | 0.01          | 1000|                   | 0.0448   | 0.0526  |
| Case 2    | 0.05          | 100 |                   | 0.0431   | 0.8216  |
|           | 0.01          | 100 |                   | 0.0552   | 0.8052  |
|           | 0.01          | 1000|                   | 0.0448   | 1.0000  |
| Case 3    | 0.05          | 100 |                   | 0.2369   | 0.0720  |
|           | 0.01          | 100 |                   | 1.0000   | 0.0672  |
|           | 0.01          | 1000|                   | 1.0000   | 0.0526  |
| Case 4    | 0.05          | 100 |                   | 0.2369   | 0.8216  |
|           | 0.01          | 100 |                   | 1.0000   | 0.8052  |
|           | 0.01          | 1000|                   | 1.0000   | 1.0000  |

For each parametric test in (8). When $(\alpha_1^*, \alpha_4^*, \beta_1^*, \beta_2^*) = (3.0, 4.0, 1.0, 0.5)$, that is Case 1, the empirical sizes of the test statistics for both $\alpha$ and $\beta$ are obtained in Table 4. When $(\alpha_1^*, \alpha_4^*, \beta_1^*, \beta_2^*) = (3.0, 4.0, 1.1, 0.6)$, that is Case 2, the empirical size of the test statistics for $\alpha$ and the power of the test statistics for $\beta$ are obtained in Table 4. When $(\alpha_1^*, \alpha_4^*, \beta_1^*, \beta_2^*) = (3.1, 4.1, 1.0, 0.5)$, that is Case 3, the power of the test statistics for $\alpha$ and the empirical size of the test statistics for $\beta$ are obtained in Table 4. When $(\alpha_1^*, \alpha_4^*, \beta_1^*, \beta_2^*) = (3.1, 4.1, 1.1, 0.6)$, that is Case 4, the empirical sizes and powers of the test statistics for both $\alpha$ and $\beta$ are obtained in Table 4. The bold numbers in the table indicate that the results are close to the theoretical results, that is, the empirical sizes are close to the significance level $\delta = 0.05$ or the empirical powers take 1.0000. When the true case is Case 1, in particular, Fig. 3 shows the histogram...
and the empirical distribution of the Type I test statistics. This Figure implies that each test statistic follows the asymptotic distribution stated in Theorem 3 under null hypothesis.

5.2 Model 2 (Case of having the same parameter)

Second, we consider the case that the model has the same parameter in the drift and diffusion coefficients, that is, the case of \( \alpha = \beta \) in the SDE (1). For example, we introduce the following SIR model with the small diffusion coefficient proposed in Guy et al. (2014, 2015): let \( X_t = (S_t, I_t), \ t \in [0, T] \), and

\[
\begin{aligned}
    dS_t &= -\beta S_t I_t \ dt + \epsilon \sqrt{\beta S_t I_t} \ dW_{t,1}, \\
    dI_t &= (\beta S_t I_t - \gamma I_t) \ dt + \epsilon \left( -\sqrt{\beta S_t I_t} \ dW_{t,1} + \sqrt{\gamma I_t} \ dW_{t,2} \right), \\
    X_0 &= (s_0, i_0)^\top,
\end{aligned}
\]

(9)

where both of the drift and diffusion coefficients have \( \theta = (\beta, \gamma) \) as unknown parameters and \( (s_0, i_0) \in (0, 1)^2 \) is the fixed value. The SIR model describes the simple epidemic spread with the three mutually exclusive health states:

\[ Springer \]
Susceptible-Infectious-Removed from the infectious chain. The parameter $\beta$ implies the transmission rate and the parameter $\gamma$ implies the recovery rate. Moreover, the basic reproduction number $R_0 := \beta / \gamma$ implies the average number of secondary cases generated by one infected person. Therefore, when the value $R_0$ is greater than one, then the epidemic spreads and vice versa. We treat the cases of $\theta_0 = (1.2, 1.0)$ and $\theta_0 = (0.9, 1.0)$. The first case corresponds to $R_0 > 1$, and the second case corresponds to $R_0 < 1$. In this simulation, we set $\epsilon = 10^{-4}$ ($= 1/\sqrt{N}, \ N = 10^8$; population size), $(s_0, i_0) = (0.99999, 0.00001)$ and the parameter space $\Theta = [0.01, 100]^2$. In order to treat the 10 days, monthly and yearly data, we set $(n, T, h_n) = (10, 1, 1/10), (30, 1, 1/30), (360, 12, 1/30)$ and determine the balance coefficient $\rho = 4$. Therefore, the Type I and Type II methods described in Sect. 3.3(ii)-(a) are used for estimation. It is remarked that this model does not have to consider the initial parameter problem. This is because the model has only two parameters, and by using the uniform + optim() method stated in Sect. 5.3 below or Method U of Kaino and Uchida (2018), we can get the suitable initial estimator to do the parameters optimization. Therefore, we set that the initial parameter is the true value. In the simulation, 10000 independent sample paths are generated by the Euler-Maruyama scheme with discretization step $1/1000$. The datasets generated during the current study are available from the corresponding author upon request.

Tables 5 and 6 show the simulation results with the two types of true parameters settings. In both tables, the sample means are close to the true value and the sample standard deviations are also close to the theoretical standard deviations. Overall, the simulations for both Type I and Type II methods have good behavior.

Next, we conduct the following parametric test:

\[
\begin{align*}
H_0 &: (\beta, \gamma) = (1.2, 1.0), \\
H_1 &: \text{not } H_0^{(1)}.
\end{align*}
\]  

(10)

In this test, we utilize Type II estimator and construct a test statistic by using the method proposed in Sect. 4.3. The test statistic is expected to converge in distribution to $\chi^2_2$ under $H_0$. Hence this test is rejected when the realization of test statistic is greater than $\chi^2_2(\delta)$. We choose $(1.2, 1.0)$, or $(1.3, 0.9)$ as true parameters $(\beta^*, \gamma^*)$, 0.05 as the significance level $\delta$ and other simulation settings are the same as the estimation case in this section. Table 7 shows the number of counts of $H_0$ or $H_1$ selected by the Type II test and empirical sizes and powers. On the whole, the Type II test has good performance.

5.3 Model 3 (Case of estimating only drift parameter)

Third, we consider the case of estimating only the drift parameter. In particular, we treat the case where the diffusion coefficient is the identity matrix.

\[
dX_t = \begin{pmatrix}
1 - \alpha_1 X_{t,1} - 5 \sin(\alpha_2 X_{t,2}^2) \\
2 - \alpha_3 X_{t,2} - 5 \sin(\alpha_4 X_{t,3}^2) \\
3 - \alpha_5 X_{t,3} - 5 \sin(\alpha_6 X_{t,1}^2)
\end{pmatrix} \ dt + \epsilon dW_t, \quad t \in [0, 1], \quad X_0 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad (11)
\]
Table 5  Mean and standard deviation (S.D.) of the estimators in the case of $\theta_0 = (1.2, 1.0)$

| n  | $T$ | Method | $\beta$ Mean | $\beta$ S.D. | $\beta$ Theoretical S.D. | $\gamma$ Mean | $\gamma$ S.D. | $\gamma$ Theoretical S.D. |
|----|-----|--------|--------------|--------------|--------------------------|--------------|--------------|--------------------------|
| 10 | 1   | Type I | 1.199507     | 0.035437     | 0.034884                 | 0.999758     | 0.032008     | 0.031845                 |
|    |     | Type II| 1.199570     | 0.035440     | 0.034884                 | 0.999802     | 0.032010     | 0.031845                 |
| 30 | 1   | Type I | 1.199555     | 0.033724     | 0.033545                 | 1.000083     | 0.030760     | 0.030622                 |
|    |     | Type II| 1.199578     | 0.033726     | 0.033545                 | 1.000099     | 0.030762     | 0.030622                 |
| 360| 12  | Type I | 1.199709     | 0.004906     | 0.004915                 | 1.000173     | 0.004553     | 0.004486                 |
|    |     | Type II| 1.199712     | 0.004906     | 0.004915                 | 1.000176     | 0.004554     | 0.004486                 |
Table 6 Mean and standard deviation (S.D.) of the estimators in the case of $\theta_0 = (0.9, 1.0)$

| $n$ | $T$ | Method | $\beta$ | $\gamma$ |
|-----|-----|--------|---------|---------|
|     |     |        | Mean    | S.D.    | Theoretical S.D. | Mean    | S.D.    | Theoretical S.D. |
| 10  | 1   | Type I | 0.899630 | 0.032820 | 0.032337 | 0.999791 | 0.034273 | 0.034086 |
|     |     | Type II| 0.899663 | 0.032819 | 0.032337 | 0.999833 | 0.034279 | 0.034086 |
| 30  | 1   | Type I | 0.899652 | 0.031437 | 0.031253 | 1.000083 | 0.033082 | 0.032944 |
|     |     | Type II| 0.899659 | 0.031440 | 0.031253 | 1.000092 | 0.033078 | 0.032944 |
| 360 | 12  | Type I | 0.899130 | 0.011284 | 0.011358 | 1.000537 | 0.012076 | 0.011972 |
|     |     | Type II| 0.899144 | 0.011286 | 0.011358 | 1.000555 | 0.012076 | 0.011972 |
Table 7 Results of the Type II test

| True case | \( n \) | \( T \) | Judgement | Empirical Size or Power |
|-----------|--------|-------|-----------|------------------------|
| \( H_0 (\beta^*, \gamma^*) = (1.2, 1.0) \) | 10 | 1 | 9471 | 529 | 0.0529 |
| | 30 | 1 | 9488 | 512 | 0.0512 |
| | 360 | 12 | 9490 | 510 | 0.0510 |
| \( H_1 (\beta^*, \gamma^*) = (1.3, 0.9) \) | 10 | 1 | 202 | 9708 | 0.9708 |
| | 30 | 1 | 133 | 9867 | 0.9867 |
| | 360 | 12 | 0 | 10000 | 1.0000 |

where \( \alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6) \) are unknown parameters. Assume that the parameter space is \( \Theta_\alpha = [0.01, 30]^6 \) and the true parameter values are \( \alpha_0 = (3, 7, 2, 8, 1, 6) \).

We treat the case of \( (\varepsilon, n) = (0.01, 100), (0.001, 100), (0.001, 1000) \) and choose \( v = 3 \) where \( v = 3 \). We estimate the parameters with the Type I method and Type II method. Moreover, we estimate the initial parameters with the following uniform + optim method:

**Step 1.** Generate 20000 uniform random numbers \( \alpha_{0,m} (m = 1, \ldots, 20000) \) on \( [0.01, 30]^6 \).

**Step 2.** Compute

\[
\begin{align*}
\hat{\alpha}^{(1)}_m &= \arg \min_{\alpha \in \Theta_\alpha} U_{\varepsilon, n, v}^{(1)} (\alpha), \quad \text{(case of Type I)}, \\
\hat{\hat{\alpha}}^{(1)}_m &= \arg \min_{\alpha \in \Theta_\alpha} V_{\varepsilon, n}^{(1)} (\alpha), \quad \text{(case of Type II)},
\end{align*}
\]

by means of optim() in the R language, where the uniform random numbers \( \alpha_{0,m} \) are used as the initial value for optimization.

**Step 3.** Define the initial estimator \( \hat{\alpha}^{(1)}_{\text{init}} \) or \( \hat{\hat{\alpha}}^{(1)}_{\text{init}} \) as

\[
\begin{align*}
\hat{\alpha}^{(1)}_{\text{init}} &= \arg \min_{\alpha \in \Theta_\alpha} \left\{ U_{\varepsilon, n, v}^{(1)} (\hat{\alpha}_1^{(1)}), U_{\varepsilon, n, v}^{(1)} (\hat{\alpha}_2^{(1)}), \ldots, U_{\varepsilon, n, v}^{(1)} (\hat{\alpha}_{20000}^{(1)}) \right\}, \quad \text{(case of Type I)}, \\
\hat{\hat{\alpha}}^{(1)}_{\text{init}} &= \arg \min_{\alpha \in \Theta_\alpha} \left\{ V_{\varepsilon, n}^{(1)} (\hat{\alpha}_1^{(1)}), V_{\varepsilon, n}^{(1)} (\hat{\alpha}_2^{(1)}), \ldots, V_{\varepsilon, n}^{(1)} (\hat{\alpha}_{20000}^{(1)}) \right\}, \quad \text{(case of Type II)}.
\end{align*}
\]

In the simulation, 10000 independent sample paths are generated by the Euler-Maruyama scheme with discretization step 1/1000. The datasets generated during the current study are available from the corresponding author upon request.

Table 8 shows the sample means, standard deviations of the simulated estimator values and the computation times of estimation for one sample path. Regarding accuracy of the estimation, we see that both methods performed well for each setting. From the viewpoint of the computation time, however, the estimators of the Type II method are computed more quickly than those of the Type I method. This is because that the contrast function of the Type II method does not optimize the higher order term but put the estimated value \( \hat{\alpha} \) into \( Q_{i,k}(\hat{\alpha}) \). As the result, we recommend using the Type II method.
| $\varepsilon$ | $n$   | Method      | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\alpha_4$ | $\alpha_5$ | $\alpha_6$ | Time (m) |
|-----------|-----|-------------|-----------|-----------|-----------|-----------|-----------|-----------|----------|
| 0.01      | 100 | Type I      | 3.0032    | 7.0127    | 1.9983    | 7.9964    | 1.0049    | 5.9979    | 58       |
|           |     |             | (0.1036)  | (0.3040)  | (0.0127)  | (0.0047)  | (0.0227)  | (0.0131)  |          |
|           |     | Type II     | 2.9891    | 7.0030    | 1.9987    | 7.9958    | 0.9989    | 6.0032    | 11       |
|           |     |             | (0.0570)  | (0.0127)  | (0.0124)  | (0.0068)  | (0.0088)  | (0.0090)  |          |
| 0.001     | 100 | Type I      | 2.9997    | 6.9990    | 1.9986    | 7.9958    | 1.0083    | 6.0003    | 53       |
|           |     |             | (0.0079)  | (0.0028)  | (0.0045)  | (0.0052)  | (0.0240)  | (0.0121)  |          |
|           |     | Type II     | 2.9977    | 7.0002    | 1.9987    | 7.9947    | 0.9994    | 6.0022    | 9        |
|           |     |             | (0.0051)  | (0.0024)  | (0.0050)  | (0.0081)  | (0.0010)  | (0.0014)  |          |
| 0.001     | 1000| Type I      | 2.99994   | 7.00001   | 1.9993    | 7.99997   | 0.99999   | 6.00005   | 204      |
|           |     |             | (0.00182) | (0.00034) | (0.00108) | (0.00018) | (0.00082) | (0.00072) |          |
|           |     | Type II     | 2.99994   | 7.00001   | 1.9993    | 7.99997   | 0.99999   | 6.00005   | 21       |
|           |     |             | (0.00182) | (0.00034) | (0.00108) | (0.00018) | (0.00082) | (0.00072) |          |
6 Proofs

In this section, we treat the case of $T = 1$ without loss of generality. Let the $\sigma$-field $G_k^n := \sigma(X^n_k : s \leq t_k^n)$, and for any vector $u$ (matrix $A$), $u^i(A^{i,j})$ denotes the $i$-th element ($i, j$)-th element) of the vector (matrix). Let $R : (0, \infty) \times \mathbb{R}^d \to \mathbb{R}$ be a function for which there exists a constant $C > 0$ such that $|R(u, x)| \leq uC(1 + |x|)^C$ for all $(u, x) \in (0, \infty) \times \mathbb{R}^d$, and $R_d$ denotes the $d$-dimensional vector whose element satisfies the definition of the function $R$. In order to prove the theorems and the lemmas proposed in Sect. 2, we introduce the following restrictive condition of $[A2']$:

$[A2']$ For all $(x, \beta) \in \mathbb{R}^d \times \Theta_1 \beta$, the matrix $[\sigma \sigma^\top](x, \beta)$ is positive definite. Moreover, the functions $\sigma$, $[\sigma \sigma^\top]^{-1}$ (respectively $b$) are bounded and smooth with bounded derivatives of any order on $\mathbb{R}^d \times \Theta_1 \beta$ (respectively $\mathbb{R}^d \times \Theta_\alpha$).

The following proposition enables us to prove the theorems and the lemmas under $[A1]$, $[A2']$, $[A3]$, $[A4]$ and $[B]$. Note that $[A2']$ implies $[A5]$ in Gloter and Sørensen (2009) by Theorem 1.2 in Freidlin and Wentzell (1998).

Proposition 4 In order to show that the conclusions of Theorems 1-2 and Lemmas 1-2 hold under $[A1]$-$[A4]$ and $[B]$, it is enough to prove that they hold under $[A1]$, $[A2']$, $[A3]$, $[A4]$ and $[B]$.

Proof This result is obtained in an analogous manner to the proof of Proposition 1 in Gloter and Sørensen (2009). We omit the detailed proof. \hfill \Box

Proof of Lemma 1 First, we show $\tilde{\alpha}_{e,n}^{(1)} \xrightarrow{P} \alpha_0$. One deduces that

$$U_{e,n,v}(\alpha) = \varepsilon^{-2}n \sum_{k=1}^n (P_{v,k}(\alpha) - P_{v,k}(\alpha_0) + P_{v,k}(\alpha_0))^\top (P_{v,k}(\alpha) - P_{v,k}(\alpha_0))$$

$$= \varepsilon^{-2}n \sum_{k=1}^n (P_{v,k}(\alpha) - P_{v,k}(\alpha_0))^\top (P_{v,k}(\alpha) - P_{v,k}(\alpha_0))$$

$$+ 2\varepsilon^{-2}n \sum_{k=1}^n (P_{v,k}(\alpha) - P_{v,k}(\alpha_0))^\top P_{v,k}(\alpha_0)$$

$$+ \varepsilon^{-2}n \sum_{k=1}^n P_{v,k}(\alpha_0)^\top P_{v,k}(\alpha_0).$$

Hence, it follows from $[A2']$ and Lemma 4 in Gloter and Sørensen (2009) that

$$\varepsilon^2 \left( U_{e,n,v}^{(1)}(\alpha) - U_{e,n,v}^{(1)}(\alpha_0) \right)$$

$$= n \sum_{k=1}^n (P_{v,k}(\alpha) - P_{v,k}(\alpha_0))^\top (P_{v,k}(\alpha) - P_{v,k}(\alpha_0))$$

\hfill \copyright \ Springer
Adaptive inference for small diffusion processes based on… 669

\[ + 2n \sum_{k=1}^{n} \left( P_{v,k}(\alpha) - P_{v,k}(\alpha_0) \right)^\top P_{v,k}(\alpha_0) \]

\[ = \frac{1}{n} \sum_{k=1}^{n} \left( b(X^n_{t, k-1}, \alpha_0) - b(X^n_{t, k-1}, \alpha) \right)^\top \left( b(X^n_{t, k-1}, \alpha_0) - b(X^n_{t, k-1}, \alpha) \right) \]

\[ + 2n \sum_{k=1}^{n} \left( b(X^n_{t, k-1}, \alpha_0) - b(X^n_{t, k-1}, \alpha) + R_d(n^{-1}, X^n_{t, k-1}) \right)^\top P_{v,k}(\alpha_0) \]

\[ + \frac{1}{n} \sum_{k=1}^{n} R(n^{-1}, X^n_{t, k-1}) \]

\[ \xrightarrow{p} U_1(\alpha, \alpha_0) \text{ uniformly in } \alpha, \quad (12) \]

where \( U_1 \) is defined by (5):

\[ U_1(\alpha, \alpha_0) = \int_0^1 \left( b(X^0_s, \alpha_0) - b(X^0_s, \alpha) \right)^\top \left( b(X^0_s, \alpha_0) - b(X^0_s, \alpha) \right) ds. \]

Let \( \omega \in \Omega \) be fixed. It follows from the compactness of \( \Theta_\alpha \) that for any sequence \((\varepsilon_m, n_m)\), there exists subsequence \((\varepsilon'_m, n'_m)\) such that

\[ \tilde{\alpha}_{\varepsilon'_m, n'_m}(\omega) \to \alpha_\infty \in \Theta_\alpha \quad (\varepsilon_m \to 0, n_m \to \infty). \quad (13) \]

From the continuity of \( U_1 \) and the definition of \( \tilde{\alpha}_{\varepsilon, n}^{(1)} \), one deduces that

\[ 0 \geq \varepsilon^2 \left( U_{\varepsilon, n, v}^{(1)}(\tilde{\alpha}_{\varepsilon, n}^{(1)}(\omega)) - U_{\varepsilon, n, v}^{(1)}(\alpha_0) \right)(\omega) \to U_1(\alpha_\infty, \alpha_0) \geq 0. \]

Hence, we have \( \alpha_\infty = \alpha_0 \) from the identifiability condition [A3], and (13) means that \( \tilde{\alpha}_{\varepsilon, n}^{(1)} \xrightarrow{p} \alpha_0 \).

Second, we prove \( \varepsilon^{-1}(\tilde{\alpha}_{\varepsilon, n}^{(1)} - \alpha_0) = O_P(1) \). It follows from Taylor’s theorem that

\[ -\varepsilon \partial_\alpha U_{\varepsilon, n, v}^{(1)}(\alpha_0) = \left( \varepsilon^2 \int_0^1 \partial_\alpha^2 U_{\varepsilon, n, v}^{(1)}(\alpha_0 + u(\tilde{\alpha}_{\varepsilon, n}^{(1)} - \alpha_0)) du \right) \varepsilon^{-1}(\tilde{\alpha}_{\varepsilon, n}^{(1)} - \alpha_0). \quad (14) \]

For \( 1 \leq \ell \leq p \) and \( 1 \leq l_1, l_2 \leq p \), we deduce from [A2*] and Lemma 4 in Gloter and Sørensen (2009) that

\[ -\varepsilon \partial_\alpha U_{\varepsilon, n, v}^{(1)}(\alpha_0) = 2\varepsilon^{-1} \sum_{k=1}^{n} \sum_{i=1}^{d} \left( \partial_\alpha b^i(X^n_{t, k-1}, \alpha_0) + R(n^{-1}, X^n_{t, k-1}) \right) P_{v,k}^i(\alpha_0) \]

\[ = O_P(1), \quad (15) \]
and
\[\varepsilon^2 \partial^2_{\alpha_t l_2} U^{(1)}_{\varepsilon,n,v}(\alpha) = -2 \sum_{k=1}^{n} \sum_{i=1}^{d} \left( \partial^2_{\alpha_t l_2} b^i(X^n_{i,k-1}, \alpha) - R(n^{-1}, X^n_{k-1}) \right) \left( P_i^{\varepsilon,k}(\alpha) - P_i^{\varepsilon,k}(\alpha_0) \right) \]
\[+ 2n^{-1} \sum_{k=1}^{n} \sum_{i=1}^{d} \left( \partial_{\alpha_t} b^i(X^n_{i,k-1}, \alpha) - R(n^{-1}, X^n_{k-1}) \right) \times \left( \partial_{\alpha_t} b^i(X^n_{i,k-1}, \alpha) - R(n^{-1}, X^n_{k-1}) \right) \]
\[= \frac{2}{n} \sum_{k=1}^{n} \sum_{i=1}^{d} \partial^2_{\alpha_t l_2} b^i(X^n_{i,k-1}, \alpha) \left( b^i(X^n_{i,k-1}, \alpha) - b^i(X^n_{i,k-1}, \alpha_0) \right) \]
\[+ \frac{2}{n} \sum_{k=1}^{n} \sum_{i=1}^{d} \partial_{\alpha_t} b^i(X^n_{i,k-1}, \alpha) \partial_{\alpha_t} b^i(X^n_{i,k-1}, \alpha) + \frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{d} R(n^{-1}, X^n_{k-1}) \]
\[\Rightarrow 2B^{l_1,l_2}_1(\alpha, \alpha_0) \text{ uniformly in } \alpha, \quad (16)\]

where
\[B^{l_1,l_2}_1(\alpha, \alpha_0) := \int_0^1 \left( \partial^2_{\alpha_t l_2} b(X^0_s, \alpha) \right)^\top \left( b(X^0_s, \alpha) - b(X^0_s, \alpha_0) \right) ds + J^{l_1,l_2}_b(\alpha).\]

By noting that for all \( \lambda \in \mathbb{R}^d \setminus \{0\} \), it follows from [A4] that
\[\eta := 2\lambda^\top B_1(\alpha_0, \alpha_0) \lambda = 2\lambda^\top J^{l_1,l_2}_b(\alpha_0) \lambda > 0, \quad (17)\]

and one deduces that
\[1 = P \left( 2\lambda^\top B_1(\alpha_0, \alpha_0) \lambda > \frac{\eta}{2} \right) \]
\[\leq P \left( \lambda^\top \left[ \int_0^1 \left\{ B_1(\alpha_0, \alpha_0) - B_1(\alpha_0 + u(\tilde{\alpha}^{(1)}_{\varepsilon,n} - \alpha_0), \alpha_0) \right\} du \right] \lambda > \frac{\eta}{12} \right) \quad (18)\]
\[+ P \left( \lambda^\top \left[ \int_0^1 \left\{ 2B_1(\alpha_0 + u(\tilde{\alpha}^{(1)}_{\varepsilon,n} - \alpha_0), \alpha_0) \right\} du \right] \lambda > \frac{\eta}{6} \right) \]
\[+ P \left( \lambda^\top \left( 2 \int_0^1 \partial^2_{\alpha} U^{(1)}_{\varepsilon,n,v}(\alpha_0 + u(\tilde{\alpha}^{(1)}_{\varepsilon,n} - \alpha_0)) du \right) \lambda > \frac{\eta}{6} \right) \quad (19)\]

For a sequence \( \{r_n\}_{n \in \mathbb{N}} \) such that \( r_n \to 0 (n \to \infty) \), we define a set \( N_{n,\alpha} \) and an event \( A_{n,\alpha} \) as
\[N_{n,\alpha} := \{ \alpha \in \Theta_\alpha | |\alpha - \alpha_0| \leq r_n \}, \quad A_{n,\alpha} := \left\{ \tilde{\alpha}^{(1)}_{\varepsilon,n} \in N_{n,\alpha} \right\}.\]
Consequently, we obtain $	ilde{a}_{\varepsilon,n} \overset{P}{\to} a_0$, there exists a sequence \( \{r_n\}_{n \in \mathbb{N}} \) such that \( P(A_{n,\alpha}) \to 1 \) \((\varepsilon \to 0, n \to \infty)\). Therefore, for the right hand side of (18), it follows from uniform continuity of \( B_1 \) that

\[
P\left(\lambda \mathbb{T} \left[ \int_0^1 \left\{ B_1(\alpha_0, \alpha_0) - B_1(\alpha_0 + u(\tilde{a}_{\varepsilon,n}^{(1)} - \alpha_0), \alpha_0) \right\} du \right] \lambda > \frac{\eta}{12} \right)
\leq P\left( \sup_{\alpha \in N_{n,\alpha}} \left| B_1(\alpha_0, \alpha_0) - B_1(\alpha, \alpha_0) \right| > \frac{\eta}{12|\lambda|^2} \right) + P(A_{n,\alpha}^c)
\leq P\left( \sup_{\alpha \in N_{n,\alpha}} \left| B_1(\alpha_0, \alpha_0) - B_1(\alpha, \alpha_0) \right| > \frac{\eta}{12|\lambda|^2} \right) + P(A_{n,\alpha}^c)
\to 0 \quad (\varepsilon \to 0, n \to \infty),
\]

and for (19), we deduce from the uniformly convergence (16) that

\[
P\left(\lambda \mathbb{T} \left[ \int_0^1 \left\{ 2B_1(\alpha_0 + u(\tilde{a}_{\varepsilon,n}^{(1)} - \alpha_0), \alpha_0) - \varepsilon^2 \partial_u^2 U_{\varepsilon,n,v}(\alpha_0 + u(\tilde{a}_{\varepsilon,n}^{(1)} - \alpha_0)) \right\} du \right] \lambda > \frac{\eta}{6} \right)
\leq P\left( \sup_{\alpha \in \Theta_u} \left| 2B_1(\alpha, \alpha_0) - \varepsilon^2 \partial_u^2 U_{\varepsilon,n,v}(\alpha) \right| > \frac{\eta}{6|\lambda|^2} \right)
\to 0 \quad (\varepsilon \to 0, n \to \infty).
\]

Consequently, we obtain

\[
P\left(\lambda \mathbb{T} \left( \varepsilon^2 \int_0^1 \partial_u^2 U_{\varepsilon,n,v}(\alpha_0 + u(\tilde{a}_{\varepsilon,n}^{(1)} - \alpha_0)) du \right) \lambda > \frac{\eta}{6} \right) \to 1 \quad (\varepsilon \to 0, n \to \infty),
\]

and hence, it follows from (14), (15) and (20) that one has $\varepsilon^{-1}(\tilde{a}_{\varepsilon,n}^{(1)} - \alpha_0) = O_P(1)$.

Third, we prove the asymptotic normality of $\tilde{a}_{\varepsilon,n}^{(1)}$. In an analogous manner to Sørensen and Uchida (2003), it is sufficient to show the following two properties:

\[
\sup_{u \in [0,1]} \left| \varepsilon^2 \partial_u^2 U_{\varepsilon,n,v}(\alpha_0 + u(\tilde{a}_{\varepsilon,n}^{(1)} - \alpha_0)) - 2J_b(\alpha_0) \right| \overset{P}{\to} 0,
\]

(21)

\[
- \varepsilon \partial_u U_{\varepsilon,n,v}(\alpha_0) \overset{d}{\to} N_p(0, 4K_b(\theta_0)).
\]

(22)

For (21), it follows from (16), the consistency of $\tilde{a}_{\varepsilon,n}^{(1)}$ and \([A2^*]\) that for all $\delta > 0$,

\[
P\left( \sup_{u \in [0,1]} \left| \varepsilon^2 \partial_u^2 U_{\varepsilon,n,v}(\alpha_0 + u(\tilde{a}_{\varepsilon,n}^{(1)} - \alpha_0)) - 2J_b(\alpha_0) \right| > \delta \right)
\leq P\left( \sup_{u \in [0,1]} \left| \varepsilon^2 \partial_u^2 U_{\varepsilon,n,v}(\alpha_0 + u(\tilde{a}_{\varepsilon,n}^{(1)} - \alpha_0)) - 2B_1(\alpha_0 + u(\tilde{a}_{\varepsilon,n}^{(1)} - \alpha_0), \alpha_0) \right| > \frac{\delta}{3} \right)
\]

\(\Box\) Springer
\[ + P \left( \sup_{u \in [0,1]} \left| 2B_1(\alpha_0 + u(\tilde{\alpha}_{\epsilon,n} - \alpha_0), \alpha_0) - 2B_1(\alpha_0, \alpha_0) \right| > \frac{\delta}{3} \right) \]
\[ + P \left( \sup_{u \in [0,1]} \left| 2 \int_0^1 \left( \frac{\partial^2}{\partial \alpha_1^2} b(X_s^0, \alpha_0 + u(\tilde{\alpha}_{\epsilon,n} - \alpha_0)) \right)^\top \right| > \frac{\delta}{3} \right) \]
\[ \leq P \left( \sup_{u \in [0,1]} \left| \epsilon^2 \frac{\partial^2}{\partial \alpha^2} \left( \tilde{\alpha}_{\epsilon,n} - \alpha_0 \right) - 2B_1(\alpha_0, \alpha_0) \right| > \frac{\delta}{3} \right) \]
\[ + P \left( \sup_{\alpha \in \tilde{\Theta}_n} \left| 2B_1(\alpha, \alpha_0) - 2B_1(\alpha_0, \alpha_0) \right| > \frac{\delta}{3} \right) \]
\[ + P \left( \sup_{\alpha \in \tilde{\Theta}_n} \left| b(X_s^0, \alpha) - b(X_s^0, \alpha_0) \right| > \frac{\delta}{6C} \right) + 3P(A_{n,\alpha}^c) \]
\[ \to 0 \quad (\epsilon \to 0, n \to \infty). \]

Regarding (22), we have for \(1 \leq l \leq p\),
\[-\epsilon \partial_{\alpha_l} U_{\epsilon,n,v}^{(1)}(\alpha_0) = \sum_{k=1}^n \xi_{k,1}^l(\alpha_0) + \sum_{k=1}^n R_d(\epsilon^{-1}n^{-1}, X_{tn}^{k-1})^\top P_{v,k}(\alpha_0), \tag{23}\]
where
\[ \xi_{k,1}^l(\alpha_0) := 2\epsilon^{-1} \left( \partial_{\alpha_l} b(X_{tn}^{k-1}, \alpha_0) \right)^\top P_{v,k}(\alpha_0). \tag{24}\]

It follows from Lemma 1 in Gloter and Sørensen (2009) that
\[ \mathbb{E} \left[ \left| R_d(\epsilon^{-1}n^{-1}, X_{tn}^{k-1})^\top P_{v,k}(\alpha_0) \right| \bigg| \mathcal{G}_k\right] \]
\[ = \mathbb{E} \left[ \left| \sum_{i=1}^d R(\epsilon^{-1}n^{-1}, X_{tn}^{k-1}) P_{v,k}(\alpha_0) \right| \bigg| \mathcal{G}_k\right] \]
\[ \leq \sum_{i=1}^d \mathbb{E} \left[ \left| P_{v,k}(\alpha_0) \big| \mathcal{G}_k\right| \right] R(\epsilon^{-1}n^{-1}, X_{tn}^{k-1}) \]
\[ \leq \sum_{i=1}^d \mathbb{E} \left[ \left| P_{v,k}(\alpha_0) \big| \mathcal{G}_k\right|^2 \right]^{\frac{1}{2}} R(\epsilon^{-1}n^{-1}, X_{tn}^{k-1}) \]
\[ \leq R(n^{-\frac{3}{2}}, X_{tn}^{k-1}), \]
and hence the second term of the right hand side in (23) converges to 0 in probability as \(\epsilon \to 0\) and \(n \to \infty\). From Theorems 3.2 and 3.4 in Hall and Heyde (1980), it is sufficient to show the following convergences: for \(1 \leq l_1, l_2 \leq p\),
\[ \mathbb{E} \left[ \left| R_d(\epsilon^{-1}n^{-1}, X_{tn}^{k-1})^\top P_{v,k}(\alpha_0) \right| \bigg| \mathcal{G}_k\right] \]
\[ \to 0 \quad (\epsilon \to 0, n \to \infty). \]
\[
\sum_{k=1}^{n} \mathbb{E} \left[ \xi_{k,1}^{i_1} (\alpha_0) | \mathcal{G}_k^n \right] \to 0
\]
\[
\sum_{k=1}^{n} \mathbb{E} \left[ \xi_{k,1}^{i_1} (\alpha_0) \xi_{k,1}^{i_2} (\alpha_0) | \mathcal{G}_k^n \right] \to 4 \kappa_b^{i_1,i_2} (\theta_0)
\]
\[
\sum_{k=1}^{n} \mathbb{E} \left[ \xi_{k,1}^{i_1} (\alpha_0) | \mathcal{G}_k^n \right] \mathbb{E} \left[ \xi_{k,1}^{i_2} (\alpha_0) | \mathcal{G}_k^n \right] \to 0
\]
\[
\sum_{k=1}^{n} \mathbb{E} \left[ \left( \xi_{k,1}^{i_1} (\alpha_0) \right)^4 | \mathcal{G}_k^n \right] \to 0.
\]

The above convergences are obtained by Lemma 1 and 4 in Gloter and Sørensen (2009). We omit the detailed proof.

**Proof of Theorem 1** 1st step. We prove the consistency of \( \tilde{\beta}_{v,n} \). By the definition of the contrast function \( U_{v,n,v}^{(2)} (\beta | \tilde{\alpha}) \), we have

\[
\frac{1}{n} \left( U_{v,n,v}^{(2)} (\beta | \tilde{\alpha}_{v,n}) - U_{v,n,v}^{(2)} (\beta_0 | \tilde{\alpha}_{v,n}) \right)
\]
\[
= \frac{1}{n} \sum_{k=1}^{n} \left\{ \log \det \left( [\sigma \sigma^\top] (X_{t_k}^n, \beta) [\sigma \sigma^\top]^{-1} (X_{t_k}^n, \beta_0) \right) \right\}
\]
\[
+ \epsilon^{-2} \sum_{k=1}^{n} P_{v,k}^\top (\alpha_0) \left( [\sigma \sigma^\top]^{-1} (X_{t_k}^n, \beta) - [\sigma \sigma^\top]^{-1} (X_{t_k}^n, \beta_0) \right) P_{v,k} (\alpha_0)
\]
\[
+ 2 \epsilon^{-2} n^{-1} \sum_{k=1}^{n} \left( n P_{v,k} (\tilde{\alpha}_{v,n}) - n P_{v,k} (\alpha_0) \right)^\top
\]
\[
\left( [\sigma \sigma^\top]^{-1} (X_{t_k}^n, \beta) - [\sigma \sigma^\top]^{-1} (X_{t_k}^n, \beta_0) \right) P_{v,k} (\alpha_0)
\]
\[
+ (\epsilon n)^{-2} \sum_{k=1}^{n} \left( n P_{v,k} (\tilde{\alpha}_{v,n}) - n P_{v,k} (\alpha_0) \right)^\top
\]
\[
\times \left( [\sigma \sigma^\top]^{-1} (X_{t_k}^n, \beta) - [\sigma \sigma^\top]^{-1} (X_{t_k}^n, \beta_0) \right)
\]
\[
\left( n P_{v,k} (\tilde{\alpha}_{v,n}) - n P_{v,k} (\alpha_0) \right).
\]

It follows from Lemmas 4 and 5 in Gloter and Sørensen (2009) that the sum of (25) and (26) converges to \( U_2 (\beta, \beta_0) \) in probability uniformly in \( \beta \), where \( U_2 \) is defined by (6):

\[
U_2 (\beta, \beta_0) = \int_0^1 \left\{ \log \det \left( [\sigma \sigma^\top] (X_s^0, \beta) [\sigma \sigma^\top]^{-1} (X_s^0, \beta_0) \right)
\]
\[
+ \text{tr} \left( [\sigma \sigma^\top]^{-1} (X_s^0, \beta) [\sigma \sigma^\top] (X_s^0, \beta_0) \right) - d \right\} ds.
\]
Noting that the condition [A2'] leads to the Lipschitz continuity of the functions $n P_{v,k}$ and $[\sigma \sigma^\top]^{-1}$, we deduce from Lemma 1-(5) in Gloter and Sørensen (2009) and Lemma 1 in this paper that (27) and (28) converge to 0 in probability uniformly in $\beta$. Therefore, one deduces that

\[
\frac{1}{n} \left( U^{(2)}_{\varepsilon,n,v}(\beta|\tilde{\alpha}_{\varepsilon,n}^{(1)}) - U^{(2)}_{\varepsilon,n,v}(\beta_0|\tilde{\alpha}_{\varepsilon,n}^{(1)}) \right) \xrightarrow{P} U_2(\beta, \beta_0) \quad \text{uniformly in } \beta. \tag{29}
\]

Let $\omega \in \Omega$ be fixed. It follows from the compactness of $\Theta$ and the consistency of $\tilde{\alpha}_{\varepsilon,n}^{(1)}$ that for any sequence $(\varepsilon_m, n_m)$, there exists a subsequence $(\varepsilon_m', n_m')$ such that

\[
\left( \tilde{\alpha}_{\varepsilon_m', n_m'}^{(1)}(\omega), \tilde{\beta}_{\varepsilon_m', n_m'}^{(1)}(\omega) \right) \to (\alpha_0, \beta_\infty) \in \Theta \quad (\varepsilon_m' \to 0, n_m' \to \infty). \tag{30}
\]

From (29), the continuity of $U_2$ and the definition of $\tilde{\beta}_{\varepsilon,n}$, we obtain that

\[
0 \geq \frac{1}{n} \left( U^{(2)}_{\varepsilon,n,v}(\tilde{\beta}_{\varepsilon_m', n_m'}^{(1)}(\omega) | \tilde{\alpha}_{\varepsilon_m', n_m'}^{(1)}(\omega)) - U^{(2)}_{\varepsilon,n,v}(\beta_0 | \tilde{\alpha}_{\varepsilon_m', n_m'}^{(1)}(\omega)) \right)(\omega) \to U_2(\beta_\infty, \beta_0) \geq 0.
\]

By [A3] and the proof of Lemma 17 in Genon-Catalot and Jacod (1993), we have $\beta_\infty = \beta_0$, and (30) means that $\tilde{\beta}_{\varepsilon,n} \xrightarrow{P} \beta_0$.

\textbf{2nd step.} Next we show the consistency of $\tilde{\alpha}_{\varepsilon,n}$. By the definition of the contrast function $U_{\varepsilon,n,v}(\alpha|^\beta)$, we have

\[
e^2 \left( U_{\varepsilon,n,v}(\alpha | \tilde{\beta}_{\varepsilon,n}) - U_{\varepsilon,n,v}(\alpha_0 | \tilde{\beta}_{\varepsilon,n}) \right)
= 2n \sum_{k=1}^{n} \left( P_{v,k}(\alpha) - P_{v,k}(\alpha_0) \right)^\top \left[ \sigma \sigma^\top \right]^{-1} (X_{t_{k-1}}^{\varepsilon}, \tilde{\beta}_{\varepsilon,n}) P_{v,k}(\alpha_0)
+ n \sum_{k=1}^{n} \left( P_{v,k}(\alpha) - P_{v,k}(\alpha_0) \right)^\top \left[ \sigma \sigma^\top \right]^{-1} (X_{t_{k-1}}^{\varepsilon}, \beta_0) \left( P_{v,k}(\alpha) - P_{v,k}(\alpha_0) \right)
+ n \sum_{k=1}^{n} \left( P_{v,k}(\alpha) - P_{v,k}(\alpha_0) \right)^\top \left[ \sigma \sigma^\top \right]^{-1} (X_{t_{k-1}}^{\varepsilon}, \tilde{\beta}_{\varepsilon,n}) - \left[ \sigma \sigma^\top \right]^{-1} (X_{t_{k-1}}^{\varepsilon}, \beta_0)
\]
\[
= 2n \sum_{k=1}^{n} \left( b(X_{t_{k-1}}^{\varepsilon}, \alpha_0) - b(X_{t_{k-1}}^{\varepsilon}, \alpha) + R_d(n^{-1}, X_{t_{k-1}}^{\varepsilon}) \right)^\top \left[ \sigma \sigma^\top \right]^{-1} (X_{t_{k-1}}^{\varepsilon}, \tilde{\beta}_{\varepsilon,n}) P_{v,k}(\alpha_0)
\]
\[
+ \frac{1}{n} \sum_{k=1}^{n} \left( b(X_{t_{k-1}}^{\varepsilon}, \alpha) - b(X_{t_{k-1}}^{\varepsilon}, \alpha_0) \right)^\top \left[ \sigma \sigma^\top \right]^{-1} (X_{t_{k-1}}^{\varepsilon}, \beta_0)
\]
\[
\left( b(X_{t_{k-1}}^{\varepsilon}, \alpha) - b(X_{t_{k-1}}^{\varepsilon}, \alpha_0) \right)
\]
\[
+ \frac{1}{n} \sum_{k=1}^{n} \left( b(X_{t_{k-1}}^{\varepsilon}, \alpha) - b(X_{t_{k-1}}^{\varepsilon}, \alpha_0) \right)^\top \left[ \sigma \sigma^\top \right]^{-1} (X_{t_{k-1}}^{\varepsilon}, \tilde{\beta}_{\varepsilon,n}) - \left[ \sigma \sigma^\top \right]^{-1} (X_{t_{k-1}}^{\varepsilon}, \beta_0)
\]
Adaptive inference for small diffusion processes based on…

\[
\left(b(X_{t_{k-1}}, \alpha) - b(X_{t_{k-1}}, \alpha_0)\right) + \frac{1}{n} \sum_{k=1}^{n} R(n^{-1}, X_{t_{k-1}}),
\]

It follows from Lemma 4-(2) in Gloter and Sørensen (2009) for (31), and the consistency of \( \tilde{\beta}_{\varepsilon, n} \) for (33) that the two terms converge to 0 in probability uniformly in \( \alpha \). Moreover, from Lemma 4-(1) in Gloter and Sørensen (2009), (32) converges to \( U_3(\alpha, \theta_0) \) in probability uniformly in \( \alpha \), where

\[
U_3(\alpha, \theta_0) = \int_0^1 \left(b(X_s^0, \alpha) - b(X_s^0, \alpha_0)\right)^{\top} \left[\sigma \sigma^\top\right]^{-1} (X_s^0, \beta_0) \left(b(X_s^0, \alpha) - b(X_s^0, \alpha_0)\right) ds.
\]

Therefore, we have

\[
\varepsilon^2 \left(U_{e,n,v}^{(3)}(\alpha|\tilde{\beta}_{\varepsilon, n}) - U_{e,n,v}^{(3)}(\alpha_0|\tilde{\beta}_{\varepsilon, n})\right) \xrightarrow{n \to \infty} U_3(\alpha, \theta_0) \quad \text{uniformly in } \alpha.
\]

In an analogous manner to the proof of the consistency for \( \tilde{\alpha}_{\varepsilon, n} \) and \( \tilde{\beta}_{\varepsilon, n} \), we have \( \tilde{\alpha}_{\varepsilon, n} \xrightarrow{n \to \infty} \alpha_0 \).

**3rd step.** We prove the asymptotic normality for \( \tilde{\theta}_{\varepsilon, n} \). From Taylor’s theorem, we have the following expansions:

\[-\varepsilon \partial_\alpha U_{e,n,v}^{(3)}(\alpha_0|\tilde{\beta}_{\varepsilon, n}) = \left(\varepsilon^2 \int_0^1 \partial_\alpha^2 U_{e,n,v}^{(3)}(\alpha_0 + u(\tilde{\alpha}_{\varepsilon, n} - \alpha_0)|\tilde{\beta}_{\varepsilon, n}) du\right) \varepsilon^{-1}(\tilde{\alpha}_{\varepsilon, n} - \alpha_0),\]

\[-\frac{1}{\sqrt{n}} \partial_\beta U_{e,n,v}^{(2)}(\beta_0|\tilde{\alpha}_{\varepsilon, n}) = \left(\frac{1}{n} \int_0^1 \partial_\beta^2 U_{e,n,v}^{(2)}(\beta_0 + u(\tilde{\beta}_{\varepsilon, n} - \beta_0)|\tilde{\alpha}_{\varepsilon, n}) du\right) \sqrt{n}(\tilde{\beta}_{\varepsilon, n} - \beta_0),\]

\[-\varepsilon \partial_\alpha U_{e,n,v}^{(3)}(\alpha_0|\tilde{\beta}_{\varepsilon, n}) - \varepsilon \partial_\beta U_{e,n,v}^{(3)}(\alpha_0|\beta_0) = \left(\frac{\varepsilon}{\sqrt{n}} \int_0^1 \partial_{\alpha\beta} U_{e,n,v}^{(3)}(\alpha_0 + u(\tilde{\alpha}_{\varepsilon, n} - \alpha_0)|\beta_0 + u(\tilde{\beta}_{\varepsilon, n} - \beta_0)) du\right) \sqrt{n}(\tilde{\beta}_{\varepsilon, n} - \beta_0).\]

Using these expressions, we calculate that

\[
\Gamma_1^{1, e,n} = C_1^{1, e,n} \Lambda_1^{1, e,n},
\]

where

\[
\Gamma_1^{1, e,n} := \left(-\varepsilon \partial_\alpha U_{e,n,v}^{(3)}(\alpha_0|\beta_0) - \frac{1}{\sqrt{n}} \partial_\beta U_{e,n,v}^{(2)}(\beta_0|\tilde{\alpha}_{\varepsilon, n})\right), \quad \Lambda_1^{1, e,n} := \left(\varepsilon^{-1}(\tilde{\alpha}_{\varepsilon, n} - \alpha_0)\right) / \sqrt{n}(\tilde{\beta}_{\varepsilon, n} - \beta_0),
\]

\[
C_1^{1, e,n} := \left(\varepsilon^2 \int_0^1 \partial_\alpha^2 U_{e,n,v}^{(3)}(\alpha_0 + u(\tilde{\alpha}_{\varepsilon, n} - \alpha_0)|\tilde{\beta}_{\varepsilon, n}) du - \frac{1}{\sqrt{n}} \int_0^1 \partial_{\alpha\beta} U_{e,n,v}^{(3)}(\alpha_0|\beta_0 + u(\tilde{\beta}_{\varepsilon, n} - \beta_0)) du\right),
\]

\[
\frac{1}{n} \int_0^1 \partial_\beta^2 U_{e,n,v}^{(2)}(\beta_0 + u(\tilde{\beta}_{\varepsilon, n} - \beta_0)|\tilde{\alpha}_{\varepsilon, n}) du.
\]

 Springer
In an analogous manner to the proof of Theorem 1 in Sørensen and Uchida (2003), it is sufficient to show the following convergences:

\[ \sup_{u \in [0, 1]} \left| \varepsilon^2 \partial^2 \alpha_{n,v}^{(3)}(\alpha_0 + u(\tilde{\alpha}_{e,n} - \alpha_0)) - 2I_p(\theta_0) \right| \xrightarrow{P} 0, \]  

(36)

\[ \sup_{u \in [0, 1]} \left| \frac{1}{n} \partial^2 \beta_{U_{n,v}}^{(2)}(\beta_0 + u(\tilde{\beta}_{e,n} - \beta_0)) - 2I_q(\theta_0) \right| \xrightarrow{P} 0, \]  

(37)

\[ \sup_{u \in [0, 1]} \left| \frac{\varepsilon}{\sqrt{n}} \partial_{\alpha_0 \beta} U_{n,v}^{(3)}(\alpha_0 | \beta_0 + u(\tilde{\beta}_{e,n} - \beta_0)) \right| \xrightarrow{P} 0, \]  

(38)

\[ \Gamma_{e,n}^1 \xrightarrow{d} N_{p+q}(0, 4I(\theta_0)). \]  

(39)

Proof of (36). By a simple calculation, it holds that for \( 1 \leq l_1, l_2 \leq p, \)

\[ \varepsilon^2 \partial^2_{\alpha_{l_1/2}^{(3)}} U_{n,v}^{(3)}(\alpha | \tilde{\beta}_{e,n}) \]

\[ = 2n \sum_{k=1}^{n} \partial^2_{\alpha_{l_1/2}^{(3)}} P_{v,k}(\alpha)[\sigma \sigma^\top]^{-1}(X_{k-1}^{n}, \tilde{\beta}_{e,n}) P_{v,k}(\alpha_0) \]

\[ + 2n \sum_{k=1}^{n} \partial^2_{\alpha_{l_1/2}^{(3)}} P_{v,k}(\alpha) [\sigma \sigma^\top]^{-1}(X_{k-1}^{n}, \tilde{\beta}_{e,n}) (P_{v,k}(\alpha) - P_{v,k}(\alpha_0)) \]

\[ + 2n \sum_{k=1}^{n} \partial^2_{\alpha_{l_1/2}^{(3)}} P_{v,k}(\alpha) \left( [\sigma \sigma^\top]^{-1}(X_{k-1}^{n}, \tilde{\beta}_{e,n}) - [\sigma \sigma^\top]^{-1}(X_{k-1}^{n}, \beta_0) \right) \]

\[ (P_{v,k}(\alpha) - P_{v,k}(\alpha_0)) \]

\[ + 2n \sum_{k=1}^{n} \partial_{\alpha_{l_1}} P_{v,k}(\alpha)[\sigma \sigma^\top]^{-1}(X_{k-1}^{n}, \beta_0) \partial_{\alpha_{l_2}} P_{v,k}(\alpha) \]

\[ + 2n \sum_{k=1}^{n} \partial_{\alpha_{l_1}} P_{v,k}(\alpha) \left( [\sigma \sigma^\top]^{-1}(X_{k-1}^{n}, \tilde{\beta}_{e,n}) - [\sigma \sigma^\top]^{-1}(X_{k-1}^{n}, \beta_0) \right) \partial_{\alpha_{l_2}} P_{v,k}(\alpha). \]

(40)

(41)

(42)

(43)

(44)

By noting that \( \partial_{\alpha_{l_1}} P_{v,k}(\alpha) = -n^{-1} \partial_{\alpha_{l_1}} b(X_{k-1}^{n}, \alpha) + R_d(n^{-2}, X_{k-1}^{n}) = R_d(n^{-1}, X_{k-1}^{n}), \) and \( P_{v,k}(\alpha) - P_{v,k}(\alpha_0) = n^{-1} \left( b(X_{k-1}^{n}, \alpha) - b(X_{k-1}^{n}, \alpha_0) \right) + R_d(n^{-2}, X_{k-1}^{n}) = R_d(n^{-1}, X_{k-1}^{n}), \) it follows from Lemma 4-(1) in Gloter and Sørensen (2009) that the sum of (41) and (43) converges to \( 2B_{2}^{l_1,l_2}(\alpha, \theta_0) \) in probability uniformly in \( \alpha, \) where

\[ B_{2}^{l_1,l_2}(\alpha, \theta_0) := \int_0^1 \left( \partial^2_{\alpha_{l_1}} b(X_s^0, \alpha) \right)^\top [\sigma \sigma^\top]^{-1}(X_s^0, \beta_0) \left( b(X_s^0, \alpha) - b(X_s^0, \alpha_0) \right) ds \]

\[ + \int_0^1 \left( \partial_{\alpha_{l_1}} b(X_s^0, \alpha) \right)^\top [\sigma \sigma^\top]^{-1}(X_s^0, \beta_0) \left( \partial_{\alpha_{l_2}} b(X_s^0, \alpha) \right) ds. \]
On the other hand, by the consistency of $\tilde{\beta}_e, n$ and Lemma 4 in Gloter and Sørensen (2009), the remaining terms (40), (42) and (44) converge to 0 in probability. Hence, we have

$$\sup_{\alpha \in \Theta_1} \left| \epsilon^2 \partial^2_{\alpha_1^2} U_{e,n,v}^{(3)}(\alpha) - 2 B^{l_1,l_2}_2(\alpha, \theta_0) \right| \xrightarrow{P} 0. \quad (45)$$

By noting that $B_2(\alpha_0, \theta_0) = I_b(\theta_0)$, it follows from the consistency of $\tilde{\alpha}_e, n$ and the uniform continuity of $B_2$ that for all $\delta > 0$,

$$P \left( \sup_{u \in [0,1]} \epsilon^2 \partial^2_{\alpha_1^2} U_{e,n,v}^{(3)}(\alpha_0 + u(\tilde{\alpha}_e, n - \alpha_0)|\tilde{\beta}_e, n) - 2 I_b(\theta_0) \right) > \frac{\delta}{2} \leq P \left( \sup_{\alpha \in \Theta_1} \epsilon^2 \partial^2_{\alpha_1^2} U_{e,n,v}^{(3)}(\alpha) - 2 B_2(\alpha, \theta_0) \right) > \frac{\delta}{2}.$$

This implies (36).

Proof of (37). We deduce that for $1 \leq m_1, m_2 \leq q$,

$$\frac{1}{n} \partial^2_{\tilde{\beta}_{m_1,m_2}^e} U_{e,n,v}^{(2)}(\beta|\tilde{\alpha}_{e,n}^{(1)}) = \frac{1}{n} \sum_{k=1}^n \partial^2_{\tilde{\beta}_{m_1,m_2}^e} \log \det \{ \sigma \sigma^\top \}(X^n_{t_{k-1}}; \beta) \quad (46)$$

$$+ \epsilon^{-2} \sum_{k=1}^n P_{v,k}(\alpha_0)^\top \left( \partial^2_{\tilde{\beta}_{m_1,m_2}^e} [\sigma \sigma^\top]^{-1}(X^n_{t_{k-1}}; \beta) \right) P_{v,k}(\alpha_0) \quad (47)$$

$$+ 2 \epsilon^{-2} \sum_{k=1}^n \left( P_{v,k}(\tilde{\alpha}_{e,n}^{(1)}) - P_{v,k}(\alpha_0) \right)^\top \left( \partial^2_{\tilde{\beta}_{m_1,m_2}^e} [\sigma \sigma^\top]^{-1}(X^n_{t_{k-1}}; \beta) \right) P_{v,k}(\alpha_0) \quad (48)$$

$$+ 2 \epsilon^{-2} \sum_{k=1}^n \left( P_{v,k}(\tilde{\alpha}_{e,n}^{(1)}) - P_{v,k}(\alpha_0) \right)^\top \left( \partial^2_{\tilde{\beta}_{m_1,m_2}^e} [\sigma \sigma^\top]^{-1}(X^n_{t_{k-1}}; \beta) \right) \left( P_{v,k}(\tilde{\alpha}_{e,n}^{(1)}) - P_{v,k}(\alpha_0) \right). \quad (49)$$

It follows from Lemma 4 and 5 in Gloter and Sørensen (2009) that the sum of (46) and (47) converges to $B_3^{r_1,r_2}(\beta, \beta_0)$ in probability uniformly in $\beta$, where

$$B_3^{m_1,m_2}(\beta, \beta_0) \quad \square$$
\[ : = - \int_0^1 \text{tr} \left( \left( \sigma \sigma^\top \right)^{-1} \left( \partial_{\beta_{m_1}} \left( \sigma \sigma^\top \right) \right) \left( \sigma \sigma^\top \right)^{-1} \left( \partial_{\beta_{m_2}} \left( \sigma \sigma^\top \right) \right) \right) (X^0_s, \beta) ds \]
\[ + \int_0^1 \text{tr} \left( \left( \sigma \sigma^\top \right)^{-1} \left( \partial_{\beta_{m_1}}^2 \left( \sigma \sigma^\top \right) \right) \right) (X^0_s, \beta) ds \]
\[ + 2 \int_0^1 \text{tr} \left[ \left( \sigma \sigma^\top \right)^{-1} \left( \partial_{\beta_{m_1}} \left( \sigma \sigma^\top \right) \right) \left( \sigma \sigma^\top \right)^{-1} \left( \partial_{\beta_{m_2}} \left( \sigma \sigma^\top \right) \right) \right] (X^0_s, \beta) ds \]
\[ - \int_0^1 \text{tr} \left[ \left( \sigma \sigma^\top \right)^{-1} \left( \partial_{\beta_{m_1}}^2 \left( \sigma \sigma^\top \right) \right) \right] (X^0_s, \beta) ds. \]

We can show from Lemma 1 in this paper and Lemma 1 in Gloter and Sørensen (2009) that (48) and (49) converge to 0 in probability uniformly in \( \beta \). Therefore, it holds that

\[ \sup_{\beta \in \Theta_1} \left| \frac{1}{n} \partial_{\beta_{m_{1,2}}}^2 U^{(2)}_{\epsilon, n, v}(\beta|\alpha^{(1)}_{\epsilon, n}) - B_{3}^{m_{1,2}}(\beta, \beta_0) \right| \xrightarrow{P} 0. \quad (50) \]

Noting that \( B_{3}(\beta_0, \beta_0) = 2I_\sigma(\beta_0) \), in an analogous manner to the proof of (36), we deduce that for all \( \delta > 0 \),

\[ P \left( \sup_{u \in [0, 1]} \left| \frac{1}{n} \partial_{\beta_{m_{1,2}}}^2 U^{(2)}_{\epsilon, n, v}(\beta_0 + u(\tilde{\beta}_{\epsilon, n} - \beta_0)|\alpha^{(1)}_{\epsilon, n}) - 2I_\sigma(\beta_0) \right| > \delta \right) \rightarrow 0 \quad (\epsilon \to 0, n \to \infty). \]

This implies (37).

Proof of (38). From Lemma 4-(2) in Gloter and Sørensen (2009), it holds that for \( 1 \leq l \leq p \) and \( 1 \leq m \leq q \),

\[ \frac{\epsilon}{\sqrt{n}} \partial_{\alpha_l \beta_m}^2 U^{(3)}_{\epsilon, n, v}(\alpha_0|\beta) \]
\[ = 2\epsilon^{-1} \sqrt{n} \sum_{k=1}^n \partial_{\alpha_l} P_{v,k}(\alpha_0)^\top \partial_{\beta_m} \left( \sigma \sigma^\top \right)^{-1} (X^n_{i_{k-1}}, \beta) P_{v,k}(\alpha_0) \]
\[ = \frac{2}{\sqrt{n}} \epsilon^{-1} \sum_{k=1}^n R_d(1, X^n_{i_{k-1}})^\top P_{v,k}(\alpha_0) \]
\[ \xrightarrow{P} 0 \quad \text{uniformly in } \beta. \]

In particular, we have

\[ \sup_{u \in [0, 1]} \left| \frac{\epsilon}{\sqrt{n}} \partial_{\alpha \beta}^2 U^{(3)}_{\epsilon, n, v}(\alpha_0|\beta_0 + u(\tilde{\beta}_{\epsilon, n} - \beta_0)) \right| \xrightarrow{P} 0. \]
Proof of (39). By using Lemma 4-(2) in Gloter and Sørensen (2009), it holds that for $1 \leq l \leq p$,

$$-\varepsilon \partial_{\alpha_l} U_{\varepsilon,n,v}^{(3)}(\alpha_0|\beta_0) = -2\varepsilon^{-1} n \sum_{k=1}^{n} \partial_{\alpha_l} P_{v,k}^{\top}(\alpha_0)[\sigma \sigma^\top]^{-1}(X_{t_{k-1}^n}, \beta_0) P_{v,k}(\alpha_0) = \sum_{k=1}^{n} \xi_{k,1}^{l}(\theta_0) + o_P(1), \quad (51)$$

where

$$\xi_{k,1}^{l}(\theta_0) := 2\varepsilon^{-1} \partial_{\alpha_l} b(X_{t_{k-1}^n}, \alpha_0)^\top [\sigma \sigma^\top]^{-1}(X_{t_{k-1}^n} - 1, \beta_0) P_{v,k}(\alpha_0). \quad (52)$$

On the other hand, it follows from Taylor's theorem that

$$-\sqrt{n} \partial_{\beta} U_{\varepsilon,n,v}^{(2)}(\beta_0|\check{\alpha}_{\varepsilon,n}^{(1)}) = -\sqrt{n} \partial_{\beta} U_{\varepsilon,n,v}^{(2)}(\beta_0|\alpha_0) - \left( \frac{\varepsilon}{\sqrt{n}} \int_0^1 \partial_{\alpha} U_{\varepsilon,n,v}^{(2)}(\beta_0|\alpha_0 + u(\check{\alpha}_{\varepsilon,n}^{(1)} - \alpha_0)) du \right) \varepsilon^{-1}(\check{\alpha}_{\varepsilon,n}^{(1)} - \alpha_0).$$

In particular, it holds from Lemma 1 in this paper and Lemma 4-(2) in Gloter and Sørensen (2009) that

$$\varepsilon \sqrt{n} \partial_{\alpha_\beta} U_{\varepsilon,n,v}^{(2)}(\beta_0|\alpha) = 2\varepsilon^{-1} \sqrt{n} \sum_{k=1}^{n} \partial_{\alpha_l} P_{v,k}(\alpha)^\top \partial_{\beta_m}[\sigma \sigma^\top]^{-1}(X_{t_{k-1}^n}, \beta_0) P_{v,k}(\alpha_0) + 2\varepsilon^{-1} \sqrt{n} \sum_{k=1}^{n} \partial_{\alpha_l} P_{v,k}(\alpha)^\top \partial_{\beta_m}[\sigma \sigma^\top]^{-1}(X_{t_{k-1}^n}, \beta_0) \left( P_{v,k}(\alpha) - P_{v,k}(\alpha_0) \right) \rightarrow 0 \quad \text{uniformly in } \alpha,$$

and we have

$$\sup_{u \in [0,1]} \left| \frac{\varepsilon}{\sqrt{n}} \partial_{\alpha_\beta} U_{\varepsilon,n,v}^{(2)}(\beta_0|\alpha_0 + u(\check{\alpha}_{\varepsilon,n}^{(1)} - \alpha_0)) \right| \rightarrow 0.$$

Therefore, by Lemma 1, we deduce that for $1 \leq m \leq q$,

$$-\sqrt{n} \partial_{\beta_m} U_{\varepsilon,n,v}^{(2)}(\beta_0|\check{\alpha}_{\varepsilon,n}^{(1)}) = -\sqrt{n} \partial_{\beta_m} U_{\varepsilon,n,v}^{(2)}(\beta_0|\alpha_0) + o_P(1) = \sum_{k=1}^{n} \left( \eta_{k,1}^{m}(\beta_0) + \eta_{k,2}^{m}(\theta_0) \right) + o_P(1), \quad (53)$$

where

$$\eta_{k,1}^{m}(\beta_0) := \frac{\varepsilon}{\sqrt{n}} \partial_{\alpha} b(X_{t_{k-1}^n}, \alpha_0)^\top [\sigma \sigma^\top]^{-1}(X_{t_{k-1}^n} - 1, \beta_0) P_{v,k}(\alpha_0).$$
where
\[
\eta_{k,1}(\beta_0) := -\frac{1}{2} n^{-\frac{1}{2}} \text{tr} \left[ \left( \sigma \sigma^T \right)^{-1} \left( \beta_m \left( \sigma \sigma^T \right) \right) (X_{k-1}, \beta_0) \right], \tag{54}
\]
\[
\eta_{k,2}(\theta_0) := \varepsilon^{-2} n^{\frac{1}{2}} P_{v,k}(\alpha_0) \left( \left( \sigma \sigma^T \right)^{-1} \left( \beta_m \left( \sigma \sigma^T \right) \right) \left( \sigma \sigma^T \right)^{-1} \right) (X_{k-1}, \beta_0) P_{v,k}(\alpha_0). \tag{55}
\]

In order to show (39), by Theorems 3.2 and 3.4 in Hall and Heyde (1980), it is sufficient to show the following convergences: for \(1 \leq l_1, l_2 \leq p\) and \(1 \leq m_1, m_2 \leq q\),
\[
\sum_{k=1}^{n} \mathbb{E} \left[ \xi_{k,1}(\theta_0) | G_{k-1}^{n} \right] \overset{P}{\rightarrow} 0,
\]
\[
\sum_{k=1}^{n} \mathbb{E} \left[ \eta_{k,1}(\beta_0) + \eta_{k,2}(\theta_0) | G_{k-1}^{n} \right] \overset{P}{\rightarrow} 0,
\]
\[
\sum_{k=1}^{n} \mathbb{E} \left[ \xi_{k,1}(\theta_0) \xi_{k,1}(\theta_0) | G_{k-1}^{n} \right] - \sum_{k=1}^{n} \mathbb{E} \left[ \eta_{k,1}(\beta_0) + \eta_{k,2}(\theta_0) | G_{k-1}^{n} \right] \mathbb{E} \left[ \eta_{k,1}(\beta_0) + \eta_{k,2}(\theta_0) | G_{k-1}^{n} \right] \overset{P}{\rightarrow} 4 I_{1}^{m_1, m_2}(\beta_0),
\]
\[
\sum_{k=1}^{n} \mathbb{E} \left[ \eta_{k,1}(\beta_0) (\eta_{k,1}(\beta_0) + \eta_{k,2}(\theta_0)) | G_{k-1}^{n} \right] \text{ and } \sum_{k=1}^{n} \mathbb{E} \left[ \xi_{k,1}(\theta_0) (\eta_{k,1}(\beta_0) + \eta_{k,2}(\theta_0)) | G_{k-1}^{n} \right] \overset{P}{\rightarrow} 0,
\]
\[
\sum_{k=1}^{n} \mathbb{E} \left[ \eta_{k,1}(\beta_0) + \eta_{k,2}(\theta_0) | G_{k-1}^{n} \right] \overset{P}{\rightarrow} 0,
\]
\[
\sum_{k=1}^{n} \mathbb{E} \left[ \left( \xi_{k,1}(\theta_0) \right)^{4} | G_{k-1}^{n} \right] \overset{P}{\rightarrow} 0,
\]
\[
\sum_{k=1}^{n} \mathbb{E} \left[ \left( \eta_{k,1}(\beta_0) + \eta_{k,2}(\theta_0) \right)^{4} | G_{k-1}^{n} \right] \overset{P}{\rightarrow} 0.
\]

The above properties are obtained by Lemmas 1 and 4 in Gloter and Sørensen (2009). We omit a detailed proof. Therefore, we have \(\Lambda_{\varepsilon,n} \overset{d}{\rightarrow} N_d(0, I(\theta_0)^{-1})\), which completes the proof. \(\square\)

**Proof of Lemma 2** First, we prove that \(\hat{\alpha}_{\varepsilon,n}^{(l)} \overset{P}{\rightarrow} \alpha_0\) for \(l = 1, \ldots, v\). When \(l = 1\), a simple computation shows that
\[
\varepsilon^{2} \left( V_{\varepsilon,n}^{(1)}(\alpha) - V_{\varepsilon,n}^{(1)}(\alpha_0) \right) = \frac{1}{n} \sum_{k=1}^{n} \left( b(X_{k-1}^{n}, \alpha_0) - b(X_{k-1}^{n}, \alpha) \right) \left( b(X_{k-1}^{n}, \alpha_0) - b(X_{k-1}^{n}, \alpha) \right)^{\top}
\]
\[
\overset{\varepsilon \rightarrow 0}{\rightarrow} V_{\varepsilon,n}^{(1)}(\alpha_0),
\]
where
\[
\eta_{k,1}(\beta_0) := -\frac{1}{2} n^{-\frac{1}{2}} \text{tr} \left[ \left( \sigma \sigma^T \right)^{-1} \left( \beta_m \left( \sigma \sigma^T \right) \right) (X_{k-1}, \beta_0) \right],
\]
\[
\eta_{k,2}(\theta_0) := \varepsilon^{-2} n^{\frac{1}{2}} P_{v,k}(\alpha_0) \left( \left( \sigma \sigma^T \right)^{-1} \left( \beta_m \left( \sigma \sigma^T \right) \right) \left( \sigma \sigma^T \right)^{-1} \right) (X_{k-1}, \beta_0) P_{v,k}(\alpha_0).
\]
Adaptive inference for small diffusion processes based on… 681

Hence, it follows from (57) that \( \hat{\alpha}_{\varepsilon,n} \rightarrow \alpha_0 \).

Second, we show \( \varepsilon^{\frac{1}{v}} (\hat{\alpha}_{\varepsilon,n} - \alpha_0) \rightarrow 0 \). In the case of \( v = 1 \), this statement is the same as that of Lemma 1. Therefore, we assume \( v \geq 2 \). It follows from Taylor’s theorem that

\[
-\varepsilon^{\frac{1}{v}} \partial_\alpha V^{(1)}_{\varepsilon,n}(\alpha_0) = \left( \varepsilon^2 \int_0^1 \partial^2_\alpha V^{(1)}_{\varepsilon,n}(\alpha_0 + u(\hat{\alpha}_{\varepsilon,n} - \alpha_0)) du \right) \varepsilon^{-\frac{1}{v}} (\hat{\alpha}_{\varepsilon,n} - \alpha_0).
\]

In a similar way to the proof of (50), we have

\[
\varepsilon^2 \left( V^{(l)}_{\varepsilon,n}(\alpha|\hat{\alpha}_{\varepsilon,n}^{(l-1)}) - V^{(l)}_{\varepsilon,n}(\alpha|\alpha_0,\hat{\alpha}_{\varepsilon,n}^{(l-1)}) \right) \overset{p}{\rightarrow} U_1(\alpha, \alpha_0) \text{ uniformly in } \alpha.
\]

Therefore, in the same manner as the proof of the consistency of \( \hat{\alpha}_{\varepsilon,n} \) in Lemma 1, we can show \( \hat{\alpha}_{\varepsilon,n} \overset{p}{\rightarrow} \alpha_0 \). When \( l \geq 2 \), assume that \( \hat{\alpha}_{\varepsilon,n}^{(l-1)} \) is consistent. Noting that \( Q_{l,k}(\alpha) = R_d(n^{-2}, X^n_{t^{-1}}) \), we calculate that

\[
\varepsilon^2 \left( V^{(l)}_{\varepsilon,n}(\alpha|\hat{\alpha}_{\varepsilon,n}^{(l-1)}) - V^{(l)}_{\varepsilon,n}(\alpha|\alpha_0,\hat{\alpha}_{\varepsilon,n}^{(l-1)}) \right)
\]

In a similar way to the proof of (50), we have

\[
\varepsilon^2 \left( V^{(l)}_{\varepsilon,n}(\alpha|\hat{\alpha}_{\varepsilon,n}^{(l-1)}) - V^{(l)}_{\varepsilon,n}(\alpha|\alpha_0,\hat{\alpha}_{\varepsilon,n}^{(l-1)}) \right) \overset{p}{\rightarrow} U_1(\alpha, \alpha_0) \text{ uniformly in } \alpha.
\]
In order to prove $\varepsilon^{-\frac{1}{v}}(\hat{\alpha}_{k,n}^{(1)} - \alpha_0) \xrightarrow{P} 0$, by $[A4]$-(i), it is sufficient to show the following properties:

\begin{equation}
-\varepsilon^{2-\frac{1}{v}} \partial_\alpha V_{\varepsilon,n}^{(1)}(\alpha_0) \xrightarrow{P} 0, \tag{58}
\end{equation}

\begin{equation}
\sup_{u \in [0,1]} \left| \varepsilon^{2} \partial^2_\alpha V_{\varepsilon,n}^{(1)}(\alpha_0 + u(\hat{\alpha}_{k,n}^{(1)} - \alpha_0)) - 2J_b(\alpha_0) \right| \xrightarrow{P} 0. \tag{59}
\end{equation}

Proof of (58). By the definition of the contrast function $V_{\varepsilon,n}^{(1)}(\alpha)$, we have

\[-\varepsilon^{2-\frac{1}{v}} \partial_\alpha V_{\varepsilon,n}^{(1)}(\alpha_0) = \sum_{k=1}^{n} \psi_k^{(1)}(\alpha_0),\]

where

\[\psi_k^{(1)}(\alpha_0) = 2\varepsilon^{-\frac{1}{v}} \partial_\alpha b(X_{t_{k-1}^n}, \alpha_0) \top P_{1,k}(\alpha_0).\]

From Lemma 9 in Genon-Catalot and Jacod (1993), it is sufficient to show the following two convergences.

\begin{equation}
\sum_{k=1}^{n} \mathbb{E} \left[ \psi_k^{(1)}(\alpha_0) \mid G_{k-1}^n \right] \xrightarrow{P} 0, \tag{60}
\end{equation}

\begin{equation}
\sum_{k=1}^{n} \mathbb{E} \left[ \left( \psi_k^{(1)}(\alpha_0) \right)^2 \mid G_{k-1}^n \right] \xrightarrow{P} 0. \tag{61}
\end{equation}

By using Lemmas 1 and 4 in Gloter and Sørensen (2009), the two terms on the left hand side in (60) and (61) are $O_P(\varepsilon^{2-\frac{1}{v}})$ and $O_P(\varepsilon^{2(1-\frac{1}{v})})$, respectively. Therefore, it follows from $v \geq 2$ that the two properties hold.

Proof of (59). By Lemma 2 in Sørensen and Uchida (2003), one deduces that

\begin{align*}
\varepsilon^{2} \partial^2_\alpha V_{\varepsilon,n}^{(1)}(\alpha) &= -\frac{2}{n} \sum_{k=1}^{n} \left( \partial_\alpha b(X_{t_{k-1}^n}, \alpha) \right) \top \left( b(X_{t_{k-1}^n}, \alpha_0) - b(X_{t_{k-1}^n}, \alpha) \right) \\
&\quad + \frac{2}{n} \sum_{k=1}^{n} \left( \partial_\alpha b(X_{t_{k-1}^n}, \alpha) \right) \top \left( \partial_\alpha b(X_{t_{k-1}^n}, \alpha) \right) \\
&\quad - 2 \sum_{k=1}^{n} \left( \partial^2_\alpha b(X_{t_{k-1}^n}, \alpha) \right) \top P_{1,k}(\alpha_0) \\
&\xrightarrow{P} 2B_1(\alpha, \alpha_0) \quad \text{uniformly in } \alpha.
\end{align*}
By noting that $B_1(\alpha_0, \alpha_0) = J_b(\alpha_0)$, it follows from the consistency of $\hat{\alpha}_{\varepsilon, n}^{(1)}$ and the uniform continuity of $B_1$ that for all $\delta > 0$,

$$
P \left( \sup_{u \in [0, 1]} \left| \varepsilon^2 \partial_{\alpha}^2 V_{\varepsilon, n}^{(1)}(\alpha_0 + u(\hat{\alpha}_{\varepsilon, n}^{(1)} - \alpha_0)) - 2J_b(\alpha_0) \right| > \delta \right) \leq P \left( \sup_{u \in \Theta_\alpha} \left| \varepsilon^2 \partial_{\alpha}^2 V_{\varepsilon, n}^{(1)}(\alpha) - 2B_1(\alpha, \alpha_0) \right| > \frac{\delta}{2} \right) + P \left( \sup_{u \in [0, 1]} \left| 2B_1(\alpha_0 + u(\hat{\alpha}_{\varepsilon, n}^{(1)} - \alpha_0), \alpha_0) - 2J_b(\alpha_0) \right| > \frac{\delta}{2} \right) \to 0 \quad (\varepsilon \to 0, n \to \infty).
$$

This implies (59).

Third, we assume $\varepsilon^{-\frac{l-1}{\nu}}(\hat{\alpha}_{\varepsilon, n}^{(l-1)} - \alpha_0) \overset{P}{\to} 0$ and prove $\varepsilon^{-\frac{l}{\nu}}(\hat{\alpha}_{\varepsilon, n}^{(l)} - \alpha_0) \overset{P}{\to} 0$ for $2 \leq l \leq \nu - 1$. From Taylor’s theorem, we deduce that

$$
- \varepsilon^{-\frac{l}{\nu}} \partial_{\alpha} V_{\varepsilon, n}^{(l)}(\alpha_0) \hat{\alpha}_{\varepsilon, n}^{(l-1)} = \left( \varepsilon^{-\frac{l}{\nu}} \int_0^1 \partial_{\alpha}^2 V_{\varepsilon, n}^{(l)}(\alpha_0 + u(\hat{\alpha}_{\varepsilon, n}^{(l)} - \alpha_0)) du \right) \varepsilon^{-\frac{l}{\nu}}(\hat{\alpha}_{\varepsilon, n}^{(l)} - \alpha_0),
$$

$$
\varepsilon^{-\frac{l}{\nu}} \partial_{\alpha} V_{\varepsilon, n}^{(l)}(\alpha_0) \hat{\alpha}_{\varepsilon, n}^{(l-1)} = \left( \varepsilon^{-\frac{l-1}{\nu}} \int_0^1 \partial_{\alpha}^2 V_{\varepsilon, n}^{(l)}(\alpha_0 + u(\hat{\alpha}_{\varepsilon, n}^{(l-1)} - \alpha_0)) du \right) \varepsilon^{-\frac{l-1}{\nu}}(\hat{\alpha}_{\varepsilon, n}^{(l-1)} - \alpha_0).
$$

Hence, one deduces that

$$
- \varepsilon^{-\frac{l}{\nu}} \partial_{\alpha} V_{\varepsilon, n}^{(l)}(\alpha_0) \alpha_0 = \left( \varepsilon^{-\frac{l-1}{\nu}} \int_0^1 \partial_{\alpha}^2 V_{\varepsilon, n}^{(l)}(\alpha_0 + u(\hat{\alpha}_{\varepsilon, n}^{(l-1)} - \alpha_0)) du \right) \varepsilon^{-\frac{l-1}{\nu}}(\hat{\alpha}_{\varepsilon, n}^{(l-1)} - \alpha_0).
$$

In an analogous manner to the proof of $\varepsilon^{-\frac{1}{\nu}}(\hat{\alpha}_{\varepsilon, n}^{(1)} - \alpha_0) \overset{P}{\to} 0$, it is sufficient to show the following properties:

$$
- \varepsilon^{-\frac{l}{\nu}} \partial_{\alpha} V_{\varepsilon, n}^{(l)}(\alpha_0) \alpha_0 \overset{P}{\to} 0, \quad (63)
$$

$$
\sup_{u \in [0, 1]} \left| \varepsilon^2 \partial_{\alpha}^2 V_{\varepsilon, n}^{(l)}(\alpha_0 + u(\hat{\alpha}_{\varepsilon, n}^{(l)} - \alpha_0)) - 2J_b(\alpha_0) \right| \overset{P}{\to} 0, \quad (64)
$$

$$
\sup_{u \in [0, 1]} \left| \varepsilon^{-\frac{l}{\nu}} \partial_{\alpha}^2 V_{\varepsilon, n}^{(l)}(\alpha_0 + u(\hat{\alpha}_{\varepsilon, n}^{(l-1)} - \alpha_0)) \right| \overset{P}{\to} 0. \quad (65)
$$
Proof of (63). By the definition of the contrast function $V_{\varepsilon, n}(\alpha | \tilde{\alpha})$, we have

$$-\varepsilon^2 \frac{1}{v} \alpha \partial \alpha V_{\varepsilon, n}(\alpha | \alpha_0) = \sum_{k=1}^{n} \psi_k^{(l)}(\alpha_0) + \varepsilon^2 \frac{1}{v} \cdot \frac{1}{n} \sum_{k=1}^{n} R_d(n^{-1}, X_{l_{k-1}^n}), \quad (66)$$

where

$$\psi_k^{(l)}(\alpha_0) = 2 \varepsilon \frac{1}{v} \partial \alpha b(X_{l_{k-1}^n}, \alpha_0) \top P_{1, k}(\alpha_0).$$

Since the second term of the right hand side of (66) converges to 0 in probability, it is sufficient to show the following two properties:

1. $$\sum_{k=1}^{n} \mathbb{E} \left[ \psi_k^{(l)}(\alpha_0) | \mathcal{G}_{k-1}^n \right] \xrightarrow{P} 0., \quad (67)$$
2. $$\sum_{k=1}^{n} \mathbb{E} \left[ \left( \psi_k^{(l)}(\alpha_0) \right)^2 | \mathcal{G}_{k-1}^n \right] \xrightarrow{P} 0. \quad (68)$$

In an analogous manner to the case of $\psi_k^{(1)}$, the two terms on the left hand side in (67) and (68) are $O_P(\varepsilon^2 \frac{1}{v})$ and $O_P(\varepsilon^2 (1 - \frac{1}{v}))$, respectively. Therefore, it follows from $l < v$ that the two properties hold.

Proof of (64). By Lemma 2 in Sørensen and Uchida (2003), one deduces that

$$\varepsilon^2 \partial^2_{\alpha} V_{\varepsilon, n}(\alpha | \hat{\alpha}^{(l-1)}_{\varepsilon, n}) = -2 \sum_{k=1}^{n} \left( \partial^2_{\alpha} b(X_{l_{k-1}^n}, \alpha) \right) \top (b(X_{l_{k-1}^n}, \alpha_0) - b(X_{l_{k-1}^n}, \alpha))$$

$$+ \frac{2}{n} \sum_{k=1}^{n} \left( \partial_{\alpha} b(X_{l_{k-1}^n}, \alpha) \right) \top \left( \partial_{\alpha} b(X_{l_{k-1}^n}, \alpha) \right)$$

$$- 2 \sum_{k=1}^{n} \left( \partial^2_{\alpha} b(X_{l_{k-1}^n}, \alpha) \right) \top P_{1, k}(\alpha_0)$$

$$- 2 \sum_{k=1}^{n} \left( \partial^2_{\alpha} b(X_{l_{k-1}^n}, \alpha) \right) \top Q_{l, k}(\hat{\alpha}^{(l-1)}_{\varepsilon, n})$$

$$\xrightarrow{P} 2B_1(\alpha, \alpha_0) \text{ uniformly in } \alpha.$$

Hence, it follows from the consistency of $\hat{\alpha}^{(l)}_{\varepsilon, n}$ and the uniform continuity of $B_1$ that for all $\delta > 0$,

$$P \left( \sup_{u \in [0, 1]} \left| \varepsilon^2 \partial^2_{\alpha} V_{\varepsilon, n}(\alpha_0 + u(\hat{\alpha}^{(l)}_{\varepsilon, n} - \alpha_0) | \hat{\alpha}^{(l-1)}_{\varepsilon, n}) - 2J_b(\alpha_0) \right| > \delta \right)$$

$$\leq P \left( \sup_{\alpha \in \Theta} \left| \varepsilon^2 \partial^2_{\alpha} V_{\varepsilon, n}(\alpha | \hat{\alpha}^{(l-1)}_{\varepsilon, n}) - 2B_1(\alpha, \alpha_0) \right| > \frac{\delta}{2} \right).$$

Springer
Adaptive inference for small diffusion processes based on… 685

In analogous manners to the proofs of (64) and (65), we can show (70) and (71). For 1 \leq l_1, l_2 \leq p,

\[ \varepsilon^{-1} \partial_{\alpha} V_{\epsilon,n}^{(v)}(\alpha | \tilde{\alpha}) = 2 \varepsilon^{-1} \int_0^1 \partial_{\alpha} V_{\epsilon,n}^{(v)}(\alpha | \alpha_0 + u(\tilde{\alpha}_{\epsilon,n} - \alpha_0)) \frac{du}{\varepsilon} (\tilde{\alpha}_{\epsilon,n} - \alpha_0) \]

\[ = \left( 2 \int_0^1 \partial_{\alpha} V_{\epsilon,n}^{(v)}(\alpha | \alpha_0 + u(\tilde{\alpha}_{\epsilon,n} - \alpha_0)) \frac{du}{\varepsilon} \right) \varepsilon^{-1} (\tilde{\alpha}_{\epsilon,n} - \alpha_0). \]

Therefore, we have (65).

Next, we prove \( \varepsilon^{-1} (\tilde{\alpha}_{\epsilon,n} - \alpha_0) = O_P(1) \). In the same manner as the derivation of (62), it holds from Taylor’s theorem that

\[ - \varepsilon \partial_{\alpha} V_{\epsilon,n}^{(v)}(\alpha_0 | \alpha_0) = 2 \varepsilon^{-1} \int_0^1 \partial_{\alpha \tilde{\alpha}} V_{\epsilon,n}^{(v)}(\alpha_0 | \alpha_0 + u(\tilde{\alpha}_{\epsilon,n} - \alpha_0)) \frac{du}{\varepsilon} (\tilde{\alpha}_{\epsilon,n} - \alpha_0) \]

\[ = \left( 2 \int_0^1 \partial_{\alpha} V_{\epsilon,n}^{(v)}(\alpha_0 + u(\tilde{\alpha}_{\epsilon,n} - \alpha_0)) \frac{du}{\varepsilon} \right) \varepsilon^{-1} (\tilde{\alpha}_{\epsilon,n} - \alpha_0). \]

In order to prove \( \varepsilon^{-1} (\tilde{\alpha}_{\epsilon,n} - \alpha_0) = O_P(1) \), it is sufficient to show the following properties:

\[ - \varepsilon \partial_{\alpha} V_{\epsilon,n}^{(v)}(\alpha_0 | \alpha_0) = O_P(1), \]  
(69)

\[ \sup_{u \in [0,1]} \left| \varepsilon^{2-1} \partial_{\alpha \alpha} V_{\epsilon,n}^{(v)}(\alpha_0 + u(\tilde{\alpha}_{\epsilon,n} - \alpha_0)) - 2J_b(\alpha_0) \right| \rightarrow 0, \]  
(70)

\[ \sup_{u \in [0,1]} \left| \varepsilon^{2-1} \partial_{\alpha \alpha} V_{\epsilon,n}^{(v)}(\alpha_0 + u(\tilde{\alpha}_{\epsilon,n} - \alpha_0)) \right| \rightarrow 0. \]  
(71)

In analogous manners to the proofs of (64) and (65), we can show (70) and (71). For (69), it follows from Lemma 4-2 in Gloter and Sørensen (2009) that

\[ - \varepsilon \partial_{\alpha} V_{\epsilon,n}^{(v)}(\alpha_0 | \alpha_0) = 2 \varepsilon^{-1} \sum_{k=1}^n \partial_{\alpha} b(X_{ik-1}, \alpha_0)^\top (P_{1,k}(\alpha_0) - Q_{v,k}(\alpha_0)) \]

\[ = 2 \varepsilon^{-1} \sum_{k=1}^n \partial_{\alpha} b(X_{ik-1}, \alpha_0)^\top P_{v,k}(\alpha_0). \]
Finally, we show the asymptotic normality for \( \hat{\alpha}_{\varepsilon, n} \) in the same manner as the proof of Lemma 1. Noting that 
\[ -\varepsilon \partial_u V_{\varepsilon, n}^{(v)}(\alpha_0|\alpha_0) = \sum_{k=1}^{n} \varepsilon \partial_k \varepsilon V_{\varepsilon, n}^{(1)}(\alpha_0) \] for \( 1 \leq l \leq p \), we can show 
\[ -\varepsilon \partial_u V_{\varepsilon, n}^{(v)}(\alpha_0|\alpha_0) \xrightarrow{d} -\varepsilon \partial_u V_{\varepsilon, n}^{(1)}(\alpha_0|\alpha_0) \] from (70) and this convergence, we complete the proof.

**Proof of Theorem 2** We prove this theorem in the same order as the proof of Theorem 1.

1st step. We show \( \hat{\alpha}_{\varepsilon, n} \xrightarrow{P} \alpha_0 \). Noting that \( V_{\varepsilon, n}^{(1)}(\beta|\alpha) = U_{\varepsilon, n}^{(1)}(\beta|\alpha) \), we can utilize the proof of Theorem 1. In particular, by replacing \( U_{\varepsilon, n}^{(1)}(\alpha, \theta_0) \) with \( V_{\varepsilon, n}^{(1)}(\alpha, \theta_0) \) and \( (\hat{\alpha}_{\varepsilon, n}, \hat{\beta}_{\varepsilon, n}) \) in the 1st step of the proof of Theorem 1, we obtain the result.

2nd step. We prove \( \hat{\alpha}_{\varepsilon, n} \xrightarrow{P} \alpha_0 \). It follows from Lemma 4 in Gloter and Sørensen (2009) and the consistency of \( \hat{\beta}_{\varepsilon, n} \) that

\[ e^2 \left( V_{\varepsilon, n}^{(v+2)}(\alpha|\hat{\alpha}_{\varepsilon, n}, \hat{\beta}_{\varepsilon, n}) - V_{\varepsilon, n}^{(v+2)}(\alpha_0|\hat{\alpha}_{\varepsilon, n}, \hat{\beta}_{\varepsilon, n}) \right) \]
\[ = 2 \sum_{k=1}^{n} \left( b(X_{t_k}, \alpha) - b(X_{t_k}, \alpha_0) \right)^\top \left[ \sigma \sigma^\top \right]^{-1} \left( X_{t_k} - \hat{\beta}_{\varepsilon, n} P_{1,k}(\alpha_0) \right) \]
\[ - 2 \sum_{k=1}^{n} \left( b(X_{t_k}, \alpha) - b(X_{t_k}, \alpha_0) \right)^\top \left[ \sigma \sigma^\top \right]^{-1} \left( X_{t_k} - \hat{\beta}_{\varepsilon, n} Q_{V,k}(\hat{\alpha}_{\varepsilon, n}) \right) \]
\[ + \frac{1}{n} \sum_{k=1}^{n} \left( b(X_{t_k}, \alpha) - b(X_{t_k}, \alpha_0) \right)^\top \left[ \sigma \sigma^\top \right]^{-1} \left( X_{t_k} - \beta_{0} \right) \]
\[ \times \left( b(X_{t_k}, \alpha) - b(X_{t_k}, \alpha_0) \right) \]
\[ + \frac{1}{n} \sum_{k=1}^{n} \left( b(X_{t_k}, \alpha) - b(X_{t_k}, \alpha_0) \right)^\top \left( \left[ \sigma \sigma^\top \right]^{-1} \left( X_{t_k} - \hat{\beta}_{\varepsilon, n} \right) - \left[ \sigma \sigma^\top \right]^{-1} \left( X_{t_k} - \beta_{0} \right) \right) \]
\[ \left( b(X_{t_k}, \alpha) - b(X_{t_k}, \alpha_0) \right) \]
\[ \xrightarrow{P} U_3(\alpha, \theta_0) \] uniformly in \( \alpha \).

Therefore, in an analogous manner to the proof of the consistency for \( \hat{\alpha}_{\varepsilon, n} \), we have \( \hat{\alpha}_{\varepsilon, n} \xrightarrow{P} \alpha_0 \).

3rd step. We prove the asymptotic normality for \( \hat{\beta}_{\varepsilon, n} \). Using Taylor’s theorem, we have the following expansions:

\[ -\varepsilon \partial_u V_{\varepsilon, n}^{(v+2)}(\alpha_0|\hat{\alpha}_{\varepsilon, n}, \hat{\beta}_{\varepsilon, n}) \]
\[ = \left( e^2 \int_0^1 \partial_u \varepsilon V_{\varepsilon, n}^{(v+2)}(\alpha_0 + u(\hat{\alpha}_{\varepsilon, n} - \alpha_0)|\hat{\alpha}_{\varepsilon, n}, \hat{\beta}_{\varepsilon, n}) du \right) \varepsilon^{-1}(\hat{\alpha}_{\varepsilon, n} - \alpha_0), \]
\[ - \frac{1}{\sqrt{n}} \partial_\beta V_{\varepsilon, n}^{(v+1)}(\beta_0|\hat{\alpha}_{\varepsilon, n}) \]

\( \triangleleft \) Springer
proof of Theorem 1.

In an analogous manner to the proof of Theorem 1 in Sørensen and Uchida (2003), it follows from Lemma 2 in Sørensen and Uchida (2003), Lemma 4 in Gloter and Sørensen (2009) and the consistency of \( \hat{\beta}_{e,n} \) that

\[
\varepsilon \partial_{\alpha} V_{e,n}^{(v+2)}(\alpha_0|\hat{\alpha}_{e,n}, \hat{\beta}_{e,n}) = \varepsilon \partial_{\alpha} V_{e,n}^{(v+2)}(\alpha_0|\hat{\alpha}_{e,n}, \hat{\beta}_{e,n}) + \left( \frac{\varepsilon}{n} \int_0^1 \partial_{\alpha} V_{e,n}^{(v+2)}(\alpha_0|\hat{\alpha}_{e,n}, \hat{\beta}_{e,n} + u(\hat{\beta}_{e,n} - \beta_0))du \right) \sqrt{n}(\hat{\beta}_{e,n} - \beta_0).
\]

Using these expressions, we calculate that

\[
\Gamma_{e,n}^2 = C_{e,n}^2 \Lambda_{e,n}^2,
\]

where

\[
\Gamma_{e,n}^2 := \left( -\varepsilon \partial_{\alpha} V_{e,n}^{(v+2)}(\alpha_0|\hat{\alpha}_{e,n}, \beta_0) - \frac{1}{\sqrt{n}} \partial_{\beta} V_{e,n}^{(v+2)}(\alpha_0|\hat{\alpha}_{e,n}, \beta_0) \right), \quad \Lambda_{e,n}^2 := \left( \frac{\varepsilon}{\sqrt{n}} \int_0^1 \partial_{\alpha} V_{e,n}^{(v+2)}(\alpha_0|\hat{\alpha}_{e,n}, \hat{\beta}_{e,n} + u(\hat{\beta}_{e,n} - \beta_0))du \right).
\]

In an analogous manner to the proof of Theorem 1 in Sørensen and Uchida (2003), it is sufficient to show the following convergences:

\[
\sup_{u \in [0, 1]} \left| \varepsilon^2 \partial_{\alpha} V_{e,n}^{(v+2)}(\alpha_0 + u(\hat{\alpha}_{e,n} - \alpha_0)|\hat{\alpha}_{e,n}, \hat{\beta}_{e,n}) - 2I_b(\theta_0) \right| \to 0, \quad (72)
\]

\[
\sup_{u \in [0, 1]} \left| \frac{1}{n} \partial_{\beta} V_{e,n}^{(v+1)}(\beta_0 + u(\hat{\beta}_{e,n} - \beta_0)|\hat{\alpha}_{e,n}, \hat{\beta}_{e,n}) - 2I_\sigma(\beta_0) \right| \to 0, \quad (73)
\]

\[
\sup_{u \in [0, 1]} \left| \varepsilon \partial_{\alpha} V_{e,n}^{(v+2)}(\alpha_0|\hat{\alpha}_{e,n}, \beta_0 + u(\hat{\beta}_{e,n} - \beta_0)) \right| \to 0, \quad (74)
\]

\[
- \varepsilon \partial_{\alpha} V_{e,n}^{(v+2)}(\alpha_0|\hat{\alpha}_{e,n}, \beta_0) \to N_p(0, 4I_b(\theta_0)), \quad (75)
\]

\[
- \varepsilon \partial_{\beta} V_{e,n}^{(v+1)}(\beta_0|\hat{\alpha}_{e,n}) \to N_q(0, 4I_\sigma(\beta_0)). \quad (76)
\]

By the definitions of \( U_{e,n,v}^{(2)} \) and \( V_{e,n}^{(v+1)} \), we have already shown (73) and (76) in the proof of Theorem 1.

Proof of (72). It follows from Lemma 2 in Sørensen and Uchida (2003), Lemma 4 in Gloter and Sørensen (2009) and the consistency of \( \hat{\beta}_{e,n} \) that

\[
\varepsilon^2 \partial_{\alpha} V_{e,n}^{(v+2)}(\alpha|\hat{\alpha}_{e,n}, \hat{\beta}_{e,n}) = -2 \sum_{k=1}^n \left( \partial_{\alpha}^2 b(X_{i_{k-1}^n}, \alpha) \right)^\top [\sigma \sigma^\top]^{-1} (X_{i_{k-1}^n}, \beta_{e,n}) P_{1,k}(\alpha_0) + 2 \sum_{k=1}^n \left( \partial_{\alpha}^2 b(X_{i_{k-1}^n}, \alpha) \right)^\top [\sigma \sigma^\top]^{-1} (X_{i_{k-1}^n}, \hat{\beta}_{e,n}) Q_{v,k}(\hat{\alpha}_{e,n}).
\]
\[ + \frac{2}{n} \sum_{k=1}^{n} \left( \partial_{\alpha} b(X_{n k}^{n}, \alpha) \right)^{\top} \left[ \sigma \sigma^{\top} \right]^{-1} (X_{n k-1}^{n}, \beta) \left( b(X_{n k}^{n}, \alpha) - b(X_{n k-1}^{n}, \alpha) \right) \]

\[ + \frac{2}{n} \sum_{k=1}^{n} \left( \partial_{\alpha} b(X_{n k}^{n}, \alpha) \right)^{\top} \left[ \sigma \sigma^{\top} \right]^{-1} (X_{n k-1}^{n}, \beta_{e,n}) - \left[ \sigma \sigma^{\top} \right]^{-1} (X_{n k-1}^{n}, \beta_{0}) \]

\[ \left( b(X_{n k}^{n}, \alpha) - b(X_{n k-1}^{n}, \alpha) \right) \]

\[ + \frac{2}{n} \sum_{k=1}^{n} \left( \partial_{\alpha} b(X_{n k}^{n}, \alpha) \right)^{\top} \left[ \sigma \sigma^{\top} \right]^{-1} (X_{n k-1}^{n}, \beta_{e,n}) - \left[ \sigma \sigma^{\top} \right]^{-1} (X_{n k-1}^{n}, \beta_{0}) \]

\[ \left( \partial_{\alpha} b(X_{n k}^{n}, \alpha) \right) \]

\[ \overset{p}{\to} 2B_{2}(\alpha, \theta_{0}) \text{ uniformly in } \alpha. \]

In an analogous manner to the proof of (36), it follows from the consistency for \( \hat{\alpha}_{e,n} \) and the uniform continuity of \( B_{2} \) that for all \( \delta > 0 \),

\[ P \left( \sup_{u \in [0, 1]} \left| 2 \partial_{\alpha}^{2} V_{e,n}^{(v+2)} (\alpha_{0} + u(\hat{\alpha}_{e,n} - \alpha_{0})) \hat{\alpha}_{e,n}^{(v)} \beta_{e,n}) - 2I_{b}(\theta_{0}) \right| > \delta \right) \]

\[ \leq P \left( \sup_{\alpha \in \Theta_{\alpha}} \left| 2 \partial_{\alpha}^{2} V_{e,n}^{(v+2)} (\alpha \hat{\alpha}_{e,n}^{(v)} \beta_{e,n}) - 2B_{2}(\alpha, \theta_{0}) \right| > \frac{\delta}{2} \right) \]

\[ + P \left( \sup_{u \in [0, 1]} \left| 2B_{2}(\alpha_{0} + u(\hat{\alpha}_{e,n} - \alpha_{0}), \theta_{0}) - 2I(\theta_{0}) \right| > \frac{\delta}{2} \right) \]

\[ \to 0 \quad (\varepsilon \to 0, n \to \infty). \]

This implies (72).

Proof of (74). By using the Lipschitz continuity of \( n^{2} Q_{v,k}(\alpha) \), Lemma 2 in this paper and Lemma 4-(2) in Gloter and Sørensen (2009), it holds that

\[ \frac{\varepsilon}{\sqrt{n}} \partial_{\alpha}^{2} V_{e,n}^{(v+2)} (\alpha_{0} \hat{\alpha}_{e,n}^{(v)} \beta) \]

\[ = -2(\varepsilon \sqrt{n})^{-1} \sum_{k=1}^{n} \left( \partial_{\alpha} b(X_{n k-1}^{n}, \alpha) \right)^{\top} \left( \partial_{\beta} [\sigma \sigma^{\top}]^{-1} (X_{n k-1}^{n}, \beta) \right) P_{v,k}(\alpha_{0}) \]

\[ + 2(\varepsilon \sqrt{n})^{-1} \sum_{k=1}^{n} \left( \partial_{\alpha} b(X_{n k-1}^{n}, \alpha) \right)^{\top} \left( \partial_{\beta} [\sigma \sigma^{\top}]^{-1} (X_{n k-1}^{n}, \beta) \right) \]

\[ \left( Q_{v,k}(\hat{\alpha}_{e,n}^{(v)}) - Q_{v,k}(\alpha_{0}) \right) \]

\[ \overset{p}{\to} 0 \text{ uniformly in } \beta. \]
Hence, we have (74).

Proof of (75). For $1 \leq l \leq p$, it follows from the Lipschitz continuity of $n^2 Q_{v,k}(\alpha)$, Lemma 2 in this paper and Lemma 4-(2) in Gloter and Sørensen (2009) that

$$-\varepsilon \partial_{\alpha l} V_{\varepsilon,n}^{(v+2)}(\alpha_0 | \hat{\alpha}_{\varepsilon,n}^{(v)}, \beta_0)$$

$$= 2\varepsilon^{-1} \sum_{k=1}^{n} \left( \partial_{\alpha l} b(X_{t_{k-1}^{n}}, \alpha_0) \right)^\top \left[ \sigma \sigma^\top \right]^{-1} (X_{t_{k-1}^{n}}, \beta_0) P_{v,k}(\alpha_0) - 2\varepsilon^{-1} \sum_{k=1}^{n} \left( \partial_{\alpha l} b(X_{t_{k-1}^{n}}, \alpha_0) \right)^\top \left[ \sigma \sigma^\top \right]^{-1} (X_{t_{k-1}^{n}}, \beta_0) \left( Q_{v,k}(\hat{\alpha}_{\varepsilon,n}^{(v)}) - Q_{v,k}(\alpha_0) \right)$$

$$= \sum_{k=1}^{n} \varepsilon^{-1} l_{k,1}(\theta_0) + o P(1).$$

Since the main term of $-\varepsilon \partial_{\alpha l} V_{\varepsilon,n}^{(v+2)}(\alpha_0 | \hat{\alpha}_{\varepsilon,n}^{(v)}, \beta_0)$ is the same as that of $-\varepsilon \partial_{\alpha l} U_{\varepsilon,n,v}^{(3)}(\alpha_0 | \beta_0)$, we utilize the results in the proof of Theorem 1 and complete the proof. □

For the proof of theorem 3, we first show the following lemma.

**Lemma 3** Assume [A1]–[A4] and [B]. Then it follows that

$$\varepsilon^{-1} (\tilde{\alpha}_{\varepsilon,n}^{(1)} - \hat{\alpha}_{\varepsilon,n}^{(1)}) \rightarrow_d (J_b^{-1}(\alpha_0) - G_1^f(\alpha_0)) Z \quad \text{(under } H_0^{(1)})$$

$$\sqrt{n}(\tilde{\beta}_{\varepsilon,n} - \bar{\tilde{\beta}}_{\varepsilon,n}) \rightarrow_d (I^{-1}_\sigma(\beta_0) - G_2^f(\beta_0)) Y \quad \text{(under } H_0^{(2)})$$

as $\varepsilon \rightarrow 0$ and $n \rightarrow \infty$, where $Z$ and $Y$ are random vectors which have normal distribution $N_p(0, K_b(\theta_0))$ and $N_q(0, I_\sigma(\beta_0))$, respectively.

**Proof of Lemma 3** It follows from Taylor’s theorem that

$$\varepsilon \partial_{\alpha} U_{\varepsilon,n,v}^{(1)}(\tilde{\alpha}_{\varepsilon,n}^{(1)}, H_0) = J_n(\tilde{\alpha}_{\varepsilon,n}^{(1)}, H_0, \alpha_0) \varepsilon^{-1} (\tilde{\alpha}_{\varepsilon,n}^{(1)}, H_0 - \alpha_0),$$

where

$$J_n(\alpha, \tilde{\alpha}) = \int_0^1 \varepsilon^2 \partial_{\alpha}^2 U_{\varepsilon,n,v}^{(1)}(\tilde{\alpha} + u(\alpha - \tilde{\alpha})) du.$$
Hence, it follows from Lemma 1 and \( G_1'(\alpha) \partial_\alpha U_{e,n,v}(\tilde{\alpha}_{e,n}, H_0) = 0 \) that under \( H_0^{(1)} \),

\[
-\varepsilon G_1'(\alpha) \partial_\alpha U_{e,n,v}(\alpha) = G_1'(\alpha) J_n(\tilde{\alpha}_{e,n}, H_0, \alpha) \varepsilon^{-1}(\tilde{\alpha}_{e,n}, H_0 - \alpha) \\
= G_1'(\alpha)(2 J_b(\alpha) + o_p(1)) \varepsilon^{-1}(\tilde{\alpha}_{e,n}, H_0 - \alpha) \\
= 2 \varepsilon^{-1}(\tilde{\alpha}_{e,n}, H_0 - \alpha) + o_p(1).
\]

Therefore, one has that under \( H_0^{(1)} \),

\[
- \varepsilon \partial_\alpha U_{e,n,v}(\tilde{\alpha}_{e,n}, H_0) = \varepsilon \partial_\alpha U_{e,n,v}(\alpha) - J_n(\tilde{\alpha}_{e,n}, H_0, \alpha) \varepsilon^{-1}(\tilde{\alpha}_{e,n}, H_0 - \alpha) \\
= \varepsilon \partial_\alpha U_{e,n,v}(\alpha) - J_b(\alpha) \cdot 2 \varepsilon^{-1}(\tilde{\alpha}_{e,n}, H_0 - \alpha) + o_p(1) \\
= \varepsilon \partial_\alpha U_{e,n,v}(\alpha) + J_b(\alpha) G_1'(\alpha) \varepsilon \partial_\alpha U_{e,n,v}(\alpha) + o_p(1) \\
= - \left( E_p - J_b(\alpha) G_1'(\alpha) \right) \varepsilon \partial_\alpha U_{e,n,v}(\alpha) + o_p(1) \\
\rightarrow \left( E_p - J_b(\alpha) G_1'(\alpha) \right) \mathcal{Z}.
\]

On the other hand, it follows from Taylor’s theorem and the consistency of estimator for \( \alpha \) that

\[
- \varepsilon \partial_\alpha U_{e,n,v}(\tilde{\alpha}_{e,n}, H_0) = J_n(\tilde{\alpha}_{e,n}, H_0, \tilde{\alpha}_{e,n}) \varepsilon^{-1}(\tilde{\alpha}_{e,n}, H_0 - \tilde{\alpha}_{e,n}) \\
= 2 J_b(\alpha) \varepsilon^{-1}(\tilde{\alpha}_{e,n}, H_0 - \tilde{\alpha}_{e,n}) + o_p(1).
\]

As a result, from \([A4]\), it holds that

\[
\varepsilon^{-1}(\tilde{\alpha}_{e,n} - \tilde{\alpha}_{e,n}, H_0) \xrightarrow{d} J_b^{-1}(\alpha) \left( E_p - J_b(\alpha) G_1'(\alpha) \right) \mathcal{Z} \\
= \left( J_b^{-1}(\alpha) - G_1'(\alpha) \right) \mathcal{Z}.
\]

For the second statement, it follows from Taylor’s theorem with respect to \( \beta \) that

\[
\frac{1}{\sqrt{n}} \partial_\beta U_{e,n,v}(\tilde{\beta}_{e,n}, H_0, \tilde{\alpha}_{e,n}) - \frac{1}{\sqrt{n}} \partial_\beta U_{e,n,v}(\beta_0, \tilde{\alpha}_{e,n}) = I_n(\beta, \tilde{\beta}_{e,n}, \beta_0, \tilde{\alpha}_{e,n}) \sqrt{n}(\tilde{\beta}_{e,n} - \beta_0),
\]

where

\[
I_n(\beta, \tilde{\beta}, \beta_0) = \int_0^1 \frac{1}{n} \partial_\beta^2 U_{e,n,v}(\beta + u(\beta - \tilde{\beta}), \alpha) du.
\]

Moreover, It follows from Taylor’s theorem with respect to \( \alpha \) that

\[
\frac{1}{\sqrt{n}} \partial_\beta U_{e,n,v}(\beta_0, \tilde{\alpha}_{e,n}) = \frac{1}{\sqrt{n}} \partial_\beta U_{e,n,v}(\beta_0, \tilde{\alpha}_{e,n}) + I_n(\alpha, \beta_0, \tilde{\alpha}_{e,n}, \alpha) \varepsilon^{-1}(\tilde{\alpha}_{e,n} - \alpha). \]
where

\[ I_{n, \alpha \beta}(\beta | \alpha, \tilde{\alpha}) = \int_{0}^{1} \frac{\varepsilon}{\sqrt{n}} \frac{\partial^{2}}{\partial \alpha \partial \beta} U_{\varepsilon, n, v}^{(2)}(\beta | \tilde{\alpha} + u(\alpha - \tilde{\alpha})) du. \]

From (37)-(39), one has that under \( H_{0}^{(2)} \),

\[ I_{n, \alpha \beta}(\tilde{\beta}_{\varepsilon, n}, \beta_{0} | \tilde{\alpha}(1)_{\varepsilon, n}) \overset{P}{\to} 2I_{\sigma}(\beta_{0}), \quad I_{n, \alpha \beta}(\beta_{0} | \tilde{\alpha}(1)_{\varepsilon, n}, \alpha_{0}) \overset{P}{\to} 0, \]

\[ - \frac{1}{\sqrt{n}} \partial_{\alpha} U_{\varepsilon, n, v}^{(2)}(\beta_{0} | \alpha_{0}) \overset{d}{\to} 2Y. \]

We can calculate

\[ G_{2}^{*}(\beta_{0}) I_{\sigma}(\beta_{0}) \sqrt{n}(\tilde{\beta}_{H_{0}}^{H_{0}} - \beta_{0}) = \sqrt{n}(\tilde{\beta}_{H_{0}}^{H_{0}} - \beta_{0}), \]

and from the facts that \( \sqrt{n}(\tilde{\beta}_{H_{0}}^{H_{0}} - \beta_{0}) = O_{P}(1) \) and \( G_{2}^{*}(\beta_{0}) \partial_{\alpha} U_{\varepsilon, n, v}^{(2)}(\tilde{\beta}_{H_{0}}^{H_{0}}, \tilde{\alpha}(1)_{\varepsilon, n}) = 0 \), it follows that under \( H_{0}^{(2)} \),

\[ - \frac{1}{\sqrt{n}} G_{2}^{*}(\beta_{0}) \partial_{\alpha} U_{\varepsilon, n, v}^{(2)}(\beta_{0} | \alpha_{0}) = G_{2}^{*}(\beta_{0}) I_{n, \beta}(\tilde{\beta}_{H_{0}}^{H_{0}}, \beta_{0} | \tilde{\alpha}(1)_{\varepsilon, n}) \sqrt{n}(\tilde{\beta}_{H_{0}}^{H_{0}} - \beta_{0}) + o_{P}(1) \]

\[ = 2G_{2}^{*}(\beta_{0}) I_{\sigma}(\beta_{0}) \sqrt{n}(\tilde{\beta}_{H_{0}}^{H_{0}} - \beta_{0}) + o_{P}(1) \]

\[ = 2 \sqrt{n}(\tilde{\beta}_{H_{0}}^{H_{0}} - \beta_{0}) + o_{P}(1). \]

Therefore, one has that under \( H_{0}^{(2)} \),

\[ - \frac{1}{\sqrt{n}} \partial_{\beta} U_{\varepsilon, n, v}^{(2)}(\tilde{\beta}_{H_{0}}^{H_{0}}, \tilde{\alpha}(1)_{\varepsilon, n}) \]

\[ = - \frac{1}{\sqrt{n}} \partial_{\beta} U_{\varepsilon, n, v}^{(2)}(\beta_{0} | \alpha_{0}) - I_{n, \beta}(\tilde{\beta}_{H_{0}}^{H_{0}}, \beta_{0} | \tilde{\alpha}(1)_{\varepsilon, n}) \sqrt{n}(\tilde{\beta}_{H_{0}}^{H_{0}} - \beta_{0}) + o_{P}(1) \]

\[ = - \frac{1}{\sqrt{n}} \partial_{\beta} U_{\varepsilon, n, v}^{(2)}(\beta_{0} | \alpha_{0}) - I_{\sigma}(\beta_{0}) \cdot 2 \sqrt{n}(\tilde{\beta}_{H_{0}}^{H_{0}} - \beta_{0}) + o_{P}(1) \]

\[ = - \left( E_{q} - I_{\sigma}(\beta_{0}) G_{2}^{*}(\beta_{0}) \right) \frac{1}{\sqrt{n}} \partial_{\beta} U_{\varepsilon, n, v}^{(2)}(\beta_{0} | \alpha_{0}) + o_{P}(1) \]

\[ \overset{d}{\to} 2 \left( E_{q} - I_{\sigma}(\beta_{0}) G_{2}^{*}(\beta_{0}) \right) Y. \]

On the other hand, it follows from Taylor’s theorem and the consistency of estimator for \((\alpha, \beta)\) that

\[ - \frac{1}{\sqrt{n}} \partial_{\beta} U_{\varepsilon, n, v}^{(2)}(\tilde{\beta}_{H_{0}}^{H_{0}}, \tilde{\alpha}(1)_{\varepsilon, n}) = I_{n, \beta}(\tilde{\beta}_{H_{0}}^{H_{0}}, \tilde{\beta}_{H_{0}}^{H_{0}}, \tilde{\alpha}(1)_{\varepsilon, n}) \sqrt{n}(\tilde{\beta}_{H_{0}}^{H_{0}} - \tilde{\beta}_{H_{0}}^{H_{0}}) \]

\[ = 2I_{\sigma}(\beta_{0}) \sqrt{n}(\tilde{\beta}_{H_{0}}^{H_{0}} - \tilde{\beta}_{H_{0}}^{H_{0}}) + o_{P}(1). \]
As a result, from [A4], it holds that
\[
\sqrt{n}(\tilde{\beta}_{e,n} - \tilde{\beta}^H_{e,n}) \overset{d}{\to} I_{\sigma}^{-1}(\beta_0) \left( E_q - I_{\sigma}(\beta_0) G^*_2(\beta_0) \right) Y
= \left( I_{\sigma}^{-1}(\beta_0) - G^*_2(\beta_0) \right) Y.
\]

**Proof of Theorem 3** One has from Taylor’s theorem that
\[
\tilde{\Lambda}_n^{(1)} = U^{(1)}_{e,n,v} \left( \tilde{\alpha}^{(1)}_{e,n}, H_0 \right) - U^{(1)}_{e,n,v} \left( \tilde{\alpha}^{(1)}_{e,n} \right)
= \left( \varepsilon^{-1}(\tilde{\alpha}^{(1)}_{e,n} - \tilde{\alpha}^{(1)}_{e,n}, H_0) \right)^T
\left( \int_0^1 (1 - u)\varepsilon^2 \partial^2_{\alpha} U^{(1)}_{e,n,v} (\tilde{\alpha}^{(1)}_{e,n} + u(\tilde{\alpha}^{(1)}_{e,n}, H_0 - \tilde{\alpha}^{(1)}_{e,n}))du \right) \varepsilon^{-1}(\tilde{\alpha}^{(1)}_{e,n} - \tilde{\alpha}^{(1)}_{e,n}, H_0).
\]

From (21), we have that under \( H^{(1)}_0 \),
\[
\int_0^1 (1 - u)\varepsilon^2 \partial^2_{\alpha} U^{(1)}_{e,n,v} (\tilde{\alpha}^{(1)}_{e,n} + u(\tilde{\alpha}^{(1)}_{e,n}, H_0 - \tilde{\alpha}^{(1)}_{e,n}))du \overset{P}{\to} J_b(\alpha_0).
\]

Therefore, Lemma 3 and the continuous mapping theorem yield that
\[
\tilde{\Lambda}_n^{(1)} \overset{d}{\to} Z^T \left( J_b^{-1}(\alpha_0) - G^*_1(\alpha_0) \right) J_b(\alpha_0) \left( J_b^{-1}(\alpha_0) - G^*_1(\alpha_0) \right) Z
= Z^T \left( J_b^{-1}(\alpha_0) - G^*_1(\alpha_0) \right) Z \sim \pi_r.
\]

Next, Taylor’s theorem implies that
\[
\tilde{\Lambda}_n^{(2)} = U^{(2)}_{e,n,v} \left( \tilde{\beta}^{H_0}_{e,n} | \tilde{\alpha}^{(1)}_{e,n} \right) - U^{(2)}_{e,n,v} \left( \tilde{\beta}_{e,n} | \tilde{\alpha}^{(1)}_{e,n} \right)
= \left( \sqrt{n}(\tilde{\beta}_{e,n} - \tilde{\beta}^{H_0}_{e,n}) \right)^T
\left( \int_0^1 \frac{1 - u}{n} \partial^2_{\beta} U^{(2)}_{e,n,v} (\tilde{\beta}_{e,n} + u(\tilde{\beta}^{H_0}_{e,n} - \tilde{\beta}_{e,n} | \tilde{\alpha}^{(1)}_{e,n}))du \right) \sqrt{n}(\tilde{\beta}_{e,n} - \tilde{\beta}^{H_0}_{e,n}).
\]

From (37), we have that under \( H^{(2)}_0 \),
\[
\int_0^1 \frac{1 - u}{n} \partial^2_{\beta} U^{(2)}_{e,n,v} (\tilde{\beta}_{e,n} + u(\tilde{\beta}^{H_0}_{e,n} - \tilde{\beta}_{e,n} | \tilde{\alpha}^{(1)}_{e,n}))du \overset{P}{\to} I_\sigma(\beta_0).
\]

Therefore, we define \( Y' \) as \( Y = I_{\sigma}^{1/2} Y' \), and Lemma 3 yields that
\[
\tilde{\Lambda}_n^{(2)} \overset{d}{\to} Y^T \left( I_{\sigma}^{-1}(\beta_0) - G^*_2(\beta_0) \right) I_\sigma(\beta_0) \left( I_{\sigma}^{-1}(\beta_0) - G^*_2(\beta_0) \right) Y
\]
\[
Y' \equiv Y'^{1 \top} I_{\sigma}^{1} (\beta_0) \left( I_{\sigma}^{-1} (\beta_0) - G_{2}^{\varepsilon} (\beta_0) \right) I_{\sigma}^{1} (\beta_0) Y'
\]
\[
= Y'^{1 \top} (E_{p} - I_{\sigma}^{1} (\beta_0) G_{2}^{\varepsilon} (\beta_0) I_{\sigma}^{1} (\beta_0)) Y' \sim \chi_{\varepsilon}^{2}.
\]

In order to prove Theorem 4, we show the following lemma.

**Lemma 4** Assume \([A1]–[A4], [B] and [C]\). Then it follows that

\[
\tilde{\alpha}_{e,n}^{(1), H_{0}} \xrightarrow{P} \alpha_{0}^{H_{0}} \quad \text{(under } H_{1}^{(1)}), \quad \tilde{\beta}_{e,n}^{H_{0}} \xrightarrow{P} \beta_{0}^{H_{0}} \quad \text{(under } H_{1}^{(2)})
\]
as \(e \to 0\) and \(n \to \infty\).

**Proof of Lemma 4** It follows from (12) that under \(H_{1}^{(1)}\),

\[
\sup_{\alpha \in \Theta_{\alpha}} \left| e^{2} \left( U_{e,n,v}^{(1)} (\alpha) - U_{e,n,v}^{(1)} (\alpha_{0}) \right) - U_{1} (\alpha, \alpha_{0}) \right| \xrightarrow{P} 0.
\]

Assumption \([C]-i\) implies the following: For any \(\delta_{1} > 0\), there exists \(\delta' > 0\) such that for any \(\alpha \in \Theta_{\alpha}^{H_{0}}\),

\[
|\alpha - \alpha_{0}^{H_{0}}| \geq \delta_{1} \implies U_{1} (\alpha, \alpha_{0}) - U_{1} (\alpha_{0}^{H_{0}}, \alpha_{0}) > \delta' . \tag{77}
\]

By the definition of \(\tilde{\alpha}_{e,n}^{(1), H_{0}}\) and (77), it holds that

\[
P \left( |\tilde{\alpha}_{e,n}^{(1), H_{0}} - \alpha_{0}^{H_{0}}| \geq \delta_{1} \right)
\leq P \left( U_{1} (\tilde{\alpha}_{e,n}^{(1), H_{0}}, \alpha_{0}) - U_{1} (\alpha_{0}^{H_{0}}, \alpha_{0}) > \delta' \right)
\]
\[
= P \left\{ U_{1} (\tilde{\alpha}_{e,n}^{(1), H_{0}}, \alpha_{0}) - \varepsilon^{2} \left( U_{e,n,v}^{(1)} (\tilde{\alpha}_{e,n}^{(1), H_{0}}) - U_{e,n,v}^{(1)} (\alpha_{0}) \right) \right\}
\]
\[
+ \varepsilon^{2} \left( U_{e,n,v}^{(1)} (\tilde{\alpha}_{e,n}^{(1), H_{0}}) - U_{e,n,v}^{(1)} (\alpha_{0}^{H_{0}}) \right)
\]
\[
+ \left\{ \varepsilon^{2} \left( U_{e,n,v}^{(1)} (\alpha_{0}^{H_{0}}) - U_{e,n,v}^{(1)} (\alpha_{0}) \right) - U_{1} (\alpha_{0}^{H_{0}}, \alpha_{0}) \right\} > \delta' \right)
\]
\[
= 2P \left( \sup_{\alpha \in \Theta_{\alpha}} e^{2} \left( U_{e,n,v}^{(1)} (\alpha) - U_{e,n,v}^{(1)} (\alpha_{0}) \right) - U_{1} (\alpha, \alpha_{0}) > \frac{\delta'_{1}}{3} \right)
\]
\[
+ P \left( \varepsilon^{2} \left( U_{e,n,v}^{(1)} (\tilde{\alpha}_{e,n}^{(1), H_{0}}) - U_{e,n,v}^{(1)} (\alpha_{0}^{H_{0}}) \right) > \frac{\delta'_{1}}{3} \right)
\]
\[
\to 0,
\]

which implies \(\tilde{\alpha}_{e,n}^{(1), H_{0}} \xrightarrow{P} \alpha_{0}^{H_{0}} \text{ under } H_{1}^{(1)}\).
Next, it follows from (29) that under \( H_1^{(2)} \),
\[
\sup_{\beta \in \Theta_{\beta}} \left| \frac{1}{n} \left( U_{e,n,v} (\beta | \tilde{\alpha}_{e,n}^{(1)} - U_{e,n,v} (\beta | \tilde{\alpha}_{e,n}^{(1)}) - U_2 (\beta, \beta_0) \right) \right|^P \to 0
\]
Assumption [C]-\( \text{(ii)} \) implies the following: For any \( \delta_2 > 0 \), there exists \( \delta'_2 > 0 \) such that for any \( \beta \in \Theta_{\beta}^{H_0} \),
\[
|\beta - \beta_0^{H_0}| \geq \delta_2 \implies U_2 (\beta; \beta_0) - U_2 (\beta_0^{H_0}, \beta_0) > \delta'_2.
\]
(78)
By the definition of \( \tilde{\alpha}_{e,n}^{H_0} \) and (78), it holds that
\[
P (|\tilde{\beta}_{e,n}^{H_0} - \beta_0^{H_0}| \geq \delta_2)
\leq P \left( U_2 (\tilde{\beta}_{e,n}^{H_0}, \beta_0) - U_2 (\beta_0^{H_0}, \beta_0) > \delta'_2 \right)
\leq P \left( \left\{ U_2 (\hat{\beta}_{e,n}^{H_0}, \beta_0) - \frac{1}{n} \left( U_{e,n,v} (\hat{\beta}_{e,n}^{H_0} | \tilde{\alpha}_{e,n}^{(1)} - U_{e,n,v} (\beta_0 | \tilde{\alpha}_{e,n}^{(1)}) \right) \right\} + \frac{1}{n} \left( U_{e,n,v} (\hat{\beta}_{e,n}^{H_0} | \tilde{\alpha}_{e,n}^{(1)} - U_{e,n,v} (\beta_0 | \tilde{\alpha}_{e,n}^{(1)}) \right) + \frac{1}{n} \left( U_{e,n,v} (\beta_0 | \tilde{\alpha}_{e,n}^{(1)} - U_{e,n,v} (\beta_0 | \tilde{\alpha}_{e,n}^{(1)}) \right) > \delta'_2 \right)
\leq 2 P \left( \sup_{\beta \in \Theta_{\beta}} \left| \frac{1}{n} \left( U_{e,n,v} (\beta | \tilde{\alpha}_{e,n}^{(1)} - U_{e,n,v} (\beta_0 | \tilde{\alpha}_{e,n}^{(1)}) \right) - U_2 (\beta, \beta_0) \right| > \frac{\delta'_2}{3} \right)
+ P \left( \frac{1}{n} \left( U_{e,n,v} (\hat{\beta}_{e,n}^{H_0} | \tilde{\alpha}_{e,n}^{(1)} - U_{e,n,v} (\beta_0 | \tilde{\alpha}_{e,n}^{(1)}) \right) > \frac{\delta'_2}{3} \right)
\to 0,
\]
which implies \( \tilde{\beta}_{e,n}^{H_0} \xrightarrow{P} \beta_0^{H_0} \) under \( H_1^{(2)} \).
\( \square \)

**Proof of Theorem 4** Under \( H_1^{(1)} \), it holds from the proof of Lemma 1 and Lemma 4 that
\[
\tilde{\alpha}_{e,n}^{(1)} \xrightarrow{P} \alpha_0, \quad \tilde{\alpha}_{e,n}^{(1), H_0} \xrightarrow{P} \alpha_0^{H_0} \neq \alpha_0.
\]
Hence, one has from (12) that
\[
\varepsilon^2 \Lambda_n^{(1)} = \varepsilon^2 \left\{ U_{e,n,v} (\tilde{\alpha}_{e,n}^{(1), H_0}) - U_{e,n,v} (\tilde{\alpha}_{e,n}^{(1)}) \right\}
= \varepsilon^2 \left\{ U_{e,n,v} (\tilde{\alpha}_{e,n}^{(1), H_0}) - U_{e,n,v} (\alpha_0) \right\} - \varepsilon^2 \left\{ U_{e,n,v} (\tilde{\alpha}_{e,n}^{(1)}) - U_{e,n,v} (\alpha_0) \right\}
\xrightarrow{P} U_1 (\alpha_0^{H_0}, \alpha_0) - U_1 (\alpha_0, \alpha_0)
= U_1 (\alpha_0^{H_0}, \alpha_0).
\]
Since it follows from [A3] that $U_1(\alpha H_0, \alpha_0) > 0$ under $H_1^{(1)}$, one has that for any $\delta \in (0, 1)$,$$P(\Lambda_n^{(1)} < \pi_r(\delta)) = P(\varepsilon^2 \Lambda_n^{(1)} < \varepsilon^2 \pi_r(\delta)) \to 0.$$Next, by the proof of Theorem 1 and Lemma 4, we obtain that under $H_1^{(2)},$$\tilde{\beta}_{\varepsilon,n} \xrightarrow{P} \beta_0, \quad \tilde{\beta}_{H_0,\varepsilon,n} \xrightarrow{P} \beta_0^{H_0} \neq \beta_0.$$Therefore, by (29),$$\frac{1}{n} \Lambda_n^{(2)} = \frac{1}{n} \left\{ U_{e,n,v}(\tilde{\beta}_{e,n}^{H_0} | \tilde{\alpha}_{e,n}^{(1)}) - U_{e,n,v}(\tilde{\beta}_{e,n} | \tilde{\alpha}_{e,n}^{(1)}) \right\} = \frac{1}{n} \left\{ U_{e,n,v}(\tilde{\beta}_{e,n}^{H_0} | \tilde{\alpha}_{e,n}^{(1)}) - U_{e,n,v}(\tilde{\beta}_{0} | \tilde{\alpha}_{e,n}^{(1)}) \right\} - \frac{1}{n} \left\{ U_{e,n,v}(\tilde{\beta}_{e,n} | \tilde{\alpha}_{e,n}^{(1)}) - U_{e,n,v}(\tilde{\beta}_{0} | \tilde{\alpha}_{e,n}^{(1)}) \right\} \xrightarrow{P} U_2(\beta_0^{H_0}, \beta_0) - U_2(\beta_0, \beta_0) = U_2(\beta_0^{H_0}, \beta_0).$$It holds from [A3] that $U_2(\beta_0^{H_0}, \beta_0) > 0$ under $H_1^{(2)}$. Consequently, we have that for any $\delta \in (0, 1)$,$$P(\Lambda_n^{(2)} < \chi^2_\delta(\delta)) = P\left(\frac{1}{n} \Lambda_n^{(1)} < \frac{1}{n} \chi^2_\delta(\delta)\right) \to 0.$$This completes the proof. \(\square\)

**Acknowledgements** The authors would like to thank referees for careful reading of the manuscript and valuable comments. This work was partially supported by JST CREST Grant Number JPMJCR2115, MEXT Grant Number JPJ010217 and Cooperative Research Program of the Institute of Statistical Mathematics.

**References**

Azencott R (1982) Formule de taylor stochastique et developpement asymptotique d’integrales de feynmann. Séminaire de Probabilités XVI. 1980/81 Supplément: Géométrie Différentielle Stochastique. Springer, Berlin Heidelberg, pp 237–285

Freidlin MI, Wentzell AD (1998) Random perturbations of dynamical systems. Springer, New York

Genon-Catalot V (1990) Maximum contrast estimation for diffusion processes from discrete observations. Statistics 21(1):99–116

Genon-Catalot V, Jacob J (1993) On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. Annales de l’IHP Probabilités et statistiques 29(1):119–151

Gloter A, Sørensen M (2009) Estimation for stochastic differential equations with a small diffusion coefficient. Stoch Process Appl 119(3):679–699

Guy R, Laréd C, Vergu E (2014) Parametric inference for discretely observed multidimensional diffusions with small diffusion coefficient. Stoch Process Appl 124(1):51–80
Guy R, Larédo C, Vergu E (2015) Approximation of epidemic models by diffusion processes and their statistical inference. J Math Biol 70(3):621–646
Hall P, Heyde CC (1980) Martingale limit theory and its application. Academic press, New York
Kaino Y, Uchida M (2018) Hybrid estimators for small diffusion processes based on reduced data. Metrika 81(7):745–773
Kawai T, Uchida M (2022) Adaptive testing method for ergodic diffusion processes based on high frequency data. J Stat Plan Inference 217:241–278
Kessler M (1995) Estimation des paramètres d’une diffusion par des contrastes corrigés. Comptes rendus de l’Académie des sciences. Série I, Mathématique 320(3):359–362
Kutoyants YA (1984) Parameter estimation for stochastic processes. Heldermann
Kutoyants YA (1994) Identification of dynamical systems with small noise. Kluwer, Dordrecht
Laredo CF (1990) A sufficient condition for asymptotic sufficiency of incomplete observations of a diffusion process. Ann Stat 18(3):1158–1171
Nakakita SH, Uchida M (2019) Adaptive test for ergodic diffusions plus noise. J Stat Plan Inference 203:131–150
Nomura R, Uchida M (2016) Adaptive Bayes estimators and hybrid estimators for small diffusion processes based on sampled data. J Jpn Stat Soc 46(2):129–154
Prakasa Rao BLS (1983) Asymptotic theory for non-linear least squares estimator for diffusion processes. Stat: J Theor Appl Stat 14(2):195–209
Prakasa Rao BLS (1988) Statistical inference from sampled data for stochastic processes. Contemp Math 80:249–284
Sørensen M, Uchida M (2003) Small-diffusion asymptotics for discretely sampled stochastic differential equations. Bernoulli 9(6):1
Tsybakov AB (2009) Introduction to nonparametric estimation. Springer
Uchida M (2003) Estimation for dynamical systems with small noise from discrete observations. J Jpn Stat Soc 33(2):157–167
Uchida M (2004) Estimation for discretely observed small diffusions based on approximate martingale estimating functions. Scand J Stat 31(4):553–566
Uchida M (2008) Approximate martingale estimating functions for stochastic differential equations with small noises. Stoch Process Appl 118(9):1706–1721
Uchida M, Yoshiida N (2004) Asymptotic expansion for small diffusions applied to option pricing. Stat Infer Stoch Process 7(3):189–223
Uchida M, Yoshiida N (2012) Adaptive estimation of an ergodic diffusion process based on sampled data. Stoch Process Appl 122(8):2885–2924
van der Vaart A (1998) Asymptotic Statistics. Cambridge University Press, Cambridge
Yoshida N (1992) Estimation for diffusion processes from discrete observation. J Multivar Anal 41(2):220–242
Yoshida N (1992) Asymptotic expansions of maximum likelihood estimators for small diffusions via the theory of Malliavin–Watanabe. Probab Theory Relat Fields 92(3):275–311
Yoshida N (1992) Asymptotic expansion for statistics related to small diffusions. J Jpn Stat Soc 22(2):139–159
Yoshida N (2003) Conditional expansions and their applications. Stoch Process Appl 107(1):53–81

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.