Ann Inst Stat Math (2013) 65:213–236
DOI 10.1007/s10463-012-0366-9

Minimum density power divergence estimator for diffusion processes

Sangyeol Lee · Junmo Song

Received: 13 March 2007 / Revised: 20 April 2012 / Published online: 21 July 2012

Abstract In this paper, we consider the robust estimation for a certain class of diffusion processes including the Ornstein–Uhlenbeck process based on discrete observations. As a robust estimator, we consider the minimum density power divergence estimator (MDPDE) proposed by Basu et al. (Biometrika 85:549–559, 1998). It is shown that the MDPDE is consistent and asymptotically normal. A simulation study demonstrates the strong robustness of the MDPDE.

Keywords Diffusion processes · The Ornstein–Uhlenbeck process · Minimum density power divergence estimator · Discretely observed sample · Robustness

1 Introduction

Let us consider the diffusion process:

$$\begin{cases} \mathrm{d}X_t = a(X_t, \theta)\mathrm{d}t + \sigma \mathrm{d}W_t, \\ X_0 = x_0, \end{cases} \tag{1}$$

where $(\theta, \sigma) \in \Theta$, a convex compact subset of $\mathbb{R}^p \times \mathbb{R}^+$, a is a known real valued function defined on $\mathbb{R} \times \mathbb{R}^p$, and W is a 1-dimensional standard Wiener process. The diffusion process has long been popular in analyzing random phenomena in finance,

S. Lee (✉)
Department of Statistics, Seoul National University,
Seoul 151-742, Korea
e-mail: jpgrslee@yahoo.com; sylee@stats.snu.ac.kr

J. Song
Department of Computer Science and Statistics,
Jeju National University, Jeju 690-756, Korea

engineering, physical and medical sciences. Particularly, the diffusion process given by (1) has wide applications and includes the Ornstein–Uhlenbeck process as a special case. Statistical inference for this process has been studied by many authors, for instance, Florens-Zmirou (1989), and the asymptotic properties for various parameter estimators are well summarized in Prakasa Rao (1999, pp 143–144 and 153–158), Yoshida (1992), Kessler (1997), Ait-Sahalia (2002) and many other researchers studied the estimation problem for a more general class of diffusion processes. In those articles, the estimation procedure is conducted based on discretely observed sample. Statistical inference for continuous samples can be found in Kutoyants (2004).

In this paper, we consider a robust estimation of (θ, σ) in the model (1) based on discretized observations. In the statistical literature, it is well known that the estimators based on Gaussian likelihood are severely influenced by outliers or extreme values. Thus, it can be guessed that the similar situations happen in the estimation procedure, based on Gaussian approximation method, for diffusion processes [for e.g., the Euler estimator in Prakasa Rao (1999, p 155) and Kessler (1997) estimator]. In fact, in our simulation study, the Euler estimator is observed to be severely damaged by outliers (cf. Tables 2, 3).

In order to construct a robust estimator for process (1), we adopt the idea of Basu et al. (1998) (BHHJ for abbreviation) which introduces a robust estimation procedure to minimize a density-based divergence measures, called the density power divergences:

$$d_\alpha(g, f) := \begin{cases} \int \left\{ f^{1+\alpha}(z) - (1 + \frac{1}{\alpha})\, g(z)\, f^\alpha(z) + \frac{1}{\alpha}\, g^{1+\alpha}(z) \right\} \mathrm{d}z, & \alpha > 0, \\ \int g(z)\,(\log g(z) - \log f(z))\,\mathrm{d}z, & \alpha = 0, \end{cases} \tag{2}$$

where f and g are density functions.

For a family of parametric distributions $\{F_\theta : \theta \in \Theta \subset \mathbb{R}^m\}$ possessing densities $\{f_\theta\}$ and for a distribution G with density g, they defined the minimum density power divergence functional $T_\alpha(\cdot)$ by

$$d_\alpha\left(g, f_{T_\alpha(g)}\right) = \min_{\theta\in\Theta} d_\alpha(g, f_\theta).$$

Note that if G belongs to $\{F_\theta\}$, $T_\alpha(g) = \theta$ for some $\theta \in \Theta$. In this case, given random sample $X_1, \ldots, X_n$ with unknown density g, the minimum density power divergence estimator (MDPDE) is defined by

$$\hat{\theta}_{\alpha,n} = \arg\min_{\theta\in\Theta} \frac{1}{n} \sum_{t=1}^{n} V_{\alpha,n}(\theta; X_t),$$

where

$$V_\alpha(\theta; X_t) := \begin{cases} \int f_\theta^{1+\alpha}(z)\mathrm{d}z - \left(1 + \frac{1}{\alpha}\right) f_\theta^\alpha(X_t), & \alpha > 0, \\ \log f_\theta(X_t), & \alpha = 0. \end{cases}$$

BHHJ showed that $\hat{\theta}_{\alpha,n}$ is weakly consistent for $T_\alpha(g)$ and asymptotically normal, and demonstrated that the estimator has strong robust properties against outliers and the misspecification of underlying models with little loss in asymptotic efficiency relative to the maximum likelihood estimator.

This approach can be extended to regression models. Let $\{f_\theta(y|x)\}$ be a parametric family of regression models indexed by the parameter $\theta \in \Theta$ and let $g(y|x)$ be the true density for Y given $X = x$. Substituting f and g in (2) by $f_\theta(\cdot|x)$ and $g(\cdot|x)$ respectively, a family of the x-conditional versions of density power divergences is obtained as follows:

$$\arg\min_\theta \begin{cases} n^{-1}\sum_{t=1}^n \int f_\theta^{1+\alpha}(y|x_t)\,\mathrm{d}y - \left(1+\frac{1}{\alpha}\right) n^{-1}\sum_{t=1}^n f_\theta^\alpha(Y_t|x_t), & \alpha > 0, \\ n^{-1}\sum_{t=1}^n -\log f_\theta(Y_t|x_t), & \alpha = 0. \end{cases} \tag{3}$$

This idea will be adopted later for our own purpose.

Compared to other existing density-based divergence methods, such as Beran (1977), Tamura and Boos (1986) and Simpson (1987), which use the Hellinger distance, and Basu and Lindsay (1994) and Cao et al. (1995), this method is known to have merit of not requiring any smoothing methods. In this case, one can avoid drawbacks and difficulties like the selection of bandwidth that necessarily follow from the kernel smoothing method.

The organization of this paper is as follows. In Sect. 2, we construct the robust estimator using (3) and address the asymptotic properties of the proposed estimator. In Sect. 3, we perform a simulation study and compare the proposed estimator with the Euler estimator. The proofs for the results in Sect. 2 are provided in Sect. 4. Finally, some auxiliary lemmas are presented in Sect. 5.

2 Main results

Let $(\theta_0, \sigma_0) \in \Theta$ be the true parameter for the diffusion process in (1). Suppose that $X_{t_i^n}$, $1 \le i \le n$, are observed, where $t_i^n = ih_n$, $h_n \to 0$, and $nh_n \to \infty$. Applying the Euler's approximation to (1), we have that

$$X_{t_i^n} = X_{t_{i-1}^n} + a(X_{t_{i-1}^n}, \theta_0)h_n + \sigma_0 Z_{n,i}\sqrt{h_n} + \Delta_{n,i},$$

where

$$Z_{n,i} = \frac{1}{\sqrt{h_n}}\left(W_{t_i^n} - W_{t_{i-1}^n}\right) \quad \text{and} \quad \Delta_{n,i} = \int_{t_{i-1}}^{t_i} \left\{a(X_s, \theta_0) - a\left(X_{t_{i-1}^n}, \theta_0\right)\right\} \mathrm{d}s.$$

Let $\mathscr{G}_i^n$ denote the sigma field generated by $\{W_s : s \le t_i^n\}$. If ignoring $\Delta_{n,i}$'s (actually, one can check that $\max_{i\le n}|\Delta_{n,i}| = o_p(h_n)$. cf. Lemma 2), by noticing that $Z_{n,1}, \ldots, Z_{n,n}$ are iid $N(0, 1)$, we can see that for large n, $X_{t_i^n}|\mathscr{G}$ behave like independent r.v.'s following $N(X_{t_{i-1}^n} + a(X_{t_{i-1}^n}, \theta_0),\ \sigma_0^2 h_n)$. Hence, viewing the observations as regression data pairs $\{(X_{t_{i-1}^n}, X_{t_i^n})\}_{i=1}^n$ and applying (3) to them (in this case the family of parametric distributions is the normal distributions), we can define the MDPDE as

$$(\hat{\theta}_n^\alpha, \hat{\sigma}_n^\alpha) = \arg\min_{\theta,\sigma} \frac{1}{n} \sum_{i=1}^n V_{n,i}^\alpha(\theta, \sigma),$$

where

$$V_{n,i}^\alpha(\theta,\sigma) = \begin{cases} \left(\frac{1}{\sigma}\right)^\alpha \Big[(1+\alpha)^{-\frac{1}{2}} - \left(1+\frac{1}{\alpha}\right) \\ \quad \times \exp\left\{-\frac{\alpha}{2}\left(X_{t_i^n} - X_{t_{i-1}^n} - a\left(X_{t_{i-1}^n}, \theta\right) h_n\right)^2 / \sigma^2 h_n\right\}\Big], & \alpha > 0, \\ \left(X_{t_i^n} - X_{t_{i-1}^n} - a\left(X_{t_{i-1}^n}, \theta\right) h_n\right)^2 / \sigma^2 h_n + \log \sigma^2, & \alpha = 0. \end{cases}$$

Remark 1 The MDPDE with $\alpha = 0$ coincides with the Euler estimator in Prakasa Rao (1999, p 155). Also, $\hat{\theta}_n^0$ is the same as the least squares estimator $\hat{\theta}_{\mathrm{LSQ}}$ [see also Florens-Zmirou 1989] and $\hat{\theta}_{\mathrm{LF}}$ that maximize the discrete approximate likelihood functions, respectively:

$$\hat{\theta}_{\mathrm{LSQ}} = \arg\min_\theta \sum_{i=1}^n \left(X_{t_i^n} - X_{t_{i-1}^n} - a\left(X_{t_{i-1}^n}, \theta\right) h_n\right)^2,$$

and

$$\hat{\theta}_{\mathrm{LF}} = \arg\max_\theta \sum_{i=1}^n \left\{a\left(X_{t_{i-1}^n}, \theta\right)\left(X_{t_i^n} - X_{t_{i-1}^n}\right)^2 - \frac{1}{2} a\left(X_{t_{i-1}^n}, \theta\right)^2 h_n\right\}.$$

Below we establish the consistency and asymptotic normality of the MDPDE. For this task, we set

$$\mathscr{P} = \left\{f(x,\theta) : |f| \le C(1+|x|)^C \text{ for some } C\right\},$$

where C does not depend on θ, and assume the conditions as follows:

(A1) There exists a constant C_1 such that for any x, y, $|a(x,\theta_0) - a(y,\theta_0)| \le C_1|x-y|$.
(A2) The process X from (1) is ergodic for (θ_0, σ_0) with its invariant measure μ_0 such that $\int x^k \mathrm{d}\mu_0(x) < \infty$ for all $k \ge 0$.
(A3) $\sup_t \mathbb{E}\{|X_t|^k\} < \infty$ for all $k \ge 0$.
(A4) The function a is continuously differentiable with respect to x for all θ and the derivatives belong to $\mathscr{P}$.

(A5) The function a and all its x-derivatives are three times differentiable with respect to θ for all x. Moreover, these derivatives up to the third order with respect to θ belong to $\mathscr{P}$.

(A6) If $a(x,\theta) = a(x,\theta_0)$ for μ_0 a.s. all x, then $\theta = \theta_0$.

(A7) $S := \int \partial_\theta a(x,\theta_0)\, \partial_{\theta^T} a(x,\theta_0) d\mu_0(x)$ is positive definite, where $\partial_\theta a = \partial a/\partial\theta$.

Here are the main results of this paper.

Theorem 1 *Assume that* (A1)–(A6) *hold. For each* $\alpha \geq 0$, *if* $h_n \to 0$, $nh_n \to \infty$ *and* $nh_n^q \to 0$ *for some* $q > 1$, *then*

$$\left(\hat{\theta}_n^\alpha,\ \hat{\sigma}_n^\alpha\right) \to (\theta_0,\ \sigma_0) \quad \textit{in probability.}$$

Remark 2 The condition that $nh_n^q \to 0$ for some $q > 1$ is not necessary for the case of $\alpha = 0$. However, in the other cases, this condition is essential to obtain the weak consistency result (cf. Lemma 2).

Theorem 2 *Assume that* (A1)–(A7) *hold and* (θ_0, σ_0) *is in the interior of* Θ. *For each* $\alpha \geq 0$, *if* $h_n \to 0$, $nh_n \to \infty$ *and* $nh_n^2 \to 0$, *then*

$$\begin{pmatrix} \sqrt{nh_n}(\hat{\theta}_n^\alpha - \theta_0) \\ \sqrt{n}(\hat{\sigma}_n^\alpha - \sigma_0) \end{pmatrix} \longrightarrow N_{p+1}(0, \Sigma_\alpha) \quad \textit{in distribution,}$$

where

$$\Sigma_\alpha = \sigma_0^2 \begin{pmatrix} \left(\frac{1+\alpha}{\sqrt{1+2\alpha}}\right)^3 S^{-1} & 0 \\ 0 & \frac{(1+\alpha)^3}{(2+\alpha^2)^2}\left(2\frac{(1+\alpha)^2(1+2\alpha^2)}{(1+2\alpha)^2\sqrt{1+2\alpha}} - \frac{\alpha^2}{1+\alpha}\right) \end{pmatrix}.$$

Remark 3 In actual practice, one may raise a question of how to choose an optimal α. In the situation of no outliers, one may select α to minimize the asymptotic variance in Theorem 2 (i.e., the case of $\alpha = 0$). However, with their existence, the outliers will damage this procedure, so an optimal α is very hard to choose. Conventionally, taking account of this difficulty, one employs a fixed α rather than seek for a suitable α, which ranges in [0.15, 0.5] since too a large α, which would have strong robust properties, may result in a big loss in efficiency when the portion of outliers is not very large as speculated. We will see this phenomenon from the simulation result presented in the next section.

So far, we have considered the case that the diffusion coefficient is a constant. However, the MDPDE can be extended to the following diffusion process:

$$\mathrm{d}X_t = a(X_t,\theta)\mathrm{d}t + \sigma\, b(X_t)\mathrm{d}W_t. \tag{4}$$

Since by transforming X_t into $Y_t = G(X_t)$, where G satisfies the relationship $\partial_x G(x) = 1/b(x)$, and using Itô's formula, we obtain

$$dY_t = \mu(Y_t,\theta,\sigma)\mathrm{d}t + \sigma \mathrm{d}W_t, \tag{5}$$

where

$$\mu(y,\theta,\sigma) = \frac{a\left(G^{-1}(y),\theta\right)}{b\left(G^{-1}(y)\right)} - \frac{\sigma^2}{2}\,\partial_x b\left(G^{-1}(y),\sigma\right),$$

and subsequently, the estimation procedure for the process in (4) is reduced to that for the diffusion process in (1). For a more general class of diffusion process, we may define the MDPDE using the Euler approximation and the contrast functions in (3). However, there are some technical difficulties in obtaining the same asymptotic properties of the estimator. We leave this task as a future study.

3 Simulation study

In this section, we compare the performance of the MDPDE with $\alpha \in \{0.05, 0.1, \ldots, 0.95\}$ and the Euler estimator (EE) for the Ornstein–Uhlenbeck process:

$$\begin{cases} \mathrm{d}X_t = -\theta X_t \mathrm{d}t + \sigma \mathrm{d}W_t \\ X_0 = 0. \end{cases} \tag{6}$$

In our simulation, the case $(\theta_0, \sigma_0) = (1, 1)$ is considered, and the sample $\{X_{o,t_i^n}\}_{i=1}^n$ is obtained discretely with sampling interval $h_n = n^{-0.55}$. The comparison is based on the numbers defined by

$$d = d_\alpha := \frac{1}{r}\sum_{i=1}^{r}\left|\left(\hat{\theta}_{n,i}^{\alpha} - \theta_0\right)^2 + \left(\hat{\sigma}_{n,i}^{\alpha} - \sigma_0\right)^2\right|^{\frac{1}{2}}, \quad d_R := \frac{d}{d \text{ of EE}},$$

where r is the number of repetitions. In fact, d is the average distance between the true parameter and its estimates, so the smaller d (or d_R) indicates that the estimator is more efficient.

First, we handle the case that the observation is not contaminated by outliers. Based on 1000 repetitions, the mean, standard deviation, d and d_R are calculated for $n = 500, 800, 1000$. Figure 1 plots the calculated d_α's, where d_0 is the one for the EE. The results presented in Table 1 and Fig. 1 show that the EE outperforms the MDPDE,

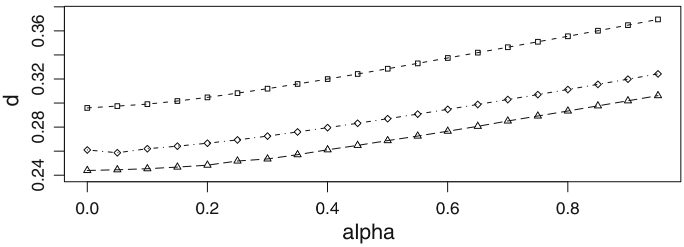

Fig. 1 d without outliers. *Square*, *diamond* and *triangle* represent for the case of n = 500, 800, 1000, respectively

Table 1 Mean (SD) and $d(d_R)$ without outliers

n	EE	MDPDE					
		0.05	0.15	0.3	0.5	0.75	0.95
500							
θ	1.091 (0.146)	1.090 (0.148)	1.089 (0.153)	1.087 (0.164)	1.084 (0.183)	1.082 (0.210)	1.080 (0.233)
σ	0.983 (0.0010)	0.983 (0.0010)	0.983 (0.0010)	0.983 (0.0011)	0.983 (0.0013)	0.983 (0.0016)	0.983 (0.0017)
$d(d_R)$	0.296 (1.000)	0.297 (1.005)	0.302 (1.019)	0.312 (1.054)	0.329 (1.110)	0.351 (1.186)	0.370 (1.249)
800							
θ	1.068 (0.110)	1.070 (0.109)	1.067 (0.113)	1.066 (0.121)	1.066 (0.134)	1.065 (0.154)	1.064 (0.172)
σ	0.987 (0.0006)	0.987 (0.0006)	0.987 (0.0007)	0.987 (0.0007)	0.987 (0.0008)	0.987 (0.0010)	0.987 (0.0011)
$d(d_R)$	0.261 (1.000)	0.259 (0.991)	0.264 (1.012)	0.273 (1.044)	0.287 (1.100)	0.307 (1.176)	0.324 (1.242)
1000							
θ	1.080 (0.098)	1.080 (0.098)	1.081 (0.099)	1.084 (0.104)	1.084 (0.115)	1.087 (0.132)	1.089 (0.146)
σ	0.988 (0.0005)	0.988 (0.0005)	0.988 (0.0005)	0.988 (0.0005)	0.988 (0.0006)	0.987 (0.0007)	0.987 (0.0008)
$d(d_R)$	0.244 (1.000)	0.245 (1.002)	0.247 (1.012)	0.254 (1.039)	0.269 (1.102)	0.289 (1.186)	0.306 (1.255)

and the MDPDE with α close to 0 performs similarly to the EE. It is also seen that the performance of the MDPDE with α not close to 0, say $\alpha = 0.15$–0.5, is not very poor.

Tables 2 and 3 and Figs. 2, 3, 4, 5, 6, and 7 summarize the results for the case that outliers are involved in the data. Here, we consider the situation that the sample $\{X_{o,t_i^n}\}_{i=1}^n$ from (6) is contaminated by the outliers $\{X_{c,t_i^n}\} \sim$ iid $N(0, \sigma_V^2)$ and the observed r.v.'s follow the scheme $X_{t_i^n} = (1 - p_i)\, X_{o,t_i^n} + p_i\, X_{c,t_i^n}$, where p_i are iid Bernoulli r.v.'s with success probability p. It is assumed that $\{p_i\}$, $\{X_{o,t_i^n}\}$ and $\{X_{c,t_i^n}\}$ are all independent. The mean, standard deviation, d and d_R based on $\{X_{t_i^n}\}$ are calculated out of 1000 repetitions for $n = 1000$, $p = 0.05$ and 0.1. From Tables 2 and 3, we can see that the d_R tends to get smaller as either σ_V^2 or p increases except for some cases such as $\alpha = 0.05, 0.15$ and $p = 0.1$. The bold phased figures denote the α's that give a minimal d_R's. Figures 2 and 5 show the plots of d_R when observations are contaminated by 5 and 10% outliers, respectively. It can be seen that the MDPDE's with the α lying in 0.15–0.3 produce very small d_R's. Figures 3 and 6 show the plots of the EE's and the MDPDE's producing the smallest d (optimal MDPDE's), and Figs. 4 and 6 are the corresponding histograms. The figures show that the EE's scatter widely whereas the optimal MDPDE's lie near the true parameter. From these results, we can conclude that the MDPDE possesses much more robust properties than the EE. It can be seen that the α's yielding minimal d varies with the cases. This indicates that choosing an optimal α is not an easy task in actual usage. Conventionally, $\alpha \in [0.1, 0.2]$ is recommended since the MDPDE with the α still keeps the efficiency when there are no outliers and are robust against outliers. The same can be applied to our case, but our simulation study suggests to us that a broader range of α's, say, in $[0.15, 0.5]$ may be employed in construction of the MDPDE.

4 Proofs

We will provide the proof for the case of $\alpha > 0$ since the proofs of the case of $\alpha = 0$ are similar to that of $\alpha > 0$. In what follows, we denote

$$a_i(\theta) = a(X_{t_i^n}, \theta), \quad Z_{n,i} = Z_i, \quad \eta = (\theta, \sigma), \quad \eta_0 = (\theta_0, \sigma_0).$$

Moreover, $C > 0$ denotes a universal constant.

Proof of Theorem 1 Note that

$$U(\sigma, \sigma_0) := \left(\frac{1}{\sigma}\right)^{\alpha} \left\{ \frac{1}{\sqrt{1+\alpha}} - \left(1 + \frac{1}{\alpha}\right)\left(1 + \alpha \frac{\sigma_0^2}{\sigma^2}\right)^{-\frac{1}{2}} \right\}, \quad \sigma > 0,$$

has a minimal value at $\sigma = \sigma_0$. Similarly to the proof of Theorem 1 of Kessler (1997), Theorem 1 is proved if we verify that

Table 2 Mean (SD), d and d_R with 5 % outliers

σ_V^2	EE	MDPDE								
		0.05	0.15	0.30	0.40	0.45	0.50	0.55	0.75	0.95
1										
θ	7.646 (4.318)	4.185 (2.404)	1.474 (0.244)	1.221 (0.167)	1.179 (0.161)	1.167 (0.161)	1.158 (0.162)	1.150 (0.163)	1.131 (0.171)	1.121 (0.181)
σ	2.601 (0.0869)	1.909 (0.0681)	1.152 (0.0028)	1.062 (0.0010)	1.050 (0.0009)	1.048 (0.0009)	1.047 (0.0009)	1.047 (0.0010)	1.050 (0.0011)	1.057 (0.0012)
d	6.8522	3.3308	0.5612	0.3566	0.3362	0.3322	0.3303	**0.3296**	0.3346	0.3444
d_R	1.0000	0.4861	0.0819	0.0520	0.0491	0.0485	0.0482	**0.0481**	0.0488	0.0503
2										
θ	10.920 (7.772)	5.407 (5.255)	1.398 (0.207)	1.197 (0.154)	1.163 (0.152)	1.153 (0.153)	1.146 (0.155)	1.140 (0.157)	1.125 (0.169)	1.118 (0.182)
σ	3.198 (0.1061)	2.155 (0.1284)	1.124 (0.0024)	1.053 (0.0010)	1.045 (0.0009)	1.044 (0.0009)	1.044 (0.0009)	1.045 (0.0009)	1.051 (0.0010)	1.059 (0.0011)
d	10.1768	4.5765	0.4914	0.3392	0.3271	**0.3259**	0.3263	0.3277	0.3390	0.3535
d_R	1.0000	0.4497	0.0483	0.0333	0.0321	**0.0320**	0.0321	0.0322	0.0333	0.0347
3										
θ	13.837 (11.793)	6.618 (11.058)	1.368 (0.185)	1.199 (0.142)	1.170 (0.139)	1.161 (0.140)	1.155 (0.141)	1.149 (0.143)	1.135 (0.151)	1.127 (0.162)
σ	3.668 (0.1453)	2.330 (0.2418)	1.103 (0.0018)	1.045 (0.0009)	1.040 (0.0009)	1.040 (0.0009)	1.041 (0.0009)	1.042 (0.0009)	1.049 (0.0010)	1.058 (0.0012)
d	13.1277	5.7954	0.4581	0.3333	0.3216	0.3199	**0.3194**	0.3200	0.3273	0.3386
d_R	1.0000	0.4415	0.0349	0.0254	0.0245	0.0244	**0.0243**	0.0244	0.0249	0.0258

Table 3 Mean (SD), d and d_R with 10 % outliers

σ_V^2		EE	MDPDE								
			0.05	0.15	0.30	0.40	0.50	0.60	0.70	0.80	0.95
1											
	θ	13.166 (7.765)	10.361 (7.014)	3.113 (2.128)	1.486 (0.273)	1.352 (0.227)	1.289 (0.211)	1.253 (0.205)	1.229 (0.204)	1.212 (0.204)	1.194 (0.207)
	σ	3.410 (0.0872)	3.011 (0.0849)	1.639 (0.0585)	1.177 (0.0025)	1.139 (0.0017)	1.128 (0.0015)	1.126 (0.0015)	1.130 (0.0015)	1.135 (0.0016)	1.145 (0.0017)
	d	12.4176	9.5922	2.2330	0.5901	0.4799	0.4381	0.4203	0.4134	**0.4119**	0.4153
	d_R	1.0000	0.7725	0.1798	0.0475	0.0386	0.0353	0.0338	0.0333	**0.0332**	0.0334
2											
	θ	18.099 (11.809)	15.020 (13.104)	2.782 (3.406)	1.383 (0.247)	1.288 (0.216)	1.242 (0.206)	1.216 (0.204)	1.199 (0.204)	1.188 (0.205)	1.175 (0.210)
	σ	4.252 (0.1226)	3.735 (0.1266)	1.536 (0.0920)	1.145 (0.0023)	1.121 (0.0017)	1.117 (0.0016)	1.120 (0.0016)	1.126 (0.0016)	1.134 (0.0017)	1.147 (0.0018)
	d	17.4199	14.3031	1.8940	0.5049	0.4346	0.4097	0.4010	**0.3994**	0.4015	0.4081
	d_R	1.0000	0.8211	0.1087	0.0290	0.0249	0.0235	0.0230	**0.0229**	0.0230	0.0234
3											
	θ	21.574 (12.397)	18.518 (15.071)	2.280 (3.180)	1.308 (0.189)	1.230 (0.171)	1.193 (0.166)	1.170 (0.166)	1.156 (0.167)	1.145 (0.170)	1.135 (0.175)
	σ	4.859 (0.1392)	4.234 (0.1500)	1.404 (0.0659)	1.126 (0.0018)	1.109 (0.0014)	1.109 (0.0013)	1.114 (0.0014)	1.122 (0.0014)	1.132 (0.0015)	1.146 (0.0016)
	d	20.9461	17.8298	1.3811	0.4427	0.3924	0.3764	**0.3725**	0.3742	0.3786	0.3878
	d_R	1.0000	0.8512	0.0659	0.0211	0.0187	0.0180	**0.0178**	0.0179	0.0181	0.0185

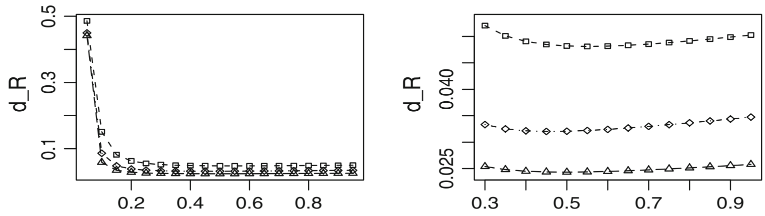

Fig. 2 d_R versus α with 5 % outliers. *Square*, *diamond* and *triangle* represent for the case of $\sigma_V^2 = 1, 2, 3$, respectively

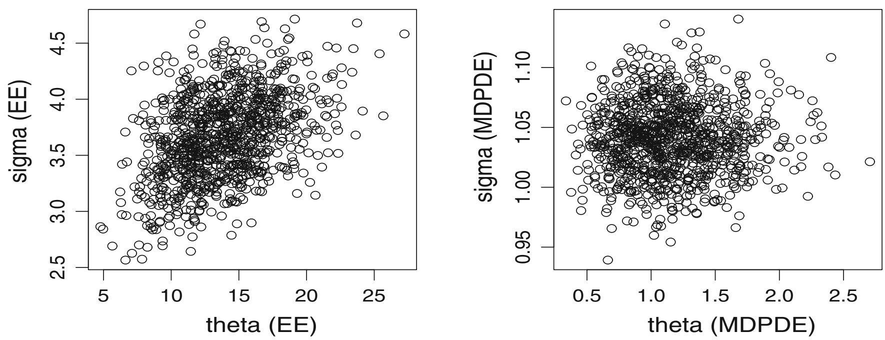

Fig. 3 Plotting $(\hat{\theta}, \hat{\sigma})$ of the EE and the optimal MDPDE with 5 % outliers when $\sigma_V^2 = 3$

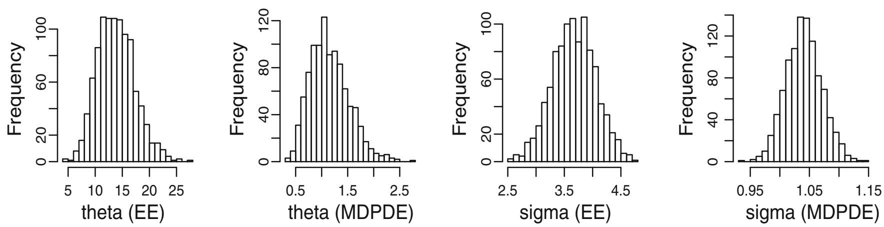

Fig. 4 Histogram of the EE and the optimal MDPDE with 5 % outliers when $\sigma_V^2 = 3$

(i) $\frac{1}{n}\sum_{i=1}^{n} V_{n,i}^{\alpha}(\theta,\sigma) \xrightarrow{P} U(\sigma,\sigma_0)$ uniformlyin η.

(ii) $\frac{1}{nh_n}\sum_{i=1}^{n} V_{n,i}^{\alpha}(\theta,\sigma) - \frac{1}{nh_n}\sum_{i=1}^{n} V_{n,i}^{\alpha}(\theta_0,\sigma)$

$$\xrightarrow{P} \frac{1}{2}(1+\alpha)\left(\frac{1}{\sigma}\right)^{\alpha+2}\left(1+\alpha\frac{\sigma_0{}^2}{\sigma^2}\right)^{-\frac{3}{2}} \int \{a(x,\theta)-a(x,\theta_0)\}^2 \,\mathrm{d}\mu_0(x)$$

uniformly in η.

Put

$$K_i(\theta,\sigma) = K_{n,i}(\theta,\sigma) := \alpha\frac{\sigma_0}{\sigma^2} Z_i A_i(\theta)\sqrt{h_n} + \frac{\alpha}{2}\frac{A_i(\theta)^2}{\sigma^2} h_n + \alpha\frac{\sigma_0}{\sigma^2}\frac{Z_i\Delta_i}{\sqrt{h_n}} + \alpha\frac{A_i(\theta)\Delta_i}{\sigma^2} + \frac{\alpha}{2}\frac{\Delta_i^2}{\sigma^2 h_n},$$

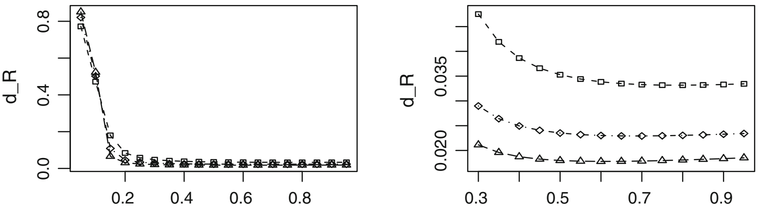

Fig. 5 d_R versus α with 10 % outliers. *Square*, *diamond* and *triangle* represent for the case of $\sigma_V^2 = 1, 2, 3$, respectively

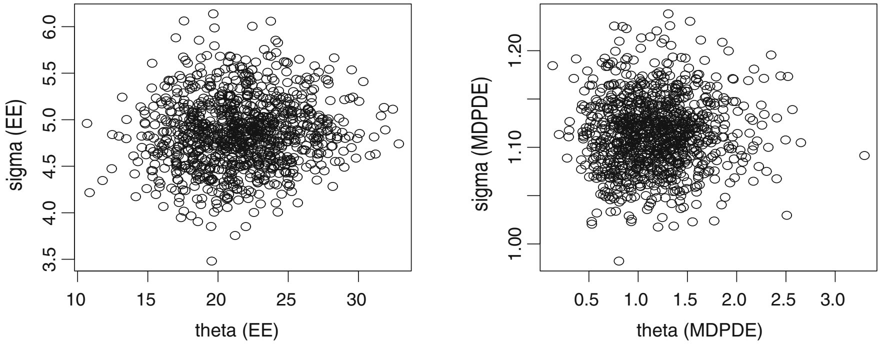

Fig. 6 Plotting $(\hat{\theta}, \hat{\sigma})$ of the EE and the optimal MDPDE with 10 % outliers when $\sigma_V^2 = 3$

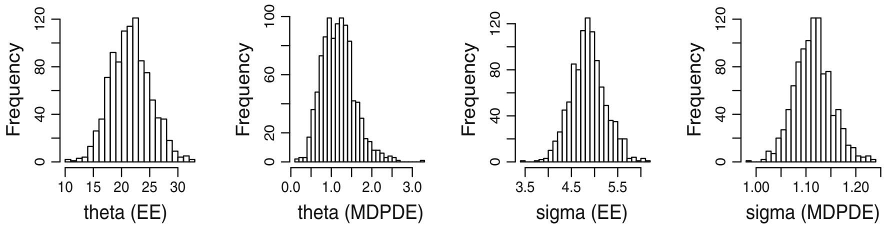

Fig. 7 Histogram of the EE and the optimal MDPDE with 10 % outliers when $\sigma_V^2 = 3$

where $A_i(\theta) = A_{n,i}(\theta) := a_{i-1}(\theta_0) - a_{i-1}(\theta)$. In view of Lemma 2 in Sect. 5, we have

$$\sup_{\eta} \max_{1 \le i \le n} |K_i(\theta, \sigma)| = o_p(h_n^{\gamma}) \qquad \text{for} \quad 0 \le \gamma < \frac{1}{2}.$$

Thus, we have that

$$\sup_{\eta} \Bigg| \left(1 + \frac{1}{\alpha}\right) \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{\sigma}\right)^{\alpha} \\ \times \left\{ \exp\left(-\frac{\alpha}{2} \frac{\left(X_{t_i^n} - X_{t_{i-1}^n} - a_{i-1}(\theta) h_n\right)^2}{\sigma^2 h_n} \right) - \exp\left(-\frac{\alpha}{2} \frac{\sigma_0^2}{\sigma^2} {Z_i}^2 \right) \right\} \Bigg|$$

$$
\begin{aligned}
&= \sup_{\eta} \left| \left(1+\frac{1}{\alpha}\right) \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{\sigma}\right)^{\alpha} \mathrm{e}^{-\frac{\alpha}{2}\frac{\sigma_0^2}{\sigma^2} Z_i{}^2} \left(\mathrm{e}^{-K_i(\theta,\sigma)} - 1\right) \right| \\
&\le \left(1+\frac{1}{\alpha}\right) \sup_{\eta} \left(\frac{1}{\sigma}\right)^{\alpha} \max_{1 \le i \le n} \left| \mathrm{e}^{-K_i(\theta,\sigma)} - 1 \right| \\
&\le C_\alpha \left\{ \exp\left(\sup_{\eta} \max_{1\le i\le n} |K_i(\theta,\sigma)| \right) - 1 \right\} = o_P(1).
\end{aligned}
$$

Using this and Lemma 4, we have that uniformly in η,

$$
\begin{aligned}
&\left(1+\frac{1}{\alpha}\right) \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{\sigma}\right)^{\alpha} \exp\left(-\frac{\alpha}{2} \frac{\left(X_{t_i^n} - X_{t_{i-1}^n} - a_{i-1}(\theta) h_n\right)^2}{\sigma^2 h_n} \right) \\
&\xrightarrow{P_0} \left(1+\frac{1}{\alpha}\right) \left(\frac{1}{\sigma}\right)^{\alpha} \left(1 + \alpha \frac{\sigma_0^2}{\sigma^2}\right)^{-\frac{1}{2}},
\end{aligned}
$$

which establish (i).

Next, we verify (ii). Using Taylor's theorem, we can write that

$$
\begin{aligned}
&\mathrm{e}^{-K_i(\theta_0,\sigma)} - \mathrm{e}^{-K_i(\theta,\sigma)} \\
&\quad = \alpha \frac{\sigma_0}{\sigma^2} Z_i A_i(\theta) \sqrt{h_n} + \frac{\alpha}{2} \frac{A_i(\theta)^2}{\sigma^2} h_n + \alpha \frac{A_i(\theta)\Delta_i}{\sigma^2} \\
&\qquad - \frac{1}{2} \left\{ \alpha^2 \frac{\sigma_0^2}{\sigma^4} Z_i^2 A_i(\theta)^2 h_n + K_i(\theta,\sigma)^2 - \alpha^2 \frac{\sigma_0^2}{\sigma^4} Z_i^2 A_i(\theta)^2 h_n \right\} \\
&\qquad + \frac{K_i(\theta_0,\sigma)^2}{2!} \mathrm{e}^{\zeta_{1,i}} + \frac{K_i(\theta,\sigma)^3}{3!} \mathrm{e}^{\zeta_{2,i}} \\
&\quad = \alpha \frac{\sigma_0}{\sigma^2} Z_i A_i(\theta) \sqrt{h_n} + \frac{\alpha}{2} \frac{1}{\sigma^2} \left(1 - \alpha \frac{\sigma_0^2}{\sigma^2} Z_i^2\right) A_i(\theta)^2 h_n + H_{i,n} \\
&\qquad + \frac{K_i(\theta_0,\sigma)^2}{2!} \mathrm{e}^{\zeta_{1,i}} + \frac{K_i(\theta,\sigma)^3}{3!} \mathrm{e}^{\zeta_{2,i}},
\end{aligned}
$$

where $|\zeta_{1,i}| \le |K_i(\theta_0,\sigma)|$, $|\zeta_{2,i}| \le |K_i(\theta,\sigma)|$, and

$$
H_i(\theta,\sigma) = H_{n,i}(\theta,\sigma) := \alpha \frac{A_i(\theta)\Delta_i}{\sigma^2} - \frac{1}{2} \left(K_i(\theta,\sigma)^2 - \alpha^2 \frac{\sigma_0^2}{\sigma^4} Z_i^2 A_i(\theta)^2 h_n \right).
$$

Owing to Lemma 2, we have that

$$\begin{aligned}\sup_{\eta}\max_{1\le i\le n}|H_i(\theta,\sigma)| &= o_p(h_n^{r_1}), \quad 0\le r_1<1.5,\\ \sup_{\eta}\max_{1\le i\le n}|K_i(\theta,\sigma)^3| &= o_p(h_n^{r_2}), \quad 0\le r_2<1.5,\\ \sup_{\eta}\max_{1\le i\le n}|K_i(\theta_0,\sigma)^2| &= o_p(h_n^{r_3}), \quad 0\le r_3<2,\end{aligned} \tag{7}$$

and

$$\begin{aligned}\sup_{\eta}\max_{1\le i\le n}|\mathrm{e}^{\zeta_{1,i}}| &\le \exp\left\{\sup_{\eta}\max_{1\le i\le n}|K_i(\theta_0,\sigma)|\right\} = O_P(1),\\ \sup_{\eta}\max_{1\le i\le n}|\mathrm{e}^{\zeta_{2,i}}| &\le \exp\left\{\sup_{\eta}\max_{1\le i\le n}|K_i(\theta,\sigma)|\right\} = O_P(1).\end{aligned} \tag{8}$$

By (7) and (8), we have that

$$\begin{aligned}\sup_{\eta}\max_{1\le i\le n} R_i(\theta,\sigma) &:= \sup_{\eta}\max_{1\le i\le n}\left\{H_{i,n}(\theta,\sigma)+\frac{K_i(\theta_0,\sigma)^2}{2!}\mathrm{e}^{\zeta_{1,i}}+\frac{K_i(\theta,\sigma)^3}{3!}\mathrm{e}^{\zeta_{2,i}}\right\}\\ &= o_P(h_n^r) \qquad \text{for } 0\le r<1.5.\end{aligned} \tag{9}$$

Thus, by Lemmas 4 and 5 and (9), we have that uniformly in (θ,σ),

$$\begin{aligned}&\frac{1}{nh_n}\sum_{i=1}^n V_{n,i}^{\alpha}(\theta,\sigma)-\frac{1}{nh_n}\sum_{i=1}^n V_{n,i}^{\alpha}(\theta_0,\sigma)\\ &=\frac{1+\alpha}{2}\left(\frac{1}{\sigma}\right)^{\alpha+2}\frac{1}{n}\sum_{i=1}^n\left(1-\alpha\frac{\sigma_0^2}{\sigma^2}Z_i^2\right)A_i(\theta)^2\,\mathrm{e}^{-\frac{\alpha}{2}\frac{\sigma_0^2}{\sigma^2}Z_i^2}+(1+\alpha)\left(\frac{1}{\sigma}\right)^{\alpha+2}\\ &\quad\times\left\{\frac{1}{n\sqrt{h_n}}\sum_{i=1}^n\sigma_0 Z_i A_i(\theta)\,\mathrm{e}^{-\frac{\alpha}{2}\frac{\sigma_0^2}{\sigma^2}Z_i^2}+\frac{1}{\alpha}\frac{1}{nh_n}\sum_{i=1}^n R_i(\theta,\sigma)\,\mathrm{e}^{-\frac{\alpha}{2}\frac{\sigma_0^2}{\sigma^2}Z_i^2}\right\}\\ &\xrightarrow{P}\frac{1}{2}(1+\alpha)\left(\frac{1}{\sigma}\right)^{\alpha+2}\left(1+\alpha\frac{{\sigma_0}^2}{\sigma^2}\right)^{-\frac{3}{2}}\int\{a(x,\theta)-a(x,\theta_0)\}^2\,\mathrm{d}\mu_0(x).\end{aligned}$$

This completes the proof. □

Proof of Theorem 2 Let $\hat{\eta}_n^{\alpha}=(\hat{\theta}_n^{\alpha},\hat{\sigma}_n^{\alpha})$, $l_n^{\alpha}(\eta)=\sum_{i=1}^n V_{n,i}^{\alpha}(\eta)$,

$$L_n^{\alpha}=\begin{pmatrix}-\frac{1}{\sqrt{nh_n}}\partial_{\theta}l_n^{\alpha}(\eta_0)\\ -\frac{1}{\sqrt{n}}\partial_{\sigma}l_n^{\alpha}(\eta_0)\end{pmatrix},\quad\text{and}\quad C_n^{\alpha}(\eta)=\begin{pmatrix}\frac{1}{nh_n}\partial^2_{\theta\theta^T}l_n^{\alpha}(\eta) & \frac{1}{n\sqrt{h_n}}\partial^2_{\theta\sigma}l_n^{\alpha}(\eta)\\ \frac{1}{n\sqrt{h_n}}\partial^2_{\sigma\theta^T}l_n^{\alpha}(\eta) & \frac{1}{n}\partial^2_{\sigma^2}l_n^{\alpha}(\eta)\end{pmatrix},$$

where $\partial^2_{\theta\theta^T} = \frac{\partial^2}{\partial\theta\,\partial\theta^T}$, $\partial^2_{\theta\sigma} = \frac{\partial^2}{\partial\theta\,\partial\sigma}$. Using Taylor's theorem, we have that

$$0 = \nabla l_n^\alpha(\eta_0) + \int_0^1 \nabla^2 l_n^\alpha\left(\eta_0 + u(\hat{\eta}_n^\alpha - \eta_0)\right)\mathrm{d}u \begin{pmatrix} \hat{\theta}_n^\alpha - \theta_0 \\ \hat{\sigma}_n^\alpha - \sigma_0 \end{pmatrix},$$

from which we can write that

$$L_n^\alpha = \int_0^1 C_{\alpha,n}\left(\eta_0 + u(\hat{\eta}_n^\alpha - \eta_0)\right)\mathrm{d}u \begin{pmatrix} \sqrt{nh_n}(\hat{\theta}_n^\alpha - \theta_0) \\ \sqrt{n}(\hat{\sigma}_n^\alpha - \sigma_0) \end{pmatrix}.$$

We intend to show that

$$L_n^\alpha \xrightarrow{d} L \in N(0,\ \Sigma_{1,\alpha}), \tag{10}$$

where

$$\Sigma_{1,\alpha} = \frac{1}{\sigma_0^{2\alpha+2}} \begin{pmatrix} \frac{(1+\alpha)^2}{(1+2\alpha)^{\frac{3}{2}}} S & 0 \\ 0 & \left(2\frac{(1+\alpha)^2(1+2\alpha^2)}{(1+2\alpha)^2\sqrt{1+2\alpha}} - \frac{\alpha^2}{1+\alpha}\right) \end{pmatrix},$$

and S is the one defined in $(A7)$. Let

$$U_{n,i}^\alpha(\eta) := \left(\frac{1}{\sigma}\right)^\alpha \left\{\frac{1}{\sqrt{1+\alpha}} - \left(1 + \frac{1}{\alpha}\right)\exp\left(-\frac{\alpha}{2}\frac{(A_{i-1}(\theta)h_n + \sigma_0 Z_i\sqrt{h_n})^2}{\sigma^2 h_n}\right)\right\}, \tag{11}$$

and

$$\zeta_i = \zeta_{n,i} := \begin{pmatrix} \frac{1}{\sqrt{nh_n}}\,\partial_\theta U_{n,i}^\alpha(\eta_0) \\ \frac{1}{\sqrt{n}}\,\partial_\sigma U_{n,i}^\alpha(\eta_0) \end{pmatrix},$$

then, due to Lemma 6, (10) can be verified if we show that

$$\sum_{i=1}^n \mathbb{E}_0\left(\zeta_i \,\middle|\, \mathscr{G}\right) \xrightarrow{P} 0, \tag{12}$$

$$\sum_{i=1}^n \mathbb{E}_0\left(\zeta_i\,\zeta_i^T \,\middle|\, \mathscr{G}\right) \xrightarrow{P} \Sigma_{1,\alpha}, \tag{13}$$

$$\sum_{i=1}^n \mathbb{E}_0\left(|\zeta_i|^4 \,\middle|\, \mathscr{G}\right) \xrightarrow{P} 0 \tag{14}$$

[cf. Theorems 3.2 and 3.4 of Hall and Heyde (1980)]. Note that

$$\partial_\theta U^\alpha_{n,i}(\eta_0) = -\sqrt{h_n}\,\frac{1+\alpha}{\sigma_0^{\alpha+1}} Z_i\,\partial_\theta a_{i-1}(\theta_0)\,\mathrm{e}-\frac{\alpha}{2}Z_i^2, \tag{15}$$

$$\partial_\sigma U^\alpha_{n,i}(\eta_0) = -\frac{\alpha}{\sqrt{1+\alpha}}\frac{1}{\sigma_0^{\alpha+1}} + \frac{1+\alpha}{\sigma_0^{\alpha+1}}\left(1 - Z_i^2\right)\mathrm{e}^{-\frac{\alpha}{2}Z_i^2}, \tag{16}$$

$$\partial_\theta U^\alpha_{n,i}\;\partial_{\theta^T} U^\alpha_{n,i}(\eta_0) = h_n\frac{(1+\alpha)^2}{\sigma_0^{2\alpha+2}} Z_i^2\,\partial_\theta a_{i-1}(\theta_0)\,\partial_{\theta^T} a_{i-1}(\theta_0)\,\mathrm{e}^{\alpha Z_i^2},$$

$$\begin{aligned}\left(\partial_\sigma U^\alpha_{n,i}(\eta_0)\right)^2 &= \frac{\alpha^2}{1+\alpha}\frac{1}{\sigma_0^{2\alpha+2}} - 2\frac{\alpha\sqrt{1+\alpha}}{\sigma_0^{2\alpha+2}}\left(1 - Z_i^2\right)\mathrm{e}^{-\frac{\alpha}{2}Z_i^2}\\ &\quad+\frac{(1+\alpha)^2}{\sigma_0^{2\alpha+2}}\left(1 - Z_i^2\right)^2\mathrm{e}^{-\alpha Z_i^2},\end{aligned}$$

and

$$\begin{aligned}\partial_\theta U^\alpha_{n,i}\;\partial_\sigma U^\alpha_{n,i}(\eta_0) &= -\frac{\alpha}{\sqrt{1+\alpha}}\frac{1}{\sigma_0^{\alpha+1}}\partial_\theta U^\alpha_{n,i}(\eta_0)\\ &\quad-\sqrt{h_n}\frac{(1+\alpha)^2}{\sigma_0^{2\alpha+2}} Z_i(1 - Z_i^2)\partial_\theta a_{i-1}(\theta_0)\,\mathrm{e}^{-\alpha Z_i^2}.\end{aligned}$$

Utilizing the arguments:

$$\begin{aligned}\mathbb{E}_0\{Z_i\mathrm{e}^{-\alpha Z_i^2/2}|\mathscr{G}\} &= 0,\\ \mathbb{E}_0\{Z_i^2\mathrm{e}^{-\alpha Z_i^2}|\mathscr{G}\} &= (1+2\alpha)^{-\frac{3}{2}},\\ \mathbb{E}_0\left\{\left(1 - Z_i^2\right)\mathrm{e}^{-\frac{\alpha}{2}Z_i^2}|\mathscr{G}\right\} &= \alpha\,(1+\alpha)^{-\frac{3}{2}},\\ \mathbb{E}_0\left\{\left(1 - Z_i^2\right)^2\mathrm{e}^{-\alpha Z_i^2}|\mathscr{G}\right\} &= 2\left(1+2\alpha^2\right)(1+2\alpha)^{-\frac{5}{2}},\end{aligned}$$

and

$$\mathbb{E}_0\left\{\left(Z_i - Z_i^3\right)\mathrm{e}^{-\alpha Z_i^2}|\mathscr{G}\right\} = 0,$$

by simple calculus, we can readily check (12) and (13). Since (14) holds due to the facts:

$$\mathbb{E}_0\left\{\left|\partial_\theta U^\alpha_{n,i}(\eta_0)\right|^4|\mathscr{G}\right\} \le h_n^2 C_\alpha\left(1+\left|X_{t^n_{i-1}}\right|\right)^C,$$

and

$$\mathbb{E}_0\left\{\left|\partial_\sigma U^\alpha_{n,i}(\eta_0)\right|^4|\mathscr{G}\right\} \le C_\alpha,$$

(10) is established.

Now, owing to (10), the theorem is asserted if we verify that

$$\int_0^1 C_{\alpha,n}\left(\eta_0 + u(\hat{\eta}_n^\alpha - \eta_0)\right) \mathrm{d}u \xrightarrow{P} \frac{1}{\sigma_0^{\alpha+2}} \begin{pmatrix} \frac{1}{\sqrt{1+\alpha}} S & 0 \\ 0 & \frac{\alpha^2+2}{(1+\alpha)^{\frac{3}{2}}} \end{pmatrix}. \tag{17}$$

For this task, it suffices to show that

$$C_{\alpha,n}(\eta_0) \xrightarrow{P} \frac{1}{\sigma_0^{\alpha+2}} \begin{pmatrix} \frac{1}{\sqrt{1+\alpha}} S & 0 \\ 0 & \frac{\alpha^2+2}{(1+\alpha)^{\frac{3}{2}}} \end{pmatrix}, \tag{18}$$

and

$$\sup_{|\eta| \le \varepsilon_n} \left| C_{\alpha,n}(\eta_0 + \eta) - C_{\alpha,n}(\eta_0) \right| \xrightarrow{P} 0, \tag{19}$$

where $\{\varepsilon_n\}$ is any positive real sequence decaying to 0.

We first deal with (18). Put

$$\begin{aligned}
J_{1,i}(\eta) &= A_i(\theta)\partial^2_{\theta\theta^T} a_{i-1}(\theta) + \left(\alpha \frac{\sigma_0^2}{\sigma^2} Z_i^2 - 1\right) \partial_\theta a_{i-1}(\theta)\, \partial_{\theta^T} a_{i-1}(\theta), \\
J_{2,i}(\eta) &= \Delta_i\, \partial^2_{\theta\theta^T} a_{i-1}(\theta) + \frac{\alpha}{\sigma^2} \Big\{ 2\sigma_0 Z_i A_i(\theta) h_n^{\frac{3}{2}} + 2\sigma_0 Z_i \Delta_i \sqrt{h_n} + A_i(\theta)^2 h_n^2 \\
&\qquad + 2A_i(\theta)\Delta_i h_n + \Delta_i^2 \Big\} \partial_\theta a_{i-1}(\theta)\, \partial_{\theta^T} a_{i-1}(\theta), \\
J_{3,i}(\eta) &= \frac{4}{\alpha} \frac{\sigma_0^2}{\sigma^2} Z_i^2 K_i(\eta) + \frac{4}{\alpha^2} K_i^2(\eta),
\end{aligned}$$

and

$$J_{4,i}(\eta) = -2\sigma_0 Z_i K_i(\eta)\sqrt{h_n} + (A_i(\theta)h_n + \Delta_i)\left(\alpha + 2 - \alpha \frac{\sigma_0^2}{\sigma^2} Z_i^2 - 2K_i(\eta)\right).$$

It is not difficult to see that

$$\left|J_{1,i}(\eta)\right| \le C_\alpha \left(1 + Z_i^2\right)\left(1 + \left|X_{t_{i-1}^n}\right|\right)^C. \tag{20}$$

Also, in view of Lemma 2, we can see that

$$\begin{aligned}
&\sup_\eta \max_{1\le i\le n} |J_{2,i}(\eta)| = o_P(h_n^r), \quad 0 \le r < 1.5, \\
&\sup_\eta \max_{1\le i\le n} |J_{3,i}(\eta)| = o_P(h_n^r), \quad 0 \le r < 0.5, \\
&\sup_\eta \max_{1\le i\le n} |J_{4,i}(\eta)| = o_P(h_n^r), \quad 0 \le r < 1.
\end{aligned} \tag{21}$$

Meanwhile, we have that

$$
\begin{aligned}
\partial^2_{\theta\theta^T} l_n^\alpha(\eta) &= -\frac{1+\alpha}{\sigma^{\alpha+2}} \sum_{i=1}^n \left\{\sigma_0 \partial^2_{\theta\theta^T} a_{i-1}(\theta) Z_i \sqrt{h_n} + J_{1,i}(\eta) h_n + J_{2,i}(\eta)\right\} \\
&\quad \times \mathrm{e}^{-\frac{\alpha}{2}\frac{\sigma_0^2}{\sigma^2} Z_i^2}\, \mathrm{e}^{-K_i(\eta)} \\
&= -\frac{1+\alpha}{\sigma^{\alpha+2}} \sum_{i=1}^n \sigma_0 \partial^2_{\theta\theta^T} a_{i-1}(\theta) Z_i \sqrt{h_n}\, \mathrm{e}^{-\frac{\alpha}{2}\frac{\sigma_0^2}{\sigma^2} Z_i^2} \\
&\quad -\frac{1+\alpha}{\sigma^{\alpha+2}} \sum_{i=1}^n \left\{ J_{1,i}(\eta) - \alpha \frac{\sigma_0^2}{\sigma^2}\, \partial^2_{\theta\theta^T} a_{i-1}(\theta) Z_i^2 A_i(\theta) \right\} h_n\, \mathrm{e}^{-\frac{\alpha}{2}\frac{\sigma_0^2}{\sigma^2} Z_i^2} \\
&\quad -\frac{1+\alpha}{\sigma^{\alpha+2}} \sum_{i=1}^n R_{1,i}(\eta), \\
\partial^2_{\sigma^2} l_n^\alpha(\eta) &= \alpha\sqrt{1+\alpha}\frac{n}{\sigma^{\alpha+2}} - \frac{1+\alpha}{\sigma^{\alpha+2}} \sum_{i=1}^n \left\{ 1+\alpha - (3+2\alpha)\frac{\sigma_0^2}{\sigma^2} Z_i^2 + \alpha \frac{\sigma_0^4}{\sigma^4} Z_i^4 \right\} \\
&\quad \times \mathrm{e}^{-\frac{\alpha}{2}\frac{\sigma_0^2}{\sigma^2} Z_i^2} + \frac{1+\alpha}{\sigma^{\alpha+2}} \sum_{i=1}^n \left\{ (3+2\alpha)\frac{2}{\alpha} K_i(\eta) - \alpha J_{3,i}(\eta) \right\} \mathrm{e}^{-\frac{\alpha}{2}\frac{\sigma_0^2}{\sigma^2} Z_i^2} \\
&\quad + \frac{1+\alpha}{\sigma^{\alpha+2}} \sum_{i=1}^n R_{2,i}(\eta)\, \mathrm{e}^{-\frac{\alpha}{2}\frac{\sigma_0^2}{\sigma^2} Z_i^2} K_i(\eta)\, \mathrm{e}^{\zeta}, \\
\partial^2_{\theta\sigma} l_n^\alpha(\eta) &= \frac{1+\alpha}{\sigma^{\alpha+3}} \sum_{i=1}^n \partial_\theta a_{i-1}(\theta) \left(\alpha + 2 - \alpha \frac{\sigma_0^2}{\sigma^2} Z_i^2 \right) \sigma_0 Z_i \sqrt{h_n}\, \mathrm{e}^{-\frac{\alpha}{2}\frac{\sigma_0^2}{\sigma^2} Z_i^2} \\
&\quad + \frac{1+\alpha}{\sigma^{\alpha+3}} \sum_{i=1}^n \partial_\theta a_{i-1}(\theta) J_{4,i}(\eta)\, \mathrm{e}^{-\frac{\alpha}{2}\frac{\sigma_0^2}{\sigma^2} Z_i^2} \\
&\quad - \frac{1+\alpha}{\sigma^{\alpha+3}} \sum_{i=1}^n \partial_\theta a_{i-1}(\theta) R_{3,i}(\eta)\, \mathrm{e}^{-\frac{\alpha}{2}\frac{\sigma_0^2}{\sigma^2} Z_i^2} K_i(\eta)\, \mathrm{e}^{\zeta},
\end{aligned}
$$

where $|\zeta| \le |K_i(\eta)|$, and

$$
\begin{aligned}
M_i(\eta) &= \sigma_0 \partial^2_{\theta\theta^T} a_{i-1}(\theta) Z_i \sqrt{h_n} + J_{1,i}(\eta) h_n + J_{2,i}(\eta), \\
R_{1,i}(\eta) &= \left\{ J_{2,i}(\eta) - \left(J_{1,i}(\eta) h_n + J_{2,i}(\eta) \right) \alpha \frac{\sigma_0}{\sigma^2} Z_i A_i(\theta) \sqrt{h_n} \right\} \mathrm{e}^{-\alpha \frac{\sigma_0^2}{\sigma^2} Z_i^2}, \\
&\quad + M_i(\eta) \left\{ \alpha \frac{\sigma_0}{\sigma^2} Z_i A_i(\theta) \sqrt{h_n} - K_{2,i}(\eta) + \frac{1}{2} K_i^2(\eta)\, \mathrm{e}^{\zeta} \right\} \mathrm{e}^{-\alpha \frac{\sigma_0^2}{\sigma^2} Z_i^2}, \\
R_{2,i}(\eta) &= 1 + \alpha - (3+2\alpha) \left\{ \frac{\sigma_0^2}{\sigma^2} Z_i^2 + \frac{2}{\alpha} K_i(\eta) \right\} + \alpha \left\{ \frac{\sigma_0^4}{\sigma^4} Z_i^4 + J_{3,i}(\eta) \right\}, \\
R_{3,i}(\eta) &= \sigma_0 \left(\alpha + 2 - \alpha \frac{\sigma_0^2}{\sigma^2} Z_i^2 \right) Z_i \sqrt{h_n} + J_{4,i}(\eta).
\end{aligned}
$$

Then, using (8), (21) and Lemma 2, we can have that

$$
\begin{aligned}
\sup_{\eta} \max_{1\le i\le n} |R_{1,i}(\eta)| &= o_p(h_n^r), \quad 0 \le r < 1.5, \\
\sup_{\eta} \max_{1\le i\le n} |R_{2,i}(\eta) K_i(\eta)\, \mathrm{e}^{\zeta}| &= o_p(h_n^r), \quad 0 \le r < 0.5, \\
\sup_{\eta} \max_{1\le i\le n} |R_{3,i}(\eta) K_i(\eta)\, \mathrm{e}^{\zeta}| &= o_p(h_n^r), \quad 0 \le r < 1.
\end{aligned} \tag{22}
$$

Therefore, by Lemmas 4 and 5, (21) and (22),

$$
\begin{aligned}
&\frac{1}{nh_n}\, \partial^2_{\theta\theta^T} l_n^{\alpha}(\eta) \xrightarrow{P} \frac{1+\alpha}{\sigma^{\alpha+2}} \left(1+\alpha\frac{\sigma_0^2}{\sigma^2}\right)^{-3/2} \\
&\quad \times \left\{ S - \int (a(x,\theta_0) - a(x,\theta))\, \partial^2_{\theta\theta^T} a(x,\theta)\, \mu_0(x) \right\}, \\
&\frac{1}{n}\, \partial^2_{\sigma^2} l_n^{\alpha}(\eta) \xrightarrow{P} \frac{\alpha\sqrt{1+\alpha}}{\sigma^{\alpha+2}} - \frac{1+\alpha}{\sigma^{\alpha+2}} \left\{ (1+\alpha)\left(1+\alpha\frac{\sigma_0^2}{\sigma^2}\right)^{-\frac{1}{2}} \right. \\
&\quad \left. -(3+2\alpha)\frac{\sigma_0^2}{\sigma^2}\left(1+\alpha\frac{\sigma_0^2}{\sigma^2}\right)^{-\frac{3}{2}} + 3\alpha\frac{\sigma_0^4}{\sigma^4}\left(1+\alpha\frac{\sigma_0^2}{\sigma^2}\right)^{-\frac{5}{2}} \right\}, \\
&\frac{1}{n\sqrt{h_n}}\, \partial^2_{\theta\sigma} l_n^{\alpha}(\eta) \xrightarrow{P} 0
\end{aligned}
$$

uniformly in η. Thus (18) is verified. Since the above limits are continuous in η by $(A2)$ and $(A5)$, we can demonstrate that (19) holds. This completes the proof. □

5 Some Lemmas

Lemma 1 *Suppose that* (A1) *holds. Then, for $k \ge 1$,*

$$
\mathbb{E}_0 \left\{ |\Delta_i|^k \middle| \mathcal{G}_{i-1}^n \right\} \le C_k h_n^{\frac{3}{2}k} \left(1 + \left|X_{t_{i-1}^n}\right|\right)^k .
$$

Since the result of Lemma 1 is well known, we omit the proof.

Lemma 2 *Suppose that $f : \mathbb{R} \times \Theta \to \mathbb{R}$ belongs to $\mathcal{P}$. Then if* (A1) *and* (A3) *hold and $nh_n^q \to 0$ for some $q > 1$,*

$$
\sup_{\eta} \max_{1\le i\le n} \left| f^j\left(X_{t_{i-1}^n}, \eta\right) Z_i^k\, \Delta_i^l\, h_n^m \right| = o_P(h_n^r) \qquad \textit{for } 0 \le r < 1.5\, l + m,
$$

where $j, k, l \in \{0, 1, 2, \cdots\}$ and $m > -1.5\, l$.

Proof For any integer p with $(1.5l + m - r)p \geq q$, we have that

$$n \max_{1 \leq i \leq n} \mathbb{E}_0 \left| \left(1 + \left|X_{t_{i-1}^n}\right|\right)^C Z_i^k \Delta_i^l \right|^p h_n^{(m-r)p} \leq Cnh_n^{(1.5l+m-r)p} = o(1),$$

which establishes the lemma. □

Lemma 3 *Suppose that* (A1) *and* (A3) *hold. Then, for* $f : \mathbb{R} \times \Theta \to \mathbb{R}$ *in* $\mathscr{P}$*,*

$$\frac{1}{n} \sum_{i=1}^{n} f\left(X_{t_{i-1}^n}, \eta\right) = O_P(1). \tag{23}$$

If in addition, $nh_n \to \infty$*,* (A2) *holds and* f *is differentiable with respect to* x *and* η *whose derivatives also belong to* $\mathcal{P}$*, then*

$$\frac{1}{n} \sum_{i=1}^{n} f\left(X_{t_{i-1}^n}, \eta\right) \xrightarrow{P} \int f(x, \eta) \mathrm{d}\mu_0(x) \tag{24}$$

uniformly in η*.*

It is easy to prove (23), so we omit the proof. The argument in (24) is due to Lemma 8 of Kessler (1997).

Lemma 4 *Suppose that* (A1)–(A3) *hold. Then if* $f : R \times \Theta \to R \in \mathscr{P}$ *is differentiable with respect to* x *and* η *with derivatives belonging to* $\mathscr{P}$*,*

$$\frac{1}{n} \sum_{i=1}^{n} f\left(X_{t_{i-1}^n}, \eta\right) \mathrm{e}^{-\frac{\alpha}{2} \frac{\sigma_0^2}{\sigma^2} Z_i^2} \xrightarrow{P} \left(1 + \alpha \frac{\sigma_0^2}{\sigma^2}\right)^{-\frac{1}{2}} \int f(x, \eta)\, \mathrm{d}\mu_0(x), \tag{25}$$

$$\frac{1}{n} \sum_{i=1}^{n} f\left(X_{t_{i-1}^n}, \eta\right) Z_i^2\, \mathrm{e}^{-\frac{\alpha}{2} \frac{\sigma_0^2}{\sigma^2} Z_i{}^2} \xrightarrow{P} \left(1 + \alpha \frac{\sigma_0^2}{\sigma^2}\right)^{-\frac{3}{2}} \int f(x, \eta)\, \mathrm{d}\mu_0(x), \tag{26}$$

$$\frac{1}{n} \sum_{i=1}^{n} f\left(X_{t_{i-1}^n}, \eta\right) Z_i^4\, \mathrm{e}^{-\frac{\alpha}{2} \frac{\sigma_0^2}{\sigma^2} Z_i{}^2} \xrightarrow{P} 3\left(1 + \alpha \frac{\sigma_0^2}{\sigma^2}\right)^{-\frac{5}{2}} \int f(x, \eta)\, \mathrm{d}\mu_0(x) \tag{27}$$

uniformly in η*.*

Proof Let

$$h_i(\eta) = \frac{1}{n} f\left(X_{t_{i-1}^n}, \eta\right) \mathrm{e}^{-\frac{\alpha}{2} \frac{\sigma_0^2}{\sigma^2} Z_i^2}.$$

In view of Lemma 9 of Genon-Catalot and Jacod (1993), the convergence result for each η is ensured by the facts:

$$\begin{aligned}\sum_{i=1}^{n} \mathbb{E}_0 \left\{h_i(\eta)|\mathcal{G}_{i-1}^n\right\} &= \left(1+\alpha\frac{\sigma_0^2}{\sigma^2}\right)^{-1/2} \frac{1}{n}\sum_{i=1}^{n} f(X_{t_{i-1}^n},\eta)\\ &\xrightarrow{P} \left(1+\alpha\frac{\sigma_0^2}{\sigma^2}\right)^{-1/2} \int f(x,\eta)\mathrm{d}\mu_0(x),\end{aligned}$$

and

$$\sum_{i=1}^{n} \mathbb{E}_0 \left\{(h_i(\eta))^2|\mathcal{G}_{i-1}^n\right\} \le \frac{1}{n^2}\sum_{i=1}^{n} f^2(X_{t_{i-1}^n},\eta) = o_P(1).$$

To establish the uniform convergence, in view of Theorem 20 of Ibragimov and Has'minskii (1981, p 378)), it suffices to verify that

$$\mathbb{E}_0\left\{\left|\sum_{i=1}^{n} h_i(\eta)\right|^{2d}\right\} \le C \quad \text{for all } \ \eta, \tag{28}$$

and

$$\mathbb{E}_0\left\{\left|\sum_{i=1}^{n} h_i(\eta_1) - \sum_{i=1}^{n} h_i(\eta_2)\right|^{2d}\right\} \le C|\eta_1-\eta_2|^{2d} \quad \text{for all } \ \eta_1,\eta_2, \tag{29}$$

where d is the dimension of Θ. We only prove (29) since (28) can be proved in essentially the same way. Note that

$$\sup_{\eta}\left|\nabla h_i(\eta)\right| \le \frac{C_\alpha}{n}|1+Z_i^2|\left(1+\left|X_{t_{i-1}^n}\right|\right)^C.$$

Then, using Cauchy's inequality and Jensen's inequality, we have that

$$\begin{aligned}&\mathbb{E}_0\left\{\left|\sum_{i=1}^{n}(h_i(\eta_1)-h_i(\eta_2))\right|^{2d}\right\}\\ &\le \mathbb{E}_0\left\{\left|\sum_{i=1}^{n}\nabla h_i(\eta^*)(\eta_1-\eta_2)\right|^{2d}\right\}\\ &\le C_\alpha|\eta_1-\eta_2|^{2d}\,\mathbb{E}_0\left\{\frac{1}{n}\sum_{i=1}^{n}\left|1+Z_i^2\right|^C\left(1+\left|X_{t_{i-1}^n}\right|\right)^C\right\}\\ &\le C_\alpha|\eta_1-\eta_2|^{2d},\end{aligned}$$

where η^* lies between η_1 and η_2. This asserts (29). In a similar fashion, we can verify (26) and (27). □

Lemma 5 *Under the conditions in Lemma 4, if $nh_n \to \infty$, then*

$$\frac{1}{n\sqrt{h_n}} \sum_{i=1}^{n} f\left(X_{t_{i-1}^n}, \eta\right) \, Z_i \, \mathrm{e}^{-\frac{\alpha}{2}\frac{\sigma_0^2}{\sigma^2} Z_i^2} = o_P(1) \tag{30}$$

uniformly in η.

Proof Let

$$h_i(\eta) = \frac{1}{n\sqrt{h_n}} f(X_{t_{i-1}^n}, \eta) \, Z_i \, \mathrm{e}^{-\frac{\alpha}{2}\frac{\sigma_0^2}{\sigma^2} Z_i^2}. \tag{31}$$

From the facts:

$$\sum_{i=1}^{n} \mathbb{E}_0 \left\{ h_i(\eta) \middle| \mathcal{G}_{i-1}^n \right\} = 0,$$

and

$$\sum_{i=1}^{n} \mathbb{E}_0 \left\{ h_i^2(\eta) \middle| \mathcal{G}_{i-1}^n \right\} \le \frac{1}{n^2 h_n} \sum_{i=1}^{n} f^2\left(X_{t_{i-1}^n}, \eta\right) = o_P(1),$$

we can see that (30) holds for each η. Thus, as in the proof of Lemma 4, the lemma is proved if we show that (28) and (29) hold for (31).

By applying Burkhoder's inequality to a sequence of martingale differences $\left\{h_i(\eta), \mathcal{G}_i^n\right\}_{i=1}^n$, one can easily see that (28) holds. Also, applying Burkhoder's inequality to a sequence of martingale differences $\left\{h_i(\eta_1) - h_i(\eta_2), \mathcal{G}_i^n\right\}_{i=1}^n$ and using the facts:

$$\sup_{\eta} \left|\nabla h_i(\eta)\right| \le \frac{C_\alpha}{n\sqrt{h_n}} \left|Z_i + Z_i^3\right| \left(1 + \left|X_{t_{i-1}^n}\right|\right)^C,$$

Cauchy's inequality and Jensen's inequality, we have that

$$\begin{aligned}
&\mathbb{E}_0 \left\{ \left| \sum_{i=1}^{n} (h_i(\eta_1) - h_i(\eta_2)) \right|^{2d} \right\} \\
&\le \mathbb{E}_0 \left\{ \left| \sum_{i=1}^{n} (h_i(\eta_1) - h_i(\eta_2))^2 \right|^{d} \right\} \\
&\le |\eta_1 - \eta_2|^{2d} \, \mathbb{E}_0 \left\{ \left| \sum_{i=1}^{n} |\nabla h_i(\eta^*)|^2 \right|^{d} \right\} \\
&\le |\eta_1 - \eta_2|^{2d} \, \mathbb{E}_0 \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{C_\alpha}{n^d h_n^d} |Z_i + Z_i^3|^C \left(1 + \left|X_{t_{i-1}^n}\right|\right)^C \right\} \\
&\le \frac{C_\alpha}{n^d h_n^d} |\eta_1 - \eta_2|^{2d}.
\end{aligned}$$

This entails (29) and completes the proof. □

Lemma 6 *Let $U^{\alpha}_{n,i}(\eta)$ be the one defined in (11). Under (A1), (A3) and (A4), if $nh^2 \to 0$, then*

$$\left|\frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n}\partial_\theta V^{\alpha}_{n,i}(\eta_0) - \frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n}\partial_\theta U^{\alpha}_{n,i}(\eta_0)\right| = o_P(1), \tag{32}$$

$$\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\partial_\sigma V^{\alpha}_{n,i}(\eta_0) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\partial_\sigma U^{\alpha}_{n,i}(\eta_0)\right| = o_P(1). \tag{33}$$

Proof We only deal with (32) since (33) can be proven similarly. Note that

$$\partial_\theta V^{\alpha}_{n,i}(\eta_0) = -\frac{1+\alpha}{\sigma_0^{\alpha+2}}\left(\sigma_0 Z_i\sqrt{h_n} + \Delta_i\right)\partial_\theta a_{i-1}(\theta_0)\,\mathrm{e}^{-\frac{\alpha}{2}Z_i^2}\,\mathrm{e}^{-K_{i,0}},$$

where $K_{i,0} = K_{n,i}(\eta_0)$. It follows from (15) that

$$\begin{aligned}
&\frac{1}{\sqrt{nh_n}}\left|\sum_{i=1}^{n}\partial_\theta V^{\alpha}_{n,i}(\eta_0) - \sum_{i=1}^{n}\partial_\theta U^{\alpha}_{n,i}(\eta_0)\right| \\
&\le \frac{C_\alpha}{\sqrt{nh_n}}\sum_{i=1}^{n}\left(1+\left|X_{t^n_{i-1}}\right|\right)^C\left|(\sigma_0 Z_i\sqrt{h_n}+\Delta_i)(1-K_{i,0}\mathrm{e}^{\zeta}) - \sigma_0 Z_i\sqrt{h_n}\right| \\
&\le \frac{C_\alpha}{\sqrt{nh_n}}\sum_{i=1}^{n}\left(1+\left|X_{t^n_{i-1}}\right|\right)^C\left\{|Z_i K_{i,0}|\sqrt{h_n}\mathrm{e}^{\zeta} + |\Delta_i|\mathrm{e}^{-K_{i,0}}\right\} \\
&\le \mathrm{e}^{\max_i |K_{i,0}|}\frac{C_\alpha}{\sqrt{nh_n}}\sum_{i=1}^{n}\left(1+\left|X_{t^n_{i-1}}\right|\right)^C\left\{|Z_i K_{i,0}|\sqrt{h_n} + |\Delta_i|\right\},
\end{aligned}$$

where $|\zeta| \le |K_{i,0}|$. Using the facts that $\mathbb{E}_0(K^2_{i,0}|\mathcal{G}^n_{i-1}) \le C\left(1+\left|X_{t^n_{i-1}}\right|\right)^C h_n^2$ and $\mathbb{E}_0(K^4_{i,0}|\mathcal{G}^n_{i-1}) \le C\left(1+\left|X_{t^n_{i-1}}\right|\right)^C h_n^4$, we have

$$\begin{aligned}
&\frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n}\mathbb{E}_0\left\{\left(1+\left|X_{t^n_{i-1}}\right|\right)^C\left(|Z_i K_{i,0}|\sqrt{h_n}+|\Delta_i|\right)|\mathcal{G}^n_{i-1}\right\} \\
&\le \frac{C}{\sqrt{nh_n}}\sum_{i=1}^{n}\left(1+\left|X_{t^n_{i-1}}\right|\right)^C\left\{\sqrt{\mathbb{E}_0(Z_i^2)\mathbb{E}_0(K^2_{i,0}|\mathcal{G}^n_{i-1})}\sqrt{h_n} + \mathbb{E}_0(|\Delta_i||\mathcal{G}^n_{i-1})\right\} \\
&\le \frac{C}{\sqrt{nh_n}}\sum_{i=1}^{n}\left(1+\left|X_{t^n_{i-1}}\right|\right)^C h_n^{1.5} = O_P(\sqrt{n}h_n)
\end{aligned}$$

and

$$
\begin{aligned}
&\frac{1}{nh_n}\sum_{i=1}^{n}\mathbb{E}_0\left\{\left(1+\left|X_{t_{i-1}^n}\right|\right)^C\left(|Z_iK_{i,0}|\sqrt{h_n}+|\Delta_i|\right)^2\Big|\mathcal{G}_{i-1}^n\right\}\\
&\quad\le\frac{C}{nh_n}\sum_{i=1}^{n}\left(1+\left|X_{t_{i-1}^n}\right|\right)^C h_n^3=O_P(h_n^2).
\end{aligned}
$$

Since $\mathrm{e}^{\max_i|K_{i,0}|}=O_P(1)$, in view of Lemma 9 of Genon-Catalot and Jacod (1993), (32) is established. This completes the proof. □

Acknowledgments We would like to thank the Associate Editor and referees for their helpful comments.

References

Ait-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica, 70*, 223–262.

Basu, A., Harris, I. R., Hjort, N. L., Jones, M. C., et al. (1998). Robust and efficient estimation by minimizing a density power divergence. *Biometrika, 85*, 549–559.

Basu, S., Lindsay, B. G. (1994). Minimum disparity estimation for continuous models: Efficiency, distributions and robustness. *Annals of the Institute of Statistical Mathematics, 48*, 683–705.

Beran, R. (1977). Minimum hellinger distance estimates for parametric models. *Annals of Statistics, 5*, 445–463.

Cao, R., Cuevas, A., Fraiman, R., et al. (1995). Minimum distance density-based estimation. *Computational Statistics and Data Analysis, 20*, 611–631.

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

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

Genon-Catalot, V., Jacod, J. (1993). On the estimation of the diffusion coefficient for multidimensional diffusion processes. *Annales de l'Institut Henri Poincaré-Probabilités et Statistiques, 29*, 119–151.

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

Ibragimov, I. A., Has'minskii, R. Z. (1981). *Statistical estimation-asymptotic theory*. New York: Springer.

Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics, 24*, 211–229.

Kutoyants, Y. (2004). *Statistical inference for ergodic diffusion processes*. New York: Springer.

Lee, S., Na, O. (2005). Test for parameter change based on the estimator minimizing density-based divergence measures. *Annals of the Institute of Statistical Mathematics, 57*, 553–573.

Prakasa Rao, B. L. S. (1999). *Statistical inference for diffusion type processes*. London: Arnold.

Simpson, D. G. (1987). Minimum hellinger distance estimation for the analysis of count data. *Journal of the American Statistical Association, 82*, 802–807.

Tamura, R. N., Boos, D. D. (1986). Minimum hellinger distance estimation for multivariate location and covariance. *Journal of the American Statistical Association, 81*, 223–239.

Yoshida, N. (1992). Estimation for diffusion processes from discrete observations. *Journal of Multivariate Analysis, 41*, 220–242.