Ann Inst Stat Math (2012) 64:545–575
DOI 10.1007/s10463-010-0323-4

Estimation of parameters for discretely observed diffusion processes with a variety of rates for information

Yasutaka Shimizu

Received: 1 March 2010 / Revised: 4 October 2010 / Published online: 24 December 2010

Abstract A specific form of stochastic differential equation with unknown parameters are considered. We do not necessarily assume ergodicity or recurrency, and any moment conditions for the true process, but some tightness conditions for an information-like quantity. The interest is to estimate the parameters from discrete observations the step size of which tends to zero. Consistency and the rate of convergence of proposed estimators are presented. The rate is deduced naturally from the rate for the information-like quantities.

Keywords Parametric inference · Non-ergodic diffusions · Discrete observations · Consistency · Rates of convergence · Normalized information

1 Introduction

1.1 Model

We consider 1-dimensional stochastic processes $X^\vartheta = (X_t^\vartheta)_{t\geq 0}$ each of which solves the following specific form of stochastic differential equation.

$$X_0^\vartheta = X_0, \quad \mathrm{d}X_t^\vartheta = \mu U(X_t^\vartheta)\,\mathrm{d}t + \sqrt{\sigma}V(X_t^\vartheta)\,\mathrm{d}W_t, \tag{1}$$

Y. Shimizu (✉)
Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka 560-8531, Japan
e-mail: yasutaka@sigmath.es.osaka-u.ac.jp

Y. Shimizu
Japan Science and Technology Agency, CREST, Sanbancho, Chiyoda-ku, Tokyo 102-0075, Japan

where W is a Wiener process; X_0 is a random variable independent of W; $\vartheta = (\mu, \sigma)$ is a 2-dimensional unknown parameter belonging to a compact convex set $\Theta := \Xi \times \Pi \subset \mathbb{R} \times \mathbb{R}_+$; U and V are measurable functions. We also denote by $\vartheta_0 = (\mu_0, \sigma_0)$ the true parameter such that $\mu_0 \in \mathrm{int}(\Xi)$ and $\sigma_0 \in \mathrm{int}(\Pi)$, $X := X^{\vartheta_0}$ the data generating process. We assume that X is observed at discrete time points $t_i^n := ih_n$ $(i = 0, 1, 2, \dots, n)$ for positive numbers h_n, and put $T_n := t_n^n$. Throughout the paper, we assume a *non-degeneracy of diffusion coefficient*:

$$\inf_{x \in \mathbb{R}} |V(x)| > 0. \tag{2}$$

This condition is not essential, but for simplicity of discussion. We can drop off this assumption by using the data such that $X_\varepsilon^n := \left\{X_{t_i^n} \mid i = 0, 1, \dots, n,\ |V(X_{t_i^n})| > \varepsilon\right\}$ for $\varepsilon > 0$ small enough, and assuming that $\# X_\varepsilon^n \sim n$ as $n \to \infty$ which would not be so restrictive in practice.

In the sequel, we denote by $C(x)$ a $[1, \infty)$-valued function such that

$$|U(x)| + |V(x)| \lesssim C(x), \tag{3}$$

where the symbol $A \lesssim B$ means that there exists an absolute constant c_0 such that $A \leq c_0 B$. For example, one of the simplest choices is $C(x) = \sqrt{1 + |U(x)|^2 + |V(x)|^2}$. Moreover, we put

$$S(x) := \frac{U^2(x)}{V^2(x)}. \tag{4}$$

1.2 Parametric inference for non-ergodic diffusions

Estimation of ϑ_0 from a discrete sample $(X_{t_i^n})_{i=0}^n$ is a fundamental problem in application of diffusion processes, and there are many works on this issue when X is ergodic. When X is non-ergodic, there are only a few works for discretely observed cases: Kasonga (1990) proposed a trajectory-fitting estimator (TFE) for a drift parameter within a nonlinear drift, and gave a sufficient condition for the strong consistency of the TFE. His condition was essentially valid for non-ergodic case, but invalid in ergodic case. Jacod (2006) proposed a moment type contrast function which includes implicit functions $\phi_n(x, \vartheta) = E_\vartheta\left[X_{h_n}^\vartheta - x\right]$ and $\phi_n'(x, \vartheta) = E_\vartheta\left[(X_{h_n}^\vartheta - x)^2\right]$ such that

$$\begin{aligned}\Psi_n(\vartheta) &= \sum_{i=1}^n \frac{1}{C^4(X_{t_{i-1}^n})}\left|\Delta_i^n X - \phi_n(X_{t_{i-1}^n}, \vartheta)\right|^2 \\ &\quad + \sum_{i=1}^n \frac{1}{C^6(X_{t_{i-1}^n})}\left|(\Delta_i^n X)^2 - \phi_n'(X_{t_{i-1}^n}, \vartheta)\right|^2,\end{aligned}$$

and showed the consistency, and that an M-estimator

$$\hat{\vartheta}_n = (\hat{\mu}_n, \hat{\sigma}_n) := \arg\min_{\vartheta \in \Theta} \Psi_n(\vartheta) \tag{5}$$

is $(\sqrt{T_n}, \sqrt{n})$-consistent if $T_n \to \infty$ and $h_n \to 0$ under some *identifiability conditions*, which are reduced in our case of (1) to that

$$\left(T_n I_{n,4}^{-1},\ T_n J_{n,6}^{-1}\right) \text{ are tight,} \tag{6}$$

where

$$I_{n,k} := \int_0^{T_n} \frac{S(X_s)}{C^k(X_s)}\,\mathrm{d}s; \quad J_{n,k} := \int_0^{T_n} \frac{1}{C^k(X_s)}\,\mathrm{d}s;$$

C is a *smooth* function satisfying (3) with bounded derivatives. The weights $C^{-k}(X_{t_{i-1}^n})$'s in Ψ_n are used to ensure some integrability for Ψ_n. Note, in this case, that $\left(T_n^{-1} I_{n,4}, T_n^{-1} J_{n,6}\right)$ is also tight.

Shimizu (2009b) proposed quadratic type contrast functions in closed form based on discrete data. For instance, one of the simplest contrasts is written as follows:

$$\Psi_{n,k}(\vartheta) := \sum_{i=1}^{n} \frac{1}{C^k(X_{t_{i-1}^n})} \left[\frac{\left(\Delta_i^n X - \mu h_n U(X_{t_{i-1}^n})\right)^2}{2 h_n \sigma V^2(X_{t_{i-1}^n})} + \frac{1}{2}\log\sigma \right].$$

Note that $\Psi_{n,k}$ is a *normalization* of a local-Gauss approximation of the log-likelihood function of $(X_{t_i^n})_{i=0}^n$. The weight $C^{-k}(X_t)$ is also used for some integrability of $\Psi_{n,k}$; that is the meaning of *normalization*. Then the M-estimator $\hat{\vartheta}_{n,k}$ via $\Psi_{n,k}$ becomes consistent if

$$\left(T_n I_{n,k}^{-1},\ T_n J_{n,k}^{-1}\right) \quad \text{and} \quad \left(T_n^{-1} I_{n,k},\ T_n^{-1} J_{n,k}\right) \text{ are tight.} \tag{7}$$

The M-estimator attains the same rate of convergence with the estimator by Jacod (2006).

It would be natural that the quantities $I_{n,k}$ and $J_{n,k}$ appear in the inference for ϑ_0 since those quantities are related to a kind of information for the distribution $(X_{t_i^n})_{i=0}^n$. For example, considering the case where X is ergodic with an invariant probability measure π, we see under some regularities that the asymptotic Fisher information matrix, say K_0, becomes

$$K_0 := \mathrm{diag}\left(\int_{\mathbb{R}} \sigma_0^{-1} S(x)\,\pi(\mathrm{d}x), 2^{-1}\sigma_0^{-2}\right) \sim \mathrm{diag}\left(T_n^{-1} I_{n,0},\ T_n^{-1} J_{n,0}\right), \quad n \to \infty.$$

Here, we should note that K_0 is also an *asymptotic observed information*:

$$K_0 = P_{\vartheta_0}\text{-}\lim_{n\to\infty} \mathrm{diag}\left(\frac{1}{T_n}\sum_{i=1}^n \partial_\mu^2 \Psi_{n,0}(\vartheta_0), \frac{1}{n}\sum_{i=1}^n \partial_\sigma^2 \Psi_{n,0}(\vartheta_0)\right).$$

As a result, it follows that the M-estimator $\hat{\vartheta}_{n,0}$ is asymptotically normal with the asymptotic variance K_0^{-1}, which is also asymptotically efficient in the sense of minimum asymptotic variance; see Kessler (1997), Theorem1 and Remark 2, or Gobet (2002) for details. The rate of convergence becomes $(\sqrt{T_n}, \sqrt{n})$, which is naturally derived from the rate of *observed information*. On the other hand, in cases treated by Jacod (2006) and Shimizu (2009b), where the contrast functions are weighted by $C^{-k}(X_{t_{i-1}^n})$'s, the information is lost due to the weights by which the information processes are normalized so that $T_n I_{n,k}^{-1}$ and $T_n J_{n,k}^{-1}$ become tight. Consequently, a *normalized information* $I_{n,k}$ and $J_{n,k}$ will appear.

Condition (6) or (7) is not so restrictive if X is ergodic, and the rate of convergence $(\sqrt{T_n}, \sqrt{n})$ is the best attainable. However, in non-ergodic cases, many examples violate (6) or (7) with a variety of rates of convergence: e.g., for the drift, one is more rapid, and another is slower than $\sqrt{T_n}$. Therefore, in this paper, we consider more variety of classes for X such that

$$\left(\beta_{n,k} I_{n,k}^{-1},\ \gamma_{n,k} J_{n,k}^{-1}\right) \quad \text{and} \quad \left(\beta_{n,k}^{-1} I_{n,k},\ \gamma_{n,k}^{-1} J_{n,k}\right) \text{ are tight}$$

for some numbers $\beta_{n,k}$ and $\gamma_{n,k}$; see Section 2. Considering such a class, one could expect that the natural rate of convergence of estimator would be $(\sqrt{\beta_{n,k}}, \sqrt{\gamma_{n,k}/h_n})$ from the above consideration in ergodic cases. This paper actually gives such an estimator, which is consistent with one of the estimator given in Shimizu (2009b) if $\beta_{n,k} = \gamma_{n,k} = T_n$.

Our plan of the paper is as follows. In Sect. 2, we introduce a class of diffusions treated in this paper with some examples. The main results are presented in Sect. 3, where two cases are discussed separately: the case where $\beta_{n,k}/\gamma_{n,k} = O(1)$, and $\beta_{n,k}/\gamma_{n,k} \to \infty$. Section 4 is devoted to some examples, where ergodic models without moments, null- and non-recurrent models are dealt with. We shall show there that our estimator is possible to attain the best attainable rate of convergence. Section 5 is for the concluding remarks. All the proofs of our results are given in Sect. 6.

Although we consider just a simple form of SDEs in this paper, it is not essential, but for the simplicity of notation and discussion. We can deal with more general form of multidimensional SDEs: see Sect. 5 on this point.

2 A class of diffusion processes

Let

$$\mathcal{D}(U; V; \Theta) := \{X^\vartheta = (X_t)_{t\ge 0} \mid U, V \text{ are measurable functions with (2)}; \vartheta \in \Theta\}$$

be a family of random elements X^ϑ satisfying (1) for each U and V.

Definition 1 For a given function $C(x)$ and a real number $k \geq 0$,

$$\mathcal{D}_k(\beta_{n,k}, \gamma_{n,k}) := \left\{ X^\vartheta \in \mathcal{D}(U; V; \Theta) \,\middle|\, \left(\mathcal{I}_{n,k}, \mathcal{J}_{n,k}\right) \text{ and } \left(\mathcal{I}_{n,k}^{-1}, \mathcal{J}_{n,k}^{-1}\right) \text{ are tight.} \right\},$$

where

$$\mathcal{I}_{n,k} := \frac{1}{\beta_{n,k}} \int_0^{T_n} \frac{S(X_s)}{C^k(X_s)}\, \mathrm{d}s; \quad \mathcal{J}_{n,k} := \frac{1}{\gamma_{n,k}} \int_0^{T_n} \frac{1}{C^k(X_s)}\, \mathrm{d}s,$$

$\beta_{n,k}$ and $\gamma_{n,k}$ are deterministic positive numbers.

Note that the tightness of $\mathcal{I}_{n,k}$ and $\mathcal{J}_{n,k}$ are automatically satisfied if $k \geq 2$ and $\beta_{n,k} = \gamma_{n,k} = T_n$, which is the class treated in Jacod (2006), or Shimizu (2009b). In what follows, we shall give examples of some classes of diffusions.

Example 1 (Ergodic models without moments) Consider a class $\mathcal{D}_{\mathrm{Erg}}$ which is a subclass of $\mathcal{D}(U; V; \Theta)$ with $|U(x)|^2 + |V(x)|^2 \lesssim 1 + |x|^{c_0}$ such that X^ϑ is ergodic (positive recurrent) with an invariant measure π_ϑ: for $f(x)$ satisfying $\int_{\mathbb{R}} f(x)\,\pi_\vartheta(\mathrm{d}x) < \infty$,

$$\frac{1}{T_n} \int_0^{T_n} f(X_t^\vartheta)\, \mathrm{d}t \xrightarrow{P_{\vartheta_0}} \int_{\mathbb{R}} f(x)\,\pi_\vartheta(\mathrm{d}x), \quad T_n \to \infty.$$

If we take $k \geq 2$ and $C(x) = \sqrt{1 + |x|^{c_0}}$, then

$$\mathcal{I}_{n,k} = \frac{1}{T_n} \int_0^{T_n} \frac{S(X_s)}{C^k(X_s)}\, \mathrm{d}s \xrightarrow{P_{\vartheta_0}} \int_{\mathbb{R}} \frac{S(x)}{C^k(x)}\,\pi(\mathrm{d}x);$$
$$\mathcal{J}_{n,k} = \frac{1}{T_n} \int_0^{T_n} \frac{1}{C^k(X_s)}\, \mathrm{d}s \xrightarrow{P_{\vartheta_0}} \int_{\mathbb{R}} \frac{1}{C^k(x)}\,\pi(\mathrm{d}x),$$

even if

$$\int_{\mathbb{R}} S(x)\,\pi(\mathrm{d}x) = \infty.$$

Hence it follows that

$$\mathcal{D}_{\mathrm{Erg}} \subset \mathcal{D}_k(\beta_{n,k} = T_n, \gamma_{n,k} = T_n),$$

for $C(x) = \sqrt{1 + |x|^{c_0}}$ and $k \geq 2$.

Example 2 (Null-recurrent models) Let

$$\mathcal{D}_{\mathrm{NR}} := \left\{ X^\vartheta \in \mathcal{D}(U; V; \Theta) \,|\, U(x) = x/(1+x^2), V(x) = 1, |\mu| < \sigma^2/2 \right\},$$

which is a null-recurrent class discussed in Höpfner and Kutoyants (2003). Then it follows that, for a constant $\alpha = 1/2 - \mu/\sigma^2 \in (0, 1)$,

$$\frac{1}{T_n^\alpha} \int_0^{T_n} U(X_s^\vartheta)^2 \, \mathrm{d}s \xrightarrow{P_{\vartheta_0}} K_{\alpha,U} W_1^\alpha, \quad T_n \to \infty,$$

where $K_{\alpha,U}$ is a constant depending on α and U, and W^α is a Mittag–Leffler process with index α. Hence

$$\mathcal{D}_{\mathrm{NR}} \subset \mathcal{D}_0(\beta_{n,k} = T_n^\alpha, \gamma_{n,k} = T_n),$$

while $\liminf_{n\to\infty} T_n > 0$.

Example 3 (OU processes) Consider the following subclass of $\mathcal{D}(U; V; \Theta)$:

$$\mathcal{D}_{\mathrm{OU}^+} := \left\{X^\vartheta \in \mathcal{D}(U; V; \Theta) \,|\, U(x) = x, V(x) = 1, \mu > 0\right\},$$

which is a family of non-recurrent Ornstein–Uhlenbeck (OU) processes. As is well known,

$$\frac{1}{e^{2\mu T_n}} \int_0^{T_n} |X_s^\vartheta|^2 \, \mathrm{d}s \to \frac{(X_0 + Z_\vartheta)^2}{2\mu\sigma} \quad a.s., \quad T_n \to \infty,$$

where $Z_\vartheta := \lim_{t\to\infty} e^{-\mu t} X_t \;\; a.s. \sim \mathcal{N}(0, \sigma(2\mu)^{-1})$; see e.g. Feigin (1976). Therefore, we see that

$$\mathcal{D}_{\mathrm{OU}^+} \subset \mathcal{D}_0(\beta_{n,0} = e^{2\mu T_n}, \gamma_{n,0} = T_n),$$

while $\liminf_{n\to\infty} T_n > 0$. Moreover, if we choose $C(x) = \sqrt{1 + x^2}$ then it also follows that

$$\mathcal{D}_{\mathrm{OU}^+} \subset \mathcal{D}_2(\beta_{n,2} = T_n, \gamma_{n,2} = 1),$$

since $\int_0^{T_n} |X_s^\vartheta|^2 (1 + |X_s^\vartheta|^2)^{-1} \, \mathrm{d}s = O(T_n)$ and $\int_0^{T_n} (1 + |X_s^\vartheta|^2)^{-1} \, \mathrm{d}s = O(1)$ a.s.; here we note that $X_t^\vartheta \sim^{a.s.} e^{\mu t} Z_\vartheta$ as $t \to \infty$. On the other hand, it follows for any $k \geq 0$ that

$$\mathcal{D}_{\mathrm{OU}^-} := \left\{X^\vartheta \in \mathcal{D}(U; V; \Theta) \,|\, U(x) = x, V(x) = 1, \mu < 0\right\} \subset \mathcal{D}_k(T_n, T_n),$$

since $\mathcal{D}_{\mathrm{OU}^-}$ is a subclass of the ergodic class with higher order moments.

Example 4 (Exponential growth models) An extension of $\mathcal{D}_{\mathrm{OU}^+}$ is the following class.

$$\mathcal{D}_{\mathrm{Exp}} := \left\{X^\vartheta \in \mathcal{D}(U; V; \Theta) \,|\, U(x) = x + r(x), V(x) = 1, \mu > 0\right\},$$

where r is a Lipschitz function with growth condition $|r(x)| \leq K(1+|x|^\kappa)$ for constants $K \geq 0$ and $\kappa \in [0, 1)$. In this case,

$$\mathcal{D}_{\text{Exp}} \subset \mathcal{D}_0(\beta_{n,0} = e^{2\mu T_n}, \gamma_{n,0} = T_n),$$

since, for a random variable $\eta_\mu := \lim_{t\to\infty} e^{-\mu t} X_t \; a.s.$,

$$\frac{1}{e^{2\mu T_n}} \int_0^{T_n} U^2(X_s)\,\mathrm{d}s \to \frac{(X_0 + \eta_\vartheta^2)}{2\mu\sigma} \quad a.s., \quad T_n \to \infty;$$

see Dietz and Kutoyants (2003) for details. Moreover, we see as in $\mathcal{D}_{OU^+}$-case that

$$\mathcal{D}_{\text{Exp}} \subset \mathcal{D}_2(\beta_{n,2} = T_n, \gamma_{n,2} = 1),$$

for $C(x) = \sqrt{1+|x|^2}$ while $\liminf_{n\to\infty} T_n > 0$.

Example 5 (Polynomial growth models) Let

$$\mathcal{D}_{\text{Po}} := \left\{ X^\vartheta \in \mathcal{D}(U; V; \Theta) \mid U(x) = |x|^\kappa, \; \kappa \in (0, 1) \text{ and } V(x) = 1 \right\}.$$

Since it follows for $\gamma = (1+\kappa)/(1-\kappa)$ that

$$\frac{1}{T_n^\gamma} \int_0^{T_n} |X_s^\vartheta|^{2\kappa}\,\mathrm{d}t \to \frac{(1-\kappa)^\gamma}{(1+\kappa)} \left(\frac{\mu^2}{\sigma}\right)^{\kappa/(1-\kappa)} \quad a.s., \quad T_n \to \infty;$$

from Kutoyants (2004) Sect. 3.5, it follows that

$$\mathcal{D}_{\text{Po}} \subset \mathcal{D}_0(\beta_{n,0} = T_n^\gamma, \gamma_{n,0} = T_n).$$

Taking $C(x) = \sqrt{1+|x|^{2\kappa}}$, we see that $\int_0^{T_n} S(X_s^\vartheta) C^{-2}(X_s^\vartheta)\,\mathrm{d}s = O(T_n) \; a.s.$ Moreover, noticing that

$$X_t \sim^{a.s.} c_0 t^{\frac{1}{1-\kappa}},$$

where $c_0 = [(1-\kappa)\mu_0]^{1/(1-\kappa)}$, and that $\int_0^T (1+s^\alpha)^{-1}\,\mathrm{d}s = O(T^{1-\alpha})$ for $\alpha \in (0, 1)$, we see that

$$\int_0^{T_n} \frac{\mathrm{d}s}{C^2(X_s)} \sim^{a.s.} \int_0^{T_n} \frac{\mathrm{d}s}{1+s^{\frac{2\kappa}{1-\kappa}}} = O\left(T_n^{\gamma'}\right), \quad \gamma' = \frac{1-3\kappa}{1-\kappa},$$

if $2\kappa/(1-\kappa) < 1 \Leftrightarrow \kappa \in (0, 1/3)$. Therefore, it follows that

$$\mathcal{D}_{Po}^\vartheta \subset \mathcal{D}_2^\vartheta(\beta_{n,2} = T_n, \gamma_{n,2} = T_n^{\gamma'}),$$

for $C(x) = \sqrt{1+|x|^2}$ and $\kappa \in (0, 1/3)$.

3 M-estimator and the asymptotic properties

If $X^\vartheta \in \mathcal{D}_k(\beta_{n,k}, \gamma_{n,k})$ then, as is described in Sect. 1, the natural rate of convergence of estimators for ϑ_0 would be $(\sqrt{\beta_{n,k}}, \sqrt{\gamma_{n,k}/h_n})$. In this section, we shall propose an M-estimator for ϑ_0 that attains such a rate.

We consider the following quadratic-type contrast functions; cf. Shimizu (2009b):

$$Q_{n,k}(\vartheta) := \sum_{i=1}^{n} \frac{1}{C_{i-1}^k} \left[\frac{\left(\Delta_i^n X - \kappa_n(\mu) U_{i-1}\right)^2}{h_n \sigma V_{i-1}^2} + \log \sigma \right], \tag{8}$$

where $G_{i-1} := G(X_{t_{i-1}^n})$ for any function G, and define the M-estimator $\hat{\vartheta}_{n,k} = (\hat{\mu}_{n,k}, \hat{\sigma}_{n,k})$ as

$$\hat{\vartheta}_{n,k} = \arg \inf_{\vartheta \in \Theta} Q_{n,k}(\vartheta). \tag{9}$$

If κ_n has an inverse for each n, then $\hat{\vartheta}_{n,k}$ can be written in an explicit form such as

$$\hat{\mu}_{n,k} = \kappa_n^{-1} \circ \left[\left(\sum_{i=1}^{n} \frac{U_{i-1} \Delta_i^n X}{C_{i-1}^k V_{i-1}^2} \right) \left(\sum_{i=1}^{n} \frac{U_{i-1}^2}{C_{i-1}^k V_{i-1}^2} \right)^{-1} \right]; \tag{10}$$

$$\hat{\sigma}_{n,k} = \left(\sum_{i=1}^{n} \frac{\left(\Delta_i^n X - \kappa_n(\hat{\mu}_{n,k}) U_{i-1}\right)^2}{h_n C_{i-1}^k V_{i-1}^2} \right) \left(\sum_{i=1}^{n} \frac{1}{C_{i-1}^k} \right)^{-1}. \tag{11}$$

When we discuss the asymptotic properties under the assumption that $X \in \mathcal{D}_k(\beta_{n,k}, \gamma_{n,k})$, we consider the following two cases separately: As $n \to \infty$,

Case 1: $\beta_{n,k}/\gamma_{n,k} = O(1)$;
Case 2: $\beta_{n,k}/\gamma_{n,k} \to \infty$.

Case 1 includes the case of $\mathcal{D}_{\mathrm{Erg}}$ and a null-recurrent case $\mathcal{D}_{\mathrm{NR}}$, where the rate for $\hat{\sigma}_{n,k}$ will be more rapid than the one for $\hat{\mu}_{n,k}$, and μ_0 should be estimated in first.

Case 2 includes non-recurrent cases such as $\mathcal{D}_{\mathrm{OU}^+}$, $\mathcal{D}_{\mathrm{Exp}}$ and $\mathcal{D}_{\mathrm{Po}}$, where the rate for $\hat{\mu}_{n,k}$ will be more rapid than the one for $\hat{\sigma}_{n,k}$, and σ_0 should be estimated in first.

The case where $T_n \to \infty$ is important since the joint estimation of (μ_0, σ_0) can be possible. We shall also add a remark describing what the asymptotic distribution should be if the joint convergence holds for the *score vector* $\nabla_\vartheta Q_{n,k}(\vartheta_0)$ and the *normalized information* $(\mathcal{I}_{n,k}, \mathcal{J}_{n,k})$; see Corollaries 1 and 2.

3.1 Assumptions

Let $\kappa_n(\mu)$ be a twice differentiable function in μ, satisfying the following assumptions.

Assumption 1 $\sup_{\mu \in \Xi} |\partial_\mu^j \kappa_n(\mu) - \partial_\mu^j (h_n \mu)| = O(h_n^2)$; *$j = 0, 1, 2$, as $n \to \infty$.*

Assumption 2 *There exist some functions $R_j(x);\ j = 1, 2, 4$, such that*

(i) $\left|E_{\vartheta_0}\left[\Delta_i^n X - \kappa_n(\mu_0)U_{i-1}\middle|X_{t_{i-1}^n}\right]\right| \lesssim R_{1,i-1}h_n^{3/2}$;

(ii) $\left|E_{\vartheta_0}\left[\left(\Delta_i^n X - \kappa_n(\mu_0)U_{i-1}\right)^2\middle|X_{t_{i-1}^n}\right] - h_n\sigma_0 V_{i-1}^2\right| \lesssim R_{2,i-1}h_n^2$;

(iii) $E_{\vartheta_0}\left[\left|\Delta_i^n X - \kappa_n(\mu_0)U_{i-1}\right|^4\middle|X_{t_{i-1}^n}\right] \lesssim R_{4,i-1}h_n^2$,

where $R_{j,i-1} := R_j(X_{t_{i-1}^n})$.

Moreover, we assume some regularities on functions R_j and C^k:

Assumption (*H*[*k*])

($H1$) $|R_1UC^{-k}|$ *is bounded;*
($H2$) $|R_2C^{-k}|$ *is bounded;*
($H3$) $|R_4C^{-2k}|$ *is bounded.*

Remark 1 The simplest choice of κ_n for $X^\vartheta \in \mathcal{D}(U; V; \Theta)$ is $\kappa_n(\mu) = \mu h_n$. Then Assumption 2 is satisfied with $R_j(x) = C^j(x)$ if $U, V \in C^2$ and the derivatives are uniformly bounded: see Jacod (2006), Sect. 3.1. In this case, Assumption ($H[2]$) holds true. Furthermore, if U and V are also uniformly bounded, then all R_j's become constants since the above C can be taken as a constant; e.g., $\mathcal{D}_{\mathrm{NR}}$-class, and Assumption ($H[0]$) holds true.

Remark 2 For $\mathcal{D}_{\mathrm{OU}^+}$, a choice $\kappa_n(\mu) = e^{\mu h_n} - 1$ is useful; see Sect. 4.3. One can calculate the conditional expectations in Assumption 2 (i)–(iii) directly since, in this case,

$$\Delta_i^n X - \kappa_n(\mu_0)U_{i-1} = X_{t_i^n} - e^{\mu_0 h_n}X_{t_{i-1}^n} \sim \mathcal{N}\left(0, \frac{\sigma_0(e^{2\mu_0} - 1)}{2\mu_0}\right),$$

which is independent of $X_{t_{i-1}^n}$. As a result, R_j's become constants, especially $R_1 \equiv 0$, and Assumption ($H[0]$) holds true.

We further require a situation that the following convergences hold true:

Assumption (LLN)

($LLN1$) $\displaystyle \frac{h_n}{\gamma_{n,k}}\sum_{i=1}^n \frac{1}{C_{i-1}^k} - \mathcal{J}_{n,k} \xrightarrow{P_{\vartheta_0}} 0$;

($LLN2$) $\displaystyle \frac{h_n}{\beta_{n,k}}\sum_{i=1}^n \frac{S_{i-1}}{C_{i-1}^k} - \mathcal{I}_{n,k} \xrightarrow{P_{\vartheta_0}} 0.$

Note that, if $\mathcal{J}_{n,k} \xrightarrow{P_{\vartheta_0}} J_k$ and $\mathcal{I}_{n,k} \xrightarrow{P_{\vartheta_0}} I_k$ for some positive tight variable I_k and J_k as in Examples 2–5, then (LLN1,2) are, so-called, the law of large numbers. However, we do not require the existence of the limits I_k and J_k. Although we often assume directly that the above (LLN1,2) hold true below, we can give some sufficient (but not necessary) conditions that ensure those convergences: see Appendix A for details.

3.2 Results for Case 1

Theorem 1 *Assume that* $X \in \mathcal{D}_k(\beta_{n,k}, \gamma_{n,k})$ *with* $\beta_{n,k}/\gamma_{n,k} = O(1)$. *Moreover, suppose Assumptions* 1, 2*,* $(H[k])$*, (LLN), and that*

$$nh_n^2\gamma_{n,k}^{-1} \to 0; \quad h_n \to 0, \tag{12}$$

as $n \to \infty$. *Then* $\hat{\sigma}_{n,k} \xrightarrow{P_{\vartheta_0}} \sigma_0$. *In addition, suppose that*

$$\beta_{n,k}^{-1}\left(1 + nh_n^{3/2}\right) \to 0. \tag{13}$$

Then $\hat{\mu}_{n,k} \xrightarrow{P_{\vartheta_0}} \mu_0$.

Remark 3 Note that Case 1 also includes the case where T_n is fixed, say $T_n \equiv 1$ without loss of generality. Indeed, $X \in \mathcal{D}_0(1, 1)$ while $\int_0^1 S(X_s)\,ds > 0$ *a.s.* In this case, the consistent estimates for σ_0 is possible. We remark that, in that case, we need not assume the tightness of $\mathcal{I}_{n,0}^{-1}$ to estimate σ_0.

Remark 4 It is required that $\beta_{n,k} \to \infty$ in (13) to estimate μ_0. In this case, $\mathcal{I}_{n,k} \xrightarrow{P_{\vartheta_0}} 0$ if $\limsup_{n\to\infty} T_n < \infty$. Therefore, it should be that $T_n \to \infty$ to ensure the tightness of $\mathcal{I}_{n,k}^{-1}$.

Theorem 2 *Suppose the same assumptions as in Theorem* 1*, and that*

$$nh_n^{3/2}\beta_{n,k}^{-1/2} \to 0; \tag{14}$$

$$nh_n\gamma_{n,k}^{-1} = O(1). \tag{15}$$

Then the sequence $\left(\sqrt{\beta_{n,k}}(\hat{\mu}_{n,k} - \mu_0), \sqrt{\gamma_{n,k}/h_n}(\hat{\sigma}_{n,k} - \sigma_0)\right)$ *is tight.*

Remark 5 We notice that Condition (14) does not need to show the tightness of the sequence $\sqrt{\gamma_{n,k}/h_n}(\hat{\sigma}_{n,k} - \sigma_0)$. In the light of Remark 3, the sequence $\sqrt{n}(\hat{\sigma}_{n,k} - \sigma_0)$ is tight even when T_n is fixed since (15) is satisfied for $\gamma_{n,k} = T_n$, where $T_n/h_n = n$. This is consistent with the well known result.

The following corollary is clear from the proof of Theorem 2.

Corollary 1 *Suppose the same assumptions as in Theorem* 2*, and that the following joint convergence holds true:*

$$\left(\frac{1}{\sqrt{\beta_{n,k}}}\partial_\mu Q_{n,k}(\vartheta_0), \sqrt{\frac{h_n}{\gamma_{n,k}}}\partial_\sigma Q_{n,k}(\vartheta_0), \mathcal{I}_{n,k}, \mathcal{J}_{n,k}\right) \xrightarrow{\mathcal{D}} (L_1, L_2, I_k, J_k)\,.$$

Then

$$\left(\sqrt{\beta_{n,k}}(\hat{\mu}_{n,k} - \mu_0), \sqrt{\gamma_{n,k}/h_n}(\hat{\sigma}_{n,k} - \sigma_0)\right) \xrightarrow{\mathcal{D}} \left(\frac{\sigma_0}{2}I_k^{-1}L_1, \sigma_0^2 J_k^{-1}L_2\right).$$

3.3 Results for Case 2

Theorem 3 *Assume that* $X \in \mathcal{D}_k(\beta_{n,k}, \gamma_{n,k})$ *with* $\beta_{n,k}/\gamma_{n,k} \to \infty$. *Moreover, suppose Assumptions* 1, 2, $(H[k])$ *and (LLN2), and that*

$$nh_n^{3/2}\beta_{n,k}^{-1} \to 0; \quad h_n \to 0. \tag{16}$$

Then $\hat{\mu}_{n,k} \xrightarrow{P_{\vartheta_0}} \mu_0$. *Moreover, the sequence* $\sqrt{\beta_{n,k}}(\hat{\mu}_{n,k} - \mu_0)$ *is tight.*

Theorem 4 *Assume that* $X \in \mathcal{D}_k(\beta_{n,k}, \gamma_{n,k})$ *with* $\beta_{n,k}/\gamma_{n,k} \to \infty$. *Moreover, suppose Assumptions* 1, 2, $(H[k])$, *(LLN) and that the sequence* $\sqrt{\beta_{n,k}}(\hat{\mu}_{n,k} - \mu_0)$ *is tight. Furthermore, suppose that*

$$h_n\gamma_{n,k}^{-1} \to 0; \quad nh_n^{3/2}\gamma_{n,k}^{-1/2} = O(1). \tag{17}$$

Then $\hat{\sigma}_{n,k} \xrightarrow{P_{\vartheta_0}} \sigma_0$. *Moreover, the sequence* $\sqrt{\gamma_{n,k}/h_n}(\hat{\sigma}_{n,k} - \sigma_0)$ *is tight.*

Corollary 2 *Suppose the same assumptions as in Theorem 4, and that the following joint convergence holds true:*

$$\left(\frac{1}{\sqrt{\beta_{n,k}}}\partial_\mu Q_{n,k}(\vartheta_0), \sqrt{\frac{h_n}{\gamma_{n,k}}}\partial_\sigma Q_{n,k}(\vartheta_0), \mathcal{I}_{n,k}, \mathcal{J}_{n,k}\right) \xrightarrow{\mathcal{D}} (L_1, L_2, I_k, J_k)\,.$$

Then

$$\left(\sqrt{\beta_{n,k}}(\hat{\mu}_{n,k} - \mu_0), \sqrt{\gamma_{n,k}/h_n}(\hat{\sigma}_{n,k} - \sigma_0)\right) \xrightarrow{\mathcal{D}} \left(\frac{\sigma_0}{2}I_k^{-1}L_1, \sigma_0^2 J_k^{-1}L_2\right).$$

4 Examples

Examples in this section are the continuation of Examples 1–5 described in Sect. 2. We shall propose some explicit forms of $\kappa_n(\mu)$ and the value of $k \geq 0$, and study the asymptotic properties of those estimators. In Sect. 4.6, we describe a relationship between the TFE and our proposal estimator.

Throughout this section, we consider the case where

$$h_n \to 0 \quad \text{and} \quad T_n \to \infty,$$

as $n \to \infty$ for the joint estimation of μ_0 and σ_0.

4.1 Ergodic models without moments

Assume that $X \in \mathcal{D}_{\mathrm{Erg}}$ with an invariant measure π, and that $U, V \in C^2(\mathbb{R})$ satisfying

$$|U(x)| + |V(x)| \lesssim 1 + |x|^{c_0}, \quad c_0 > 0,$$

with bounded derivatives. Here, we do not assume any moment condition on X.

In this case, we can choose $C(x) = 1 + |x|^{c_0}$ for some $c_0 > 0$. Then

$$\mathcal{D}_{\mathrm{Erg}} \subset \mathcal{D}_2(\beta_{n,2} = T_n, \gamma_{n,2} = T_n).$$

Therefore, this is Case 1.

- When $\kappa_n(\mu) = \mu h_n$ then Assumptions 1 and 2 are satisfied with $R_j(x) = C^j(x)$.
- Assumption $(H[2])$ is satisfied.
- Assumption (LLN) hold by Corollaries 3 and 4 if $h_n \to 0$:

$$\mathcal{I}_{n,2} \xrightarrow{P_{\vartheta_0}} \int_{\mathbb{R}} \frac{S(x)}{C^2(x)} \pi(\mathrm{d}x) =: I_2; \quad \mathcal{J}_{n,2} \xrightarrow{P_{\vartheta_0}} \int_{\mathbb{R}} \frac{1}{C^2(x)} \pi(\mathrm{d}x) =: J_2,$$

  the limits of which are constants.
- Conditions (12)–(15) in Theorems 1 and 2 are satisfied if $nh_n^2 \to 0$.

Hence we see that $\left(\sqrt{T_n}(\hat{\mu}_{n,2} - \mu_0), \sqrt{n}(\hat{\sigma}_{n,2} - \sigma_0)\right)$ is tight. Moreover, by Corollary 1,

$$\left(\sqrt{T_n}(\hat{\mu}_{n,2} - \mu_0), \sqrt{n}(\hat{\sigma}_{n,2} - \sigma_0)\right) \xrightarrow{\mathcal{D}} \mathcal{N}_2\left(0, \mathrm{diag}\left(\sigma_0^2 I_2^{-1}, 2\sigma_0^2 J_2^{-1}\right)\right).$$

4.2 Null-recurrent models

Assume that $X \in \mathcal{D}_{\mathrm{NR}}$. Then it follows that

$$\mathcal{D}_{\mathrm{NR}} \subset \mathcal{D}_0(\beta_{n,k} = T_n^{\alpha}, \gamma_{n,k} = T_n), \quad \alpha := 1/2 - \mu_0/\sigma_0^2 \in (0, 1).$$

Therefore, this is Case 1. We see the following:

- When $\kappa_n(\mu) = \mu h_n$ then Assumptions 1 and 2 are satisfied with $R_j(x)$'s being constants.
- Assumption $(H[0])$ is satisfied; see Remark 1.
- Convergences (LLN1,2) hold if

$$nh_n^{3/2}\beta_{n,k}^{-1} = h_n^{1/2}T_n^{1-\alpha} \to 0; \tag{18}$$

  see Remark 8.
- Conditions (12) and (13) in Theorem 1 are satisfied under (18).

Therefore, it follows from (18) that $\hat{\vartheta}_{n,0} \to \vartheta_0$ if

$$nh_n^{\frac{3}{2}+\frac{\alpha}{2(1-\alpha)}} \to 0.$$

For the tightness of $D_n(\hat{\vartheta}_{n,0} - \vartheta_0)$ with $D_n = \mathrm{diag}(T_n^{\alpha/2}, \sqrt{n})$, we need a further condition

$$nh_n^{1+\frac{1}{2-\alpha}} \to 0.$$

The rate D_n is the best attainable since $T_n^{\alpha/2}$ is consistent with the rate of the MLE based on continuous observations which is asymptotically efficient in the asymptotic minimax sense; see Höpfner and Kutoyants (2003) for details.

4.3 Non-recurrent OU processes

Assume that $X \in \mathcal{D}_{\mathrm{OU}^+}$. As in Example 3, it follows that

$$\mathcal{D}_{OU^+} \subset \mathcal{D}_0(\beta_{n,k} = e^{2\mu_0 T_n}, \gamma_{n,k} = T_n).$$

Therefore, this is Case 2. In this case, a naive choice $\kappa_n(\mu) = \mu h_n$ is not suitable since Assumption $(H[0])$, e.g., (H1), is not satisfied. Actually, the M-estimator for μ_0 via the above choice is $\sqrt{T_n}$-consistent but not $e^{\mu_0 T_n}$-consistent which is the optimal rate of convergence in this model. However, it implies that we can obtain at least $(\sqrt{T_n}, \sqrt{n})$-consistent estimator by such a naive choice even if the sign of μ is unknown (ergodic, or non-ergodic). See Shimizu (2009a), Theorems 1–3 for details.

To obtain $e^{\mu_0 T_n}$-consistent estimator for μ_0, the TFE proposed by Kasonga (1990) is one of the candidates. We can obtain the TFE-type estimator in this model by setting $\kappa_n(\mu) = e^{\mu h_n} - 1$; see also Sect. 4.6. Then we see the following:

- If $\kappa_n(\mu) = e^{\mu h_n} - 1$ then Assumptions 1 and 2 are satisfied with $R_1(x) \equiv 0$ and R_2, R_4 are constants.
- Assumption $(H[0])$ is satisfied; see Remark 2.
- Convergences (LLN1,2) hold true if $h_n \to 0$: see Remark 9. In particular,

$$\mathcal{I}_{n,0} \xrightarrow{P_{\vartheta_0}} \frac{(X_0 + Z)^2}{2\mu_0\sigma_0} := I_0; \quad \mathcal{J}_{n,0} \equiv 1,$$

  where $Z \sim \mathcal{N}(0, \sigma_0(2\mu_0)^{-1})$.
- Condition (16): $e^{\mu_0 T_n}/n \to \infty$ and $T_n e^{-\mu_0 T_n} \to 0$, which are satisfied if there exists some $\delta > 0$ such that $nh_n^{1+\delta} \to \infty$, which is a mild condition in practice.

Hence it follows from Theorem 3 that $e^{\mu_0 T_n}(\hat{\mu}_{n,0} - \mu_0)$ is tight if $nh_n^{1+\delta} \to \infty$ for some $\delta > 0$. Moreover, if $nh_n^2 = O(1)$ then

- Condition (17) is satisfied since $nh_n^{3/2}\gamma_{n,k}^{-1/2} = \sqrt{nh_n^2} = O(1)$,

which implies that $\sqrt{n}(\hat{\sigma}_{n,0} - \sigma_0)$ is tight by Theorem 4. Therefore, the joint estimation of (μ_0, σ_0) is possible. Note that $\hat{\vartheta}_{n,k}$ can be written in explicit form by (10) since $\kappa_n^{-1}(\mu) = h_n^{-1}\log(1 + \mu)$.

According to Shimizu (2009c), Corollary 1 type result is also obtained, and the asymptotic distribution becomes a mixture of normal:

$$\left(e^{\mu_0 T_n}(\hat{\mu}_{n,0} - \mu_0), \sqrt{n}(\hat{\sigma}_{n,0} - \sigma_0)\right) \xrightarrow{\mathcal{D}} \mathcal{N}_2\left(0, \mathrm{diag}(I_0^{-1}, 2\sigma_0^2)\right).$$

Furthermore, $\hat{\vartheta}_{n,0}$ becomes asymptotically efficient in the sense of *maximum concentration probability: Wolfowitz's-sense*.

Remark 6 In Jacod (2006), or Shimizu (2009b), the identifiability conditions described in (6), or (7) are invalid for joint estimation of (μ_0, σ_0); see the comment for OU processes in the end of Sect. 2, p. 388 in Jacod (2006), or Remark 4.2 in Shimizu (2009b). This is because the normalization rate for $I_{n,k}$ and $J_{n,k}$ is T_n in both cases. However, in this paper, we admit more flexible rate of normalization according to k: $\beta_{n,k}$ and $\gamma_{n,k}$, which enables us the joint estimation.

Remark 7 Condition (17) is actually redundant for the tightness of $\sqrt{n}(\hat{\sigma}_{n,k} - \sigma_0)$ in the case where $X \in \mathcal{D}_{OU^+}$ since the condition is used in the proof of Theorem 4 to show, e.g., the convergence in (47): $\sum_{i=1}^n |E_{\vartheta_0}[p_i^n | X_{t_{i-1}^n}]| \xrightarrow{P_{\vartheta_0}} 0$. However, in OU cases with $\kappa_n(\mu) = e^{\mu T_n} - 1$, $\sum_{i=1}^n |E_{\vartheta_0}[p_i^n | X_{t_{i-1}^n}]| = 0$, exactly.

4.4 Exponential growth models

Assume that $X \in \mathcal{D}_{\text{Exp}}$, which is a general class including $\mathcal{D}_{\text{OU}^+}$. Moreover, we assume that $r \in C^2$ with bounded derivatives. Note from Example 4 that

$$\mathcal{D}_{\text{Exp}} \subset \mathcal{D}_0(e^{2\mu_0 T_n}, T_n) \cap \mathcal{D}_2(T_n, 1),$$

if we choose $C(x) = \sqrt{1+x^2}$. In this case, we could not take a *good* $\kappa_n(\mu)$ for which $R_1(x) \equiv 0$ and R_2, R_4 are constants as in the case where $X \in \mathcal{D}_{\text{OU}^+}$. Therefore, Assumptions (H1) with $k = 0$ may collapse. Now, we shall regard that

$$\mathcal{D}_{\text{Exp}} \subset \mathcal{D}_2(T_n, 1).$$

Then this is Case 2, and we see the following:

- For $\kappa_n(\mu) = \mu h_n$, Assumptions 1 and 2 are satisfied with $R_j(x) = C^j(x)$.
- Assumption $(H[2])$ is satisfied.
- (LLN1) is checked by an easy calculation; (LLN2) is also be checked by Corollary 4.
- Condition (16) is clear: $nh_n^{3/2}\beta_{n,k}^{-1} = \sqrt{h_n} \to 0$.
- Condition (17) is satisfied if $nh_n^{3/2} = O(1)$.

Therefore, if $nh_n^{3/2} = O(1)$, then

$$\left(\sqrt{T_n}(\hat{\mu}_{n,2} - \mu_0), \sqrt{1/h_n}(\hat{\sigma}_{n,2} - \sigma_0)\right) \text{ is tight.}$$

4.5 Polynomial growth models

Assume that $X \in \mathcal{D}_{\text{Po}}$ with $U(x) = |x|^\kappa$, $\kappa \in (0, 1/3)$. In this case, as in the exponential growth model, $k = 0$ would not be suitable for our purpose. However, we know by Example 5 that it also holds that, for $C(x) = \sqrt{1+|x|^2}$,

$$\mathcal{D}_{\text{Po}} \subset \mathcal{D}_2(\beta_{n,k} = T_n, \gamma_{n,k} = T_n^{\gamma'}), \quad \gamma' = \frac{1-3\kappa}{1-\kappa} \in (0, 1).$$

Therefore, this is Case 2, and we see the following:

- For $\kappa_n(\mu) = \mu h_n$, Assumptions 1 and 2 are satisfied with $R_j(x) = C^j(x)$.
- Assumption $(H[2])$ is satisfied.
- Assumption (LLN) is checked by Corollaries 3 and 4 with slight modifications.
- Condition (16) is clear: $nh_n^{3/2}\beta_{n,k}^{-1} = \sqrt{h_n} \to 0$.
- Condition (17) is satisfied if $nh_n^{1+\frac{1}{2-\gamma'}} = O(1)$.

Therefore, if $\kappa \in (0, 1/5)$ and $nh_n^{1+\frac{1}{2-\gamma'}} = O(1)$, then

$$\left(\sqrt{T_n}(\hat{\mu}_{n,2} - \mu_0), \sqrt{T_n^{\gamma'}/h_n}(\hat{\sigma}_{n,2} - \sigma_0)\right) \quad \textit{is tight.}$$

4.6 The TFE and our contrast function

The trajectory-fitting estimator (TFE) based on discrete sample was introduced by Kasonga (1990) in the context of drift estimation for non-ergodic diffusions. When we can find the explicit solution to the following ordinary differential equations:

$$\mathrm{d}x_t^\mu = \mu U(x_t^\mu)\,\mathrm{d}t, \quad x_{t_{i-1}^n}^\mu = X_{t_{i-1}^n}, \tag{19}$$

for each $i = 1, \dots, n$, the contrast function is given by

$$\Psi_n^{\mathrm{TFE}}(\vartheta) = \sum_{i=1}^n \left[\frac{|X_{t_i^n} - x_{t_i^n}^\mu|^2}{h_n \sigma V_{i-1}^2} + \log\sigma \right].$$

This is an extended version of Kasonga (1990)'s contrast functions for joint estimation of (μ, σ).

The TFE can be a candidate that has a *good* rate of convergence. Actually, in the case of $\mathcal{D}_{OU+}$, our contrast function with $\kappa_n(\mu) = e^{\mu h_n} - 1$ gives the TFE, which is asymptotically efficient as is described in Sect. 4.3.

In the case of $\mathcal{D}_{\mathrm{Po}}$, where $U(x) = |x|^\kappa$, (19) is also explicitly solved such that

$$|x_t^\mu|^{1-\kappa} = \mathrm{sgn}(X_{t_{i-1}^n})\mu(1-\kappa)(t - t_{i-1}^n) + |X_{t_{i-1}^n}|^{1-\kappa}, \tag{20}$$

in a neighborhood of $t = t_{i-1}^n$. Then the TFE for μ is expected to have more rapid rate than $\sqrt{T_n}$ given in Sect. 4.5 although one can not get $\hat{\mu}$ in explicit form. However, the contrast Ψ_n^{TFE} with (20) does not belong to the class of our contrast functions at the present form, and it seems difficult to find a suitable choice of $\kappa_n(\mu)$ by which $Q_{n,k}$ is *close* to Ψ_n^{TFE}. For example, assuming for simplicity that $X_{t_{i-1}^n} > 0$ for each i, we see from (20) that

$$x_{t_i^n}^\mu = X_{t_{i-1}^n}\left[1 + \mu(1-\kappa)h_n X_{t_{i-1}^n}^{1-\kappa}\right]^{\frac{1}{1-\kappa}} \sim X_{t_{i-1}^n} + \mu h_n X_{t_{i-1}^n}^\kappa, \quad h_n \to 0,$$

which just leads us a naive $\kappa_n(\mu) = \mu h_n$ as in Sect. 4.5. This consideration gives us an insight that it seems difficult to find some $\kappa_n(\mu)$ that gives a better rate for $\hat{\mu}$ than $\sqrt{T_n}$ in our class of contrasts. Research of the convergence rate of the TFE is a separate issue.

As in the case of $\mathcal{D}_{\mathrm{Exp}}$, it would be difficult to solve (19) in general. When we could not find the solution explicitly, a simple approximation of the solution, say $\tilde{x}_t^\mu$, is as follows.

$$\tilde{x}_t^\mu = X_{t_{i-1}^n} + \mu U_{i-1}(t - t_{i-1}^n),$$

which yields a version of our contrast function $Q_{n,0}$ with a simple choice $\kappa_n(\mu) = \mu h_n$. As a result, we need weights, e.g., C_{i-1}^{-2}'s, to obtain some limit theorems. That is why, we naturally obtain our contrast function $Q_{n,k}$. The case of $\mathcal{D}_{\mathrm{OU+}}$ is a rare example where we can choose an explicit $\kappa_n(\mu)$ yielding the TFE.

5 Concluding remarks

We investigated parametric inference for discretely observed diffusion processes which may not be ergodic (positive recurrent), but may be null- or non-recurrent. To specify the rate for information, we restricted the class of diffusions to $\mathcal{D}_k(\beta_{n,k}, \gamma_{n,k})$ defined in Definition 1, where $(\beta_{n,k}, \gamma_{n,k})$ is the rate for some *normalized information*. We proposed a simple contrast function, which is due to a (*weighted*) *local-Gauss approximation* for the likelihood function, and gave a pair of sufficient conditions depending on $k \geq 0$ to ensure that a proposed estimator $\hat{\vartheta}_{n,k} = (\hat{\mu}_{n,k}, \hat{\sigma}_{n,k})$ becomes $(\sqrt{\beta_{n,k}}, \sqrt{\gamma_{n,k}/h_n})$-consistent:

$$\left(\sqrt{\beta_{n,k}}(\hat{\mu}_{n,k} - \mu_0), \sqrt{\gamma_{n,k}/h_n}(\hat{\sigma}_{n,k} - \sigma_0)\right) \quad \textit{is tight.}$$

If the conditions hold true for $k = 0$, the rate can be *the best attainable (rate-optimal)*. Indeed, we gave, in the previous section, some concrete examples where the estimator becomes rate-optimal. Otherwise, we can often check a condition for some $k > 0$, and can find an estimator the rate of convergence of which is at least $(\sqrt{\beta_{n,k}}, \sqrt{\gamma_{n,k}/h_n})$ for some $k > 0$ though it may not be optimal. Obtaining more sharp results: to present a contrast function which always gives a rate-optimal estimator, is an open problem.

In the paper, we considered only 1-dimensional, and a special form of stochastic differential equation as in (1) for simplicity of notation and discussion. As is commented in Introduction, we can extend the result to more general cases. Now, suppose that X is a 1-dimensional diffusion process satisfying

$$\mathrm{d}X_t^\vartheta = U(X_t^\vartheta, \mu)\,\mathrm{d}t + V(X_t^\vartheta, \sigma)\,\mathrm{d}W_t \tag{21}$$

where U and V are measurable functions defined on $\mathbb{R} \times \Xi$ and $\mathbb{R} \times \Pi$, respectively. Then it would be easy to imagine due to the argument as in Shimizu (2009b) that we should consider the class $\tilde{\mathcal{D}}_k(\beta_{n,k}, \gamma_{n,k})$ such that

$$\tilde{\mathcal{I}}_{n,k}(\mu_0) + \tilde{\mathcal{J}}_{n,k}(\sigma_0) = O_{P_{\vartheta_0}}(1); \tag{22}$$

$$\lim_{\eta\to 0} \limsup_{n\to\infty} P_{\vartheta_0}\left\{\inf_{\mu:|\mu-\mu_0|>\varepsilon}\left\{\tilde{\mathcal{I}}_{n,k}(\mu)\right\} \le \eta\right\} = 0; \tag{23}$$

$$\lim_{\eta\to 0} \limsup_{n\to\infty} P_{\vartheta_0}\left\{\inf_{\sigma:|\sigma-\sigma_0|>\varepsilon}\left\{\tilde{\mathcal{J}}_{n,k}(\sigma)\right\} \le \eta\right\} = 0, \tag{24}$$

where

$$\tilde{\mathcal{I}}_{n,k}(\mu) \quad := \frac{1}{\beta_{n,k}}\int_0^{T_n} \frac{\tilde{u}^2(X_s,\mu,\mu_0)}{C^k(X_s)V^2(X_s,\sigma_0)}\,ds; \quad \tilde{u}(x,\mu,\mu_0) := U(x,\mu_0) - U(x,\mu);$$

$$\tilde{\mathcal{J}}_{n,k}(\sigma) := \frac{1}{\gamma_{n,k}}\int_0^{T_n} \frac{\tilde{v}(X_s,\sigma,\sigma_0)}{C^k(X_s)}\,ds; \quad \tilde{v}(x,\sigma,\sigma_0) := \frac{V^2(x,\sigma_0)}{V^2(x,\sigma)} - \log\frac{V^2(x,\sigma_0)}{V^2(x,\sigma)} - 1.$$

It would be clear that Condition (22) is reduced to the tightness of $(\mathcal{I}_{n,k}, \mathcal{J}_{n,k})$, and that Conditions (23) and (24) are reduced to the tightness of $(\mathcal{I}_{n,k}^{-1}, \mathcal{J}_{n,k}^{-1})$ in the case of (1). In this case, a *local-Gauss type* contrast function with weights C_{i-1}^{-k} is also utilizable with some further regularities on U and V to ensure some *uniform convergence in the parameter* ϑ of the contrast function and its derivatives; the regularities are not needed in our simple case since the parameters and functions U, V are separated. Similarly, our methods would also easily extend to a higher dimensional framework.

6 Proofs

6.1 Auxiliary results

Let $\Psi_n(\vartheta)$ be contrast functions on Θ: a compact convex subset of $\mathbb{R}^q$, and $\hat{\vartheta}_n \in \Theta$ be an M-estimator such that

$$\Psi_n(\hat{\vartheta}_n) = \inf_{\vartheta\in\Theta} \Psi_n(\vartheta).$$

The following propositions given by Shimizu (2009b), are useful to show our main theorems. One can find the essence of the proofs in the paper by Jacod (2006). Here, we present the statements without proofs: the first result is for consistency of $\hat{\vartheta}_n$ for some value $\vartheta_0 \in \Theta$.

Proposition 1 *Suppose that there exist functions $F_n(\vartheta)$, possibly random, such that*

$$\sup_{\vartheta\in\Theta} |\Psi_n(\vartheta) - F_n(\vartheta)| \xrightarrow{P_{\vartheta_0}} 0 \tag{25}$$

as $n \to \infty$, and that there exists some $\vartheta_0 \in \Theta$ such that, for each $\varepsilon > 0$,

$$\left[\inf_{\vartheta:|\vartheta-\vartheta_0|>\varepsilon}(F_n(\vartheta)-F_n(\vartheta_0))\right]^{-1} \quad \textit{is uniformly tight.} \tag{26}$$

Then $\hat{\vartheta}_{n,k}$ *is consistent for* ϑ_0: $\hat{\vartheta}_n \xrightarrow{P_{\vartheta_0}} \vartheta_0$ *as* $n \to \infty$.

The next result is for the rate of convergence of $\hat{\vartheta}_n$. Hereafter, for $\vartheta = (\vartheta^1, \dots, \vartheta^q)$, we denote by $\nabla_\vartheta = (\partial_{\vartheta^1}, \dots, \partial_{\vartheta^q})^\top$ the differential operator, and by $\nabla_\vartheta^2 = \nabla_\vartheta^\top \nabla_\vartheta$.

Proposition 2 *Suppose that* $\vartheta_0 \in int(\Theta)$, *and that* $\hat{\vartheta}_n$ *is consistent for* ϑ_0. *Moreover, suppose that* $\Psi_n(\vartheta)$ *is twice differentiable in* ϑ *for each n, and that there exist some* $q \times q$*-matrices* K_n, *invertible* $q \times q$*-matrices* D_n *and* $\tilde{F}_n(\vartheta)$, *all of which are possibly random, such that, as* $n \to \infty$,

$$K_n \nabla_\vartheta \Psi_n(\vartheta_0) \quad \textit{is uniformly tight;} \tag{27}$$

$$\left|\det(\tilde{F}_n(\vartheta_0))\right|^{-1} \quad \textit{is uniformly tight;} \tag{28}$$

$$\sup_{\vartheta\in\Theta}\left|K_n \nabla_\vartheta^2 \Psi_n(\vartheta) D_n^{-1} - \tilde{F}_n(\vartheta)\right| \xrightarrow{P_{\vartheta_0}} 0; \tag{29}$$

$$\sup_{|w|\le\varepsilon_n}\left|\tilde{F}_n(w+\vartheta_0) - \tilde{F}_n(\vartheta_0)\right| \xrightarrow{P_{\vartheta_0}} 0, \tag{30}$$

where ε_n *is arbitrary sequence tending to zero. Then* $D_n(\hat{\vartheta}_n - \vartheta_0)$ *is uniformly tight.*

The next lemma is obtained immediately from Lemma 9 by Genon-Catalot and Jacod (1993) and its proof, so we omit the proof.

Lemma 1 *Let* g_i^n $(i = 1, \dots, n)$ *be* $\sigma(X_{t_i^n})$*-measurable random variables, and let*

$$A_n := \sum_{i=1}^n E_{\vartheta_0}\left[g_i^n \middle| X_{t_{i-1}^n}\right], \quad B_n := \sum_{i=1}^n E_{\vartheta_0}\left[|g_i^n|^2 \middle| X_{t_{i-1}^n}\right].$$

If both A_n *and* B_n *are uniformly tight, then* $\left\{\sum_{i=1}^n g_i^n\right\}_{n\in\mathbb{N}}$ *is also uniformly tight. In particular, if* $A_n \xrightarrow{P_{\vartheta_0}} A$ *for a random variable A, and* $B_n \xrightarrow{P_{\vartheta_0}} 0$, *then* $\sum_{i=1}^n g_i^n \xrightarrow{P_{\vartheta_0}} A$.

Lemma 2 *Assume that* $X \in \mathcal{D}_k(\beta_{n,k}, \gamma_{n,k})$. *Moreover, suppose Assumptions 1, 2, (H[k]) and (LLN2). Furthermore, suppose that* $h_n \to 0$, *and that a positive sequence* $\{\alpha_n\}$ *satisfies that*

$$nh_n^{3/2}\alpha_n^{-1} + \beta_{n,k}\alpha_n^{-2} \to 0, \tag{31}$$

as $n \to \infty$. *Then*

$$\frac{1}{\alpha_n}\sum_{i=1}^n \frac{U_{i-1}}{C_{i-1}^k V_{i-1}^2}\left(\Delta_i^n X - \kappa_n(\mu_0) U_{i-1}\right) \xrightarrow{P_{\vartheta_0}} 0. \tag{32}$$

In addition that,

$$\beta_{n,k}\alpha_n^{-1} \to 0. \tag{33}$$

Then

$$\sup_{\mu\Xi}\left|\frac{1}{\alpha_n}\sum_{i=1}^n \frac{U_{i-1}}{C_{i-1}^k V_{i-1}^2}\left(\Delta_i^n X - \kappa_n(\mu)U_{i-1}\right)\right| \xrightarrow{P_{\vartheta_0}} 0. \tag{34}$$

Proof Note the following decomposition:

$$\begin{aligned}\frac{1}{\alpha_n}\sum_{i=1}^n \frac{U_{i-1}}{C_{i-1}^k V_{i-1}^2}\left(\Delta_i^n X - \kappa_n(\mu)U_{i-1}\right) &= \frac{1}{\alpha_n}\sum_{i=1}^n \frac{U_{i-1}}{C_{i-1}^k V_{i-1}^2}\left(\Delta_i^n X - \kappa_n(\mu_0)U_{i-1}\right)\\ &\quad + \frac{1}{\alpha_n}\sum_{i=1}^n \frac{U_{i-1}^2}{C_{i-1}^k V_{i-1}^2}\left(\kappa_n(\mu_0) - \kappa_n(\mu)\right)\\ &=: \sum_{i=1}^n p_i^n + \sum_{i=1}^n q_i^n.\end{aligned}$$

It follow from Assumptions 2 (i), (ii) and 1 that

$$\begin{aligned}\left|\sum_{i=1}^n E_{\vartheta_0}\left[p_i^n | X_{t_{i-1}^n}\right]\right| &\lesssim \frac{h_n^{3/2}}{\alpha_n}\sum_{i=1}^n \frac{|U_{i-1}|R_{1,i-1}}{C_{i-1}^k} = O_{P_{\vartheta_0}}\left(\frac{nh_n^{3/2}}{\alpha_n}\right) \xrightarrow{P_{\vartheta_0}} 0;\\ \sum_{i=1}^n E_{\vartheta_0}\left[|p_i^n|^2 | X_{t_{i-1}^n}\right] &\lesssim \frac{\beta_{n,k}}{\alpha_n^2}\left(\frac{h_n}{\beta_{n,k}}\sum_{i=1}^n \frac{S_{i-1}}{C_{i-1}^k}\right) + \frac{h_n\beta_{n,k}}{\alpha_n^2}\left(\frac{h_n}{\beta_{n,k}}\sum_{i=1}^n \frac{S_{i-1}R_{2,i-1}}{C_{i-1}^{2k}}\right)\\ &= O_{P_{\vartheta_0}}\left(\frac{\beta_{n,k}}{\alpha_n^2}\right) \xrightarrow{P_{\vartheta_0}} 0.\end{aligned}$$

Note that here we used (H1,2) and (LLN2). Hence, we see that $\sum_{i=1}^n p_i^n \xrightarrow{P_{\vartheta_0}} 0$ by Lemma 1. Moreover, it follows from Assumption 1 that

$$\begin{aligned}\sup_{\mu\in\Xi}\left|\sum_{i=1}^n q_i^n\right| &\lesssim \frac{\beta_{n,k}}{\alpha_n}\left(\frac{h_n}{\beta_{n,k}}\sum_{i=1}^n \frac{S_{i-1}}{C_{i-1}^k}\right)\mathbf{1}_{\{\mu\neq\mu_0\}}\\ &= O_{P_{\vartheta_0}}\left(\frac{\beta_{n,k}}{\alpha_n}\mathbf{1}_{\{\mu\neq\mu_0\}}\right) \xrightarrow{P_{\vartheta_0}} 0.\end{aligned}$$

This completes the proof. □

Lemma 3 *Assume that* $X \in \mathcal{D}_k(\beta_{n,k}, \gamma_{n,k})$*. Moreover, suppose Assumptions* 1, 2*,* (*H*2,3) *in* ($H[k]$)*, (LLN1), and that*

$$h_n\gamma_{n,k}^{-1} \to 0.$$

Then

$$\frac{1}{\gamma_{n,k}}\sum_{i=1}^{n}\frac{\left(\Delta_i^n X-\kappa_n(\mu_0)U_{i-1}\right)^2}{C_{i-1}^k V_{i-1}^2}-\sigma_0\mathcal{J}_{n,k}\xrightarrow{P_{\vartheta_0}}0.$$

Proof Let

$$g_i^n := \frac{1}{\gamma_{n,k}}\frac{\left(\Delta_i^n X-\kappa_n(\mu_0)h_n U_{i-1}\right)^2}{C_{i-1}^k V_{i-1}^2}-\frac{h_n}{\gamma_{n,k}}\frac{\sigma_0}{C_{i-1}^k}.$$

It follow from (H2,3) and (LLN1) that

$$\left|\sum_{i=1}^{n}E_{\vartheta_0}\left[g_i^n\middle|X_{t_{i-1}^n}\right]\right|\lesssim\frac{h_n^2}{\gamma_{n,k}}\sum_{i=1}^{n}\frac{R_{2,i-1}}{C_{i-1}^k}\xrightarrow{P_{\vartheta_0}}0;$$

$$\sum_{i=1}^{n}E_{\vartheta_0}\left[|g_i^n|^2\middle|X_{t_{i-1}^n}\right]\lesssim\frac{h_n^2}{\gamma_{n,k}^2}\sum_{i=1}^{n}\frac{R_{4,i-1}}{C_{i-1}^{2k}}+\frac{h_n}{\gamma_{n,k}}\left(\frac{h_n}{\gamma_{n,k}}\sum_{i=1}^{n}\frac{1}{C_{i-1}^{2k}}\right)\xrightarrow{P_{\vartheta_0}}0.$$

Hence, by Lemma 1, we have the consequence. □

Lemma 4 *Assume that* $X\in\mathcal{D}_k(\beta_{n,k},\gamma_{n,k})$ *with* $\beta_{n,k}/\gamma_{n,k}=O(1)$*. Moreover, suppose Assumptions* 1, 2, $(H[k])$*, (LLN), and that*

$$nh_n^2\gamma_{n,k}^{-1}\to 0.$$

Then, uniformly in $\mu\in\Xi$,

$$\frac{1}{\gamma_{n,k}}\sum_{i=1}^{n}\frac{1}{C_{i-1}^k V_{i-1}^2}\left(\Delta_i^n X-\kappa_n(\mu)U_{i-1}\right)^2-\sigma_0\mathcal{J}_{n,k}\xrightarrow{P_{\vartheta_0}}0. \tag{35}$$

Proof Note the decomposition:

$$\frac{1}{\gamma_{n,k}}\sum_{i=1}^{n}\frac{\left(\Delta_i^n X-\kappa_n(\mu)U_{i-1}\right)^2}{C_{i-1}^k V_{i-1}^2}=\frac{1}{\gamma_{n,k}}\sum_{i=1}^{n}\frac{\left(\Delta_i^n X-\kappa_n(\mu_0)U_{i-1}\right)^2}{C_{i-1}^k V_{i-1}^2} \tag{36}$$

$$+\frac{1}{\gamma_{n,k}}\sum_{i=1}^{n}\frac{U_{i-1}^2}{C_{i-1}^k V_{i-1}^2}\left(\kappa_n(\mu_0)-\kappa_n(\mu)\right)^2 \tag{37}$$

$$+\frac{2}{\gamma_{n,k}}\sum_{i=1}^{n}\frac{\left(\kappa_n(\mu_0)-\kappa_n(\mu)\right)U_{i-1}}{C_{i-1}^k V_{i-1}^2}\left(\Delta_i^n X-\kappa_n(\mu_0)U_{i-1}\right). \tag{38}$$

The term (36) is asymptotically equivalent to $\sigma_0 \mathcal{J}_{n,k}$ by Lemma 3, and it follows for (37) that

$$\sup_{\mu \in \Xi} \left| \frac{1}{\gamma_{n,k}} \sum_{i=1}^{n} \frac{U_{i-1}^2}{C_{i-1}^k V_{i-1}^2} (\kappa_n(\mu_0) - \kappa_n(\mu))^2 \right| \lesssim \frac{h_n \beta_{n,k}}{\gamma_{n,k}} \frac{h_n}{\beta_{n,k}} \sum_{i=1}^{n} \frac{S_{i-1}}{C_{i-1}^k} \xrightarrow{P_{\vartheta_0}} 0,$$

by (LLN2) and the assumption $\beta_{n,k}/\gamma_{n,k} = O(1)$. Moreover, the term (38) also converges to zero in probability uniformly in μ by Lemma 2 (34) since (31) and (33) hold with $\alpha_n = \gamma_{n,k}/h_n$. This completes the proof. □

Lemma 5 *Assume that $X \in \mathcal{D}_k(\beta_{n,k}, \gamma_{n,k})$. Moreover, suppose Assumptions 1, 2, $(H[k])$ and (LLN2). Then, uniformly in $\mu \in \Xi$,*

$$T_{n,k}(\mu) - (\mu_0 - \mu)^2 \mathcal{I}_{n,k} \xrightarrow{P_{\vartheta_0}} 0, \tag{39}$$

where

$$\begin{aligned} T_{n,k}(\mu) := & \sum_{i=1}^{n} \frac{1}{h_n C_{i-1}^k V_{i-1}^2} \left(\Delta_i^n X - \kappa_n(\mu) U_{i-1}\right)^2 \\ & - \sum_{i=1}^{n} \frac{1}{h_n C_{i-1}^k V_{i-1}^2} \left(\Delta_i^n X - \kappa_n(\mu_0) U_{i-1}\right)^2. \end{aligned}$$

Proof Note that

$$\frac{1}{\beta_{n,k}} \sum_{i=1}^{n} \frac{\left(\Delta_i^n X - \kappa_n(\mu) U_{i-1}\right)^2}{h_n C_{i-1}^k V_{i-1}^2} = \frac{1}{h_n \beta_{n,k}} \sum_{i=1}^{n} \frac{U_{i-1}^2}{C_{i-1}^k V_{i-1}^2} (\kappa_n(\mu_0) - \kappa_n(\mu))^2 \tag{40}$$

$$+ \frac{2}{h_n \beta_{n,k}} \sum_{i=1}^{n} \frac{U_{i-1} (\kappa_n(\mu_0) - \kappa_n(\mu))}{C_{i-1}^k V_{i-1}^2} \left(\Delta_i^n X - \kappa_n(\mu_0) U_{i-1}\right) \tag{41}$$

$$=: \sum_{i=1}^{n} p_i^n + \sum_{i=1}^{n} q_i^n. \tag{42}$$

It follows by Assumption 1 and (H2) that

$$\left| \sum_{i=1}^{n} p_i^n - \frac{h_n}{\beta_{n,k}} \sum_{i=1}^{n} \frac{S_{i-1}}{C_{i-1}^k} (\mu - \mu_0)^2 \right| \xrightarrow{P_{\vartheta_0}} 0,$$

uniformly in μ. Since

$$\frac{h_n}{\beta_{n,k}} \sum_{i=1}^{n} \frac{S_{i-1}}{C_{i-1}^k} (\mu - \mu_0)^2 - (\mu_0 - \mu)^2 \mathcal{I}_{n,k} \xrightarrow{P_{\vartheta_0}} 0,$$

by (LLN2), we see that $\sum_{i=1}^n p_i^n - (\mu_0-\mu)^2\mathcal{I}_{n,k} \xrightarrow{P_{\vartheta_0}} 0$. Finally, it follows from Lemma 2 (32) with $\alpha_n = \beta_{n,k}$ that $\sum_{i=1}^n q_i^n \xrightarrow{P_{\vartheta_0}} 0$ uniformly in μ; the uniformity is obvious since $(\kappa_n(\mu_0)-\kappa_n(\mu))$ is separated. This completes the proof. □

6.2 Proofs of main results

6.2.1 Proof of Theorem 1

Note that an estimator $\hat{\sigma}_{n,k}$ is a minimizer of

$$\Psi_{1,n,k}(\sigma) := \frac{h_n}{\gamma_{n,k}}\left\{Q_{n,k}(\hat{\mu}_{n,k},\sigma) - Q_{n,k}(\hat{\mu}_{n,k},\sigma_0)\right\}.$$

It follows from Lemma 4 and (LLN1) that

$$\sup_{\sigma\in\Pi}\left|\Psi_{1,n,k}(\sigma) - F_{1,n,k}(\sigma)\right| \xrightarrow{P_{\vartheta_0}} 0,$$

where $F_{1,n,k}(\sigma) := \left(\frac{\sigma_0}{\sigma} + \log\frac{\sigma}{\sigma_0} - 1\right)\mathcal{J}_{n,k}$. Moreover, it follows for any $\varepsilon > 0$ that

$$\inf_{\sigma:|\sigma-\sigma_0|>\varepsilon}\left(F_{1,n,k}(\sigma) - F_{1,n,k}(\sigma_0)\right) = \inf_{\sigma:|\sigma-\sigma_0|>\varepsilon}\left(\frac{\sigma_0}{\sigma} + \log\frac{\sigma}{\sigma_0} - 1\right)\mathcal{J}_{n,k},$$

the inverse of which is tight since $\mathcal{J}_{n,k}^{-1}$ is tight from the assumption. Hence, applying Proposition 1, we obtain that $\hat{\sigma}_{n,k} \xrightarrow{P_{\vartheta_0}} \sigma_0$. Next, we consider the following contrast function:

$$\Psi_{2,n,k}(\mu) := \frac{1}{\beta_{n,k}}\left\{Q_{n,k}(\mu,\hat{\sigma}_{n,k}) - Q_{n,k}(\mu_0,\hat{\sigma}_{n,k})\right\}.$$

Note that a statistic $\hat{\mu}_{n,k}$ is a minimizer of $\Psi_{2,n,k}$ in Ξ, and that

$$\begin{aligned}\Psi_{2,n,k}(\mu) = &\frac{1}{\hat{\sigma}_{n,k}h_n\beta_{n,k}}\sum_{i=1}^n \frac{2\left(\kappa_n(\mu_0)-\kappa_n(\mu)\right)U_{i-1}}{C_{i-1}^k V_{i-1}^2}\left(\Delta_i^n X - \kappa_n(\mu_0)U_{i-1}\right)\\ &+\frac{1}{\hat{\sigma}_{n,k}h_n\beta_{n,k}}\sum_{i=1}^n \frac{U_{i-1}^2}{C_{i-1}^k V_{i-1}^2}\left(\kappa_n(\mu_0)-\kappa_n(\mu)\right)^2.\end{aligned}$$

Using Lemma 2 (32) with $\alpha_n = \beta_{n,k}$ and (LLN2), we obtain that

$$\sup_{\mu\in\Xi}\left|\Psi_{2,n,k}(\mu) - F_{2,n,k}(\mu)\right| \xrightarrow{P_{\vartheta_0}} 0,$$

where $F_{2,n,k}(\mu) := \sigma_0^{-1}(\mu_0 - \mu)^2 \mathcal{I}_{n,k}$. Moreover, it follows for any $\varepsilon > 0$ that

$$\inf_{\mu:|\mu-\mu_0|>\varepsilon} \left(F_{2,n,k}(\mu) - F_{2,n,k}(\mu_0)\right) = \inf_{\mu:|\mu-\mu_0|>\varepsilon} \sigma_0^{-1}(\mu_0 - \mu)^2 \mathcal{I}_{n,k},$$

the inverse of which is tight since $\mathcal{I}_{n,k}^{-1}$ is tight from the assumption. Hence, applying Proposition 1, we obtain that $\hat{\mu}_{n,k} \xrightarrow{P_{\vartheta_0}} \mu_0$.

6.2.2 Proof of Theorem 2

Note that the assumptions ensure that $\hat{\vartheta}_{n,k} \xrightarrow{P_{\vartheta_0}} \vartheta_0$ by Theorem 1. Let

$$\tilde{Q}_{n,k}(\vartheta) = D_n^{-1} \nabla_\vartheta^2 Q_{n,k}(\vartheta) D_n^{-1}; \qquad D_n = \begin{pmatrix} \sqrt{\beta_{n,k}} & 0 \\ 0 & \sqrt{\gamma_{n,k}/h_n} \end{pmatrix}.$$

According to Proposition 2, the proof ends if the following (a) and (b) are shown:

(a) *There exist random matrices $\tilde{F}_n(\vartheta)$ satisfying* (28) *and* (30)*-type conditions in Proposition* 2 *such that* $|\tilde{Q}_{n,k}(\vartheta) - \tilde{F}_n(\vartheta)| \xrightarrow{P_{\vartheta_0}} 0$ *uniformly in* $\vartheta \in \Theta$*, which corresponds to* (29)*;*

(b) *Sequence* $\left\{D_n^{-1} \nabla_\vartheta Q_{n,k}(\vartheta_0)\right\}_{n\in\mathbb{N}}$ *is uniformly tight.*

Note that

$$\begin{aligned}
\partial_\mu Q_{n,k}(\vartheta) &= -2 \sum_{i=1}^n \frac{\partial_\mu \kappa_n(\mu) U_{i-1}}{h_n \sigma C_{i-1}^k V_{i-1}^2} \left(\Delta_i^n X - \kappa_n(\mu) U_{i-1}\right); \\
\partial_\sigma Q_{n,k}(\vartheta) &= \sum_{i=1}^n \left\{ \frac{1}{\sigma C_{i-1}^k} - \frac{\left(\Delta_i^n X - \kappa_n(\mu) U_{i-1}\right)^2}{h_n \sigma^2 C_{i-1}^k V_{i-1}^2} \right\}; \\
\partial_\mu^2 Q_{n,k}(\vartheta) &= 2 \sum_{i=1}^n \frac{(\partial_\mu \kappa_n(\mu) U_{i-1})^2 - \partial_\mu^2 \kappa_n(\mu) U_{i-1} \left(\Delta_i^n X - \kappa_n(\mu) U_{i-1}\right)}{h_n \sigma C_{i-1}^k V_{i-1}^2}; \\
\partial_\sigma^2 Q_{n,k}(\vartheta) &= \sum_{i=1}^n \left\{ \frac{2\left(\Delta_i^n X - \kappa_n(\mu) U_{i-1}\right)^2}{h_n \sigma^3 C_{i-1}^k V_{i-1}^2} - \frac{1}{\sigma^2 C_{i-1}^k} \right\}; \\
\partial_\mu \partial_\sigma Q_{n,k}(\vartheta) &= 2 \sum_{i=1}^n \frac{\partial_\mu \kappa_n(\mu) U_{i-1} \left(\Delta_i^n X - \kappa_n(\mu) U_{i-1}\right)}{h_n \sigma^2 C_{i-1}^k V_{i-1}^2}.
\end{aligned}$$

Then it is easy to see the point (a) since, by Lemmas 2 (32) with $\alpha_n = \beta_{n,k}/h_n$ and $\sqrt{\beta_{n,k}\gamma_{n,k}/h_n}$ and Lemma 4, we can take

$$\tilde{F}_n(\vartheta) = \mathrm{diag}\left(\frac{2}{\sigma_0} I_{n,k}, \left(\frac{2\sigma_0}{\sigma_3} - \frac{1}{\sigma^2} \right) J_{n,k} \right),$$

which clearly satisfies (30), and also (28) because of the tightness of $\mathcal{I}_{n,k}^{-1}$ and $\mathcal{J}_{n,k}^{-1}$.

Finally, we shall show the point (b). Putting

$$
\begin{aligned}
p_i^n &:= \frac{-2}{\sqrt{\beta_{n,k}}} \frac{\partial_\mu \kappa_n(\mu) U_{i-1}}{h_n \sigma C_{i-1}^k V_{i-1}^2} \left(\Delta_i^n X - \kappa_n(\mu_0) U_{i-1} \right); \\
q_i^n &:= \sqrt{\frac{h_n}{\gamma_{n,k}}} \left\{ \frac{1}{\sigma C_{i-1}^k} - \frac{\left(\Delta_i^n X - \kappa_n(\mu) U_{i-1} \right)^2}{h_n \sigma^2 C_{i-1}^k V_{i-1}^2} \right\},
\end{aligned}
$$

we see by Assumption 2 that

$$
\begin{aligned}
&\sum_{i=1}^n \left| E_{\vartheta_0} \left[p_i^n \middle| X_{t_{i-1}^n} \right] \right| \lesssim \frac{h_n^{3/2}}{\sqrt{\beta_{n,k}}} \sum_{i=1}^n \frac{R_{1,i-1} U_{i-1}}{C_{i-1}^k} = O_{P_{\vartheta_0}} \left(\frac{n h_n^{3/2}}{\sqrt{\beta_{n,k}}} \right) \xrightarrow{P_{\vartheta_0}} 0; \\
&\sum_{i=1}^n E_{\vartheta_0} \left[|p_i^n|^2 \middle| X_{t_{i-1}^n} \right] \lesssim \frac{h_n}{\beta_{n,k}} \sum_{i=1}^n \frac{S_{i-1}^2}{C_{i-1}^k} + \frac{h_n^2}{\beta_{n,k}} \sum_{i=1}^n \frac{R_{2,i-1}}{C_{i-1}^{2k}} = O_{P_{\vartheta_0}} \left(\mathcal{I}_{n,k} + \frac{n h_n^2}{\beta_{n,k}} \right); \qquad (43) \\
&\sum_{i=1}^n \left| E_{\vartheta_0} \left[q_i^n \middle| X_{t_{i-1}^n} \right] \right| \lesssim \frac{h_n^{3/2}}{\sqrt{\gamma_{n,k}}} \sum_{i=1}^n \frac{R_{2,i-1}}{C_{i-1}^k V_{i-1}^2} \xrightarrow{P_{\vartheta_0}} 0; \\
&\sum_{i=1}^n E_{\vartheta_0} \left[|q_i^n|^2 \middle| X_{t_{i-1}^n} \right] \lesssim \frac{h_n}{\gamma_{n,k}} \sum_{i=1}^n \frac{1}{C_{i-1}^k} + \frac{h_n}{\gamma_{n,k}} \sum_{i=1}^n \frac{R_{4,i-1}}{C_{i-1}^{2k}} = O_{P_{\vartheta_0}} \left(\mathcal{J}_{n,k} + \frac{n h_n}{\gamma_{n,k}} \right).
\end{aligned}
$$

Therefore, by Lemma 1, we see the tightness of $\sum_{i=1}^n p_i^n$ and $\sum_{i=1}^n q_i^n$. This completes the proof.

6.3 Proof of Theorem 3

Consider the following contrast function for μ:

$$
\begin{aligned}
T_{n,k}(\mu) := &\sum_{i=1}^n \frac{1}{h_n C_{i-1}^k V_{i-1}^2} \left(\Delta_i^n X - \kappa_n(\mu) U_{i-1} \right)^2 \\
&- \sum_{i=1}^n \frac{1}{h_n C_{i-1}^k V_{i-1}^2} \left(\Delta_i^n X - \kappa_n(\mu_0) U_{i-1} \right)^2.
\end{aligned}
$$

Then an M-estimator for $T_{n,k}$ is consistent with $\hat{\mu}_{n,k}$ defined in (9), that is,

$$
T_{n,k}(\hat{\mu}_{n,k}) = \inf_{\vartheta \in \Theta} T_{n,k}(\mu).
$$

By Lemma 5, it follows that

$$
\sup_{\mu \in \Xi} \left| \frac{1}{\beta_{n,k}} T_{n,k}(\mu) - \tilde{T}_{n,k}(\mu) \right| \xrightarrow{P_{\vartheta_0}} 0,
$$

where $\tilde{T}_{n,k}(\mu) := (\mu_0 - \mu)^2 \mathcal{I}_{n,k}$. Moreover, it follows for any $\varepsilon > 0$ that

$$\left[\inf_{\mu:|\mu-\mu_0|>\varepsilon} \left(\tilde{T}_{n,k}(\mu) - \tilde{T}_{n,k}(\mu_0)\right)\right]^{-1} \le \varepsilon^{-2}\mathcal{I}_{n,k}^{-1},$$

which is tight. Hence, by Proposition 1, it follows that $\hat{\mu}_{n,k} \xrightarrow{P_{\vartheta_0}} \mu_0$.

To show the tightness of $\sqrt{\beta_{n,k}}(\hat{\mu}_{n,k} - \mu_0)$, we use Proposition 2. We will show the following:

(a) *There exist random variables* $\tilde{F}_n(\mu)$ *satisfying* (28) *and* (30)*-type conditions in Proposition* 2 *such that* $|\beta_{n,k}^{-1}\partial_\mu^2 T_{n,k}(\vartheta) - \tilde{F}_n(\mu)| \xrightarrow{P_{\vartheta_0}} 0$ *uniformly in* $\mu \in \Theta$*, which corresponds to* (29)*;*

(b) *Sequence* $\left\{\beta_{n,k}^{-1/2}\partial_\mu T_{n,k}(\mu_0)\right\}_{n\in\mathbb{N}}$ *is uniformly tight.*

The point (a) is easy to see since

$$\begin{aligned}\frac{1}{\beta_{n,k}}\partial_\mu^2 T_{n,k}(\mu) &= \frac{2}{\beta_{n,k}}\sum_{i=1}^n \frac{U_{i-1}}{h_n C_{i-1}^k V_{i-1}^2} \\ &\times \left[\left(\partial_\mu \kappa_n(\mu)\right)^2 U_{i-1} - \partial_\mu^2 \kappa_n(\mu)\left(\Delta_i^n X - \kappa_n(\mu)U_{i-1}\right)\right] \sim 2\mathcal{I}_{n,k} \text{ under } P_{\vartheta_0},\end{aligned}$$

uniformly in $\mu \in \Xi$ by Lemma 2 (34) with $\alpha_n = \beta_{n,k}/h_n$. The point (b) is clear by the same argument as in the proof of Theorem 2. This completes the proof.

6.4 Proof of Theorem 4

Consider the contrast function

$$\bar{Q}_{n,k}(\sigma) := \left\{\frac{1}{\sigma}T_{n,k}(\hat{\mu}_{n,k}) - \frac{1}{\sigma_0}T_{n,k}(\hat{\mu}_{n,k})\right\} + \sum_{i=1}^n \frac{1}{C_{i-1}^k}\log\frac{\sigma}{\sigma_0},$$

where $T_{n,k}$ is given in the proof of Theorem 3. Then the estimator $\hat{\sigma}_{n,k}$ is a minimizer of $\bar{Q}_{n,k}$. Here, we note that $\sqrt{\beta_{n,k}}\left(\hat{\mu}_{n,k} - \mu_0\right)$ is tight by Theorem 3, and that

$$\frac{h_n}{\gamma_{n,k}}T_{n,k}(\hat{\mu}_{n,k}) = \frac{1}{\gamma_{n,k}}\sum_{i=1}^n \frac{1}{C_{i-1}^k V_{i-1}^2}\left(\Delta_i^n X - \kappa_n(\mu_0)U_{i-1}\right)^2 \tag{44}$$

$$+\frac{1}{\beta_{n,k}\gamma_{n,k}}\sum_{i=1}^n \frac{U_{i-1}^2}{C_{i-1}^k V_{i-1}^2}\left\{\sqrt{\beta_{n,k}}\left(\kappa_n(\mu_0) - \kappa_n(\hat{\mu}_{n,k})\right)\right\}^2 \tag{45}$$

$$\begin{aligned}&+\frac{1}{\sqrt{\beta_{n,k}}\gamma_{n,k}}\sum_{i=1}^n \frac{U_{i-1}}{C_{i-1}^k V_{i-1}^2}\left(\Delta_i^n X - \kappa_n(\mu_0)U_{i-1}\right) \\ &\sqrt{\beta_{n,k}}\left(\kappa_n(\mu_0) - \kappa_n(\hat{\mu}_{n,k})\right).\end{aligned} \tag{46}$$

Then it follows from Lemma 3 that the term (44) is asymptotically equivalent to $\sigma_0 \mathcal{J}_{n,k}$ under P_{ϑ_0}, and the term (45) tends to zero in probability by the fact that

$$\sqrt{\beta_{n,k}}\left(\kappa_n(\mu_0) - \kappa_n(\hat{\mu}_{n,k})\right) = O_{P_{\vartheta_0}}(h_n),$$

and (LLN2). Moreover, the term (46) also tends to zero in probability by Lemma 2 (32) with $\alpha_n = \sqrt{\beta_{n,k}}\gamma_{n,k}$. Consequently, we have

$$\sup_{\sigma\in\Pi}\left|\frac{h_n}{\gamma_{n,k}}\bar{Q}_{n,k}(\sigma) - \left(\frac{\sigma_0}{\sigma} + \log\frac{\sigma}{\sigma_0} - 1\right)\mathcal{J}_{n,k}\right| \xrightarrow{P_{\vartheta_0}} 0.$$

Hence, by the same argument as in the proof of Theorem 1, we have the consistence.

Next, along Proposition 2, we shall show the following:

(a) *There exists a random variable $\tilde{F}_n(\sigma)$ satisfying (30) and (28)-type conditions in Proposition 2 such that $|h_n\gamma_{n,k}^{-1}\partial_\sigma^2\bar{Q}_{n,k}(\sigma) - \tilde{F}_n(\sigma)| \xrightarrow{P_{\vartheta_0}} 0$ uniformly in $\sigma\in\Pi$, which corresponds to (29);*

(b) *Sequence $\left\{\sqrt{h_n/\gamma_{n,k}}\partial_\sigma\bar{Q}_{n,k}(\sigma_0)\right\}_{n\in\mathbb{N}}$ is uniformly tight.*

The point (a) is easy to see since, by the same argument as above,

$$\begin{aligned}\frac{h_n}{\gamma_{n,k}}\partial_\sigma^2\bar{Q}_{n,k}(\sigma) &= \frac{2}{\sigma^3}\frac{h_n}{\gamma_{n,k}}T_{n,k}(\hat{\mu}_{n,k}) - \frac{h_n}{\gamma_{n,k}}\sum_{i=1}^n\frac{1}{\sigma^2 C_{i-1}^k}\\ &\sim \left(\frac{2\sigma_0}{\sigma^3} - \frac{1}{\sigma^2}\right)\mathcal{J}_{n,k} \text{ under } P_{\vartheta_0},\end{aligned}$$

uniformly in $\sigma\in\Pi$. To show the point (b), note that

$$\begin{aligned}\sqrt{\frac{h_n}{\gamma_{n,k}}}\partial_\sigma\bar{Q}_{n,k}(\sigma_0) &= \sum_{i=1}^n\sqrt{\frac{h_n}{\gamma_{n,k}}}\left[\frac{1}{\sigma_0 C_{i-1}^k} - \frac{1}{\sigma_0^2 h_n C_{i-1}^k V_{i-1}^2}\left(\Delta_i^n X - \kappa_n(\mu_0)U_{i-1}\right)^2\right]\\ &\quad+\sum_{i=1}^n\sqrt{\frac{h_n}{\gamma_{n,k}}}\frac{U_{i-1}^2}{\sigma_0^2 h_n C_{i-1}^k V_{i-1}^2}\left(\kappa_n(\mu_0) - \kappa_n(\hat{\mu}_{n,k})\right)^2\\ &\quad+\sum_{i=1}^n\sqrt{\frac{h_n}{\gamma_{n,k}}}\frac{U_{i-1}}{\sigma_0^2 h_n C_{i-1}^k V_{i-1}^2}\left(\Delta_i^n X - \kappa_n(\mu_0)U_{i-1}\right)\left(\kappa_n(\mu_0) - \kappa_n(\hat{\mu}_{n,k})\right)\\ &=: \sum_{i=1}^n p_i^n + \sum_{i=1}^n q_i^n + \sum_{i=1}^n r_i^n \times h_n^{-1}\sqrt{\beta_{n,k}}\left(\kappa_n(\mu_0) - \kappa_n(\hat{\mu}_{n,k})\right),\end{aligned}$$

where it means that

$$r_i^n = \sqrt{\frac{h_n}{\beta_{n,k}\gamma_{n,k}}}\frac{U_{i-1}}{\sigma_0^2 C_{i-1}^k V_{i-1}^2}\left(\Delta_i^n X - \kappa_n(\mu_0)U_{i-1}\right),$$

and note that $h_n^{-1}\sqrt{\beta_{n,k}}\left(\kappa_n(\mu_0)-\kappa_n(\hat{\mu}_{n,k})\right)$ is tight. It follows that

$$\sum_{i=1}^{n}\left|E_{\vartheta_0}\left[p_i^n|X_{t_{i-1}^n}\right]\right| \lesssim \frac{h_n^{3/2}}{\sqrt{\gamma_{n,k}}}\sum_{i=1}^{n}\frac{R_{2,i-1}}{C_{i-1}^kV_{i-1}^2} = O_{P_{\vartheta_0}}\left(\frac{nh_n^{3/2}}{\sqrt{\gamma_{n,k}}}\right);$$
$$\sum_{i=1}^{n}E_{\vartheta_0}\left[|p_i^n|^2|X_{t_{i-1}^n}\right]\lesssim \frac{h_n}{\gamma_{n,k}}\sum_{i=1}^{n}\frac{1}{C_{i-1}^k}+\frac{h_n}{\gamma_{n,k}}\sum_{i=1}^{n}\frac{R_{4,i-1}}{C_{i-1}^{2k}}=O_{P_{\vartheta_0}}\left(\mathcal{J}_{n,k}+\frac{nh_n}{\gamma_{n,k}}\right), \tag{47}$$

which are tight. Therefore, $\sum_{i=1}^{n}p_i^n$ is tight by Lemma 1. Moreover,

$$\left|\sum_{i=1}^{n}q_i^n\right| \lesssim \sqrt{\frac{h_n}{\gamma_{n,k}}\frac{h_n}{\beta_{n,k}}\sum_{i=1}^{n}\frac{S_{i-1}}{C_{i-1}^k}} \times \beta_{n,k}\left(\kappa_n(\mu_0)-\kappa_n(\hat{\mu}_{n,k})\right)^2 \xrightarrow{P_{\vartheta_0}} 0,$$

and $\sum_{i=1}^{n}r_i^n \xrightarrow{P_{\vartheta_0}} 0$ by Lemma 2 (34) with $\alpha_n=\sqrt{\beta_{n,k}\gamma_{n,k}/h_n}$. This completes the proof.

6.5 Proof of Corollary 2

Note that, for $T_{n,k}$ given in the proof of Theorem 3,

$$Q_{n,k}(\vartheta)=\frac{1}{\sigma}T_{n,k}(\mu)+\sum_{i=1}^{n}\frac{\log\sigma}{C_{i-1}^k},$$

and that

$$-D_n^{-1}\nabla_\vartheta Q_{n,k}(\vartheta_0)=\tilde{Q}_{n,k}(\vartheta_n^*)\cdot D_n(\hat{\vartheta}_{n,k}-\vartheta_0),$$

where $\vartheta_n^*=(\mu_n^*,\sigma_n^*)$ is between $\hat{\vartheta}_{n,k}$ and ϑ_0, $\tilde{Q}_{n,k}(\vartheta)=D_n^{-1}\nabla_\vartheta^2 Q_{n,k}(\vartheta)D_n^{-1}$ and

$$D_n=\begin{pmatrix}\sqrt{\beta_{n,k}} & 0\\ 0 & \sqrt{\gamma_{n,k}/h_n}\end{pmatrix}.$$

We have already shown in the proofs of Theorems 3 and 4 that $D_n^{-1}\nabla_\vartheta Q_{n,k}(\vartheta_0)$ is tight, and that

$$\operatorname{diag}\left(\tilde{Q}_{n,k}(\vartheta_n^*)\right)\sim\operatorname{diag}\left(\frac{2}{\sigma_0}\mathcal{I}_{n,k},\frac{1}{\sigma_0^2}\mathcal{J}_{n,k}\right).$$

Therefore, the proof ends if we show that

$$\sqrt{\frac{h_n}{\beta_{n,k}\gamma_{n,k}}}\partial_\mu\partial_\sigma Q_{n,k}(\vartheta_n^*) \xrightarrow{P_{\vartheta_0}} 0.$$

Here, we note that

$$\begin{aligned}
\sqrt{\frac{h_n}{\beta_{n,k}\gamma_{n,k}}}\partial_\mu\partial_\sigma Q_{n,k}(\vartheta_n^*) &= 2\sqrt{\frac{h_n}{\beta_{n,k}\gamma_{n,k}}}\sum_{i=1}^n \frac{U_{i-1}}{(\sigma_n^*)^2 C_{i-1}^k V_{i-1}^2} \\
&\quad \times \left(\Delta_i^n X - \kappa_n(\mu_n^*)U_{i-1}\right) + o_{P_{\vartheta_0}}(1) \\
&= \sqrt{\frac{h_n}{\beta_{n,k}\gamma_{n,k}}}\sum_{i=1}^n \frac{U_{i-1}}{C_{i-1}^k V_{i-1}^2}\left(\Delta_i^n X - \kappa_n(\mu_0)U_{i-1}\right) \\
&\quad + \sqrt{\frac{h_n}{\beta_{n,k}\gamma_{n,k}}}\sum_{i=1}^n \frac{U_{i-1}^2}{C_{i-1}^k V_{i-1}^2} \\
&\quad \times \left(\kappa_n(\mu_0) - \kappa_n(\mu_n^*)\right) + o_{P_{\vartheta_0}}(1) \\
&=: \sum_{i=1}^n p_i^n + \sum_{i=1}^n q_i^n + o_{P_{\vartheta_0}}(1).
\end{aligned}$$

Then it follows from Lemma 2 (34) with $\alpha_n = \sqrt{\beta_{n,k}\gamma_{n,k}/h_n}$ that $\sum_{i=1}^n p_i^n \xrightarrow{P_{\vartheta_0}} 0$. Moreover, noticing that $h_n^{-1}\sqrt{\beta_{n,k}}\left(\kappa_n(\vartheta_n^*) - \kappa_n(\vartheta_0)\right)$ is tight, we see that

$$\sum_{i=1}^n q_i^n = \sqrt{\frac{h_n}{\gamma_{n,k}}}\left(\frac{h_n}{\beta_{n,k}}\sum_{i=1}^n \frac{U_{i-1}^2}{C_{i-1}^k V_{i-1}^2}\right) \times h_n^{-1}\sqrt{\beta_{n,k}}\left(\kappa_n(\mu_0) - \kappa_n(\mu_n^*)\right) \xrightarrow{P_{\vartheta_0}} 0.$$

This completes the proof.

Appendix A: The law of large numbers

In this section, we use the following notation: for $p > 0$ and a measurable function G,

$$\|G\|_{L_p}^* := \sup_{t\geq 0} E_{\vartheta_0}\left[|G(X_t)|^p\right].$$

The following result is an L^p-version of the law of large numbers.

Proposition 3 *Let $f(x,\vartheta) : \mathbb{R}\times\Theta \to \mathbb{R}$ be a function which is twice differentiable in x such that the following three quantities are bounded: for some $p \geq 1$,*

$$\|U\partial_x f(\cdot,\vartheta)\|_{L_p}^*; \quad \|V\partial_x f(\cdot,\vartheta)\|_{L_{2p}}^*; \quad \|V^2\partial_x^2 f(\cdot,\vartheta)\|_{L_p}^*. \tag{48}$$

In addition, suppose that $nh_n^{1/2+1/p}\alpha_n^{-1} \to 0$ *as* $n \to \infty$. *Then*

$$\lim_{n\to\infty} \sup_{\vartheta\in\Theta} E_{\vartheta_0}\left[\left|\frac{h_n}{\alpha_n}\sum_{i=1}^n f_{i-1}(\vartheta) - \frac{1}{\alpha_n}\int_0^{T_n} f(X_s,\vartheta)\,\mathrm{d}s\right|^p\right] = 0.$$

Proof Let

$$\Delta_n(f;\vartheta) := \left|\frac{h_n}{\alpha_n}\sum_{i=1}^n f_{i-1}(\vartheta) - \frac{1}{\alpha_n}\int_0^{T_n} f(X_s,\vartheta)\,\mathrm{d}s\right|.$$

Thanks to Jensen's and Minkovskii's inequality, we can deduce that

$$\|\Delta_n(f;\vartheta)\|_{L_p(P_{\vartheta_0})} \le \frac{1}{\alpha_n}\sum_{i=1}^n\left(\int_{t_{i-1}^n}^{t_i^n} E_{\vartheta_0}\left[|f_{i-1}(\vartheta) - f(X_s,\vartheta)|^p\right]\mathrm{d}s\right)^{1/p}.$$

By Itô's formula, it follows for any $t > t_{i-1}^n$ and $p \ge 1$ that

$$\begin{aligned}\left|f(X_t,\vartheta) - f(X_{t_{i-1}^n},\vartheta)\right|^p &\le \left|\int_{t_{i-1}^n}^t \left(U(X_s)\partial_x f(X_s,\vartheta) + \frac{1}{2}V^2(X_s)\partial_x^2 f(X_s,\vartheta)\right)\mathrm{d}s\right|^p \\ &\quad + \left|\int_{t_{i-1}^n}^t V(X_s)\partial_x f(X_s,\vartheta)\,\mathrm{d}W_s\right|^p.\end{aligned}$$

Thanks to Condition (48), Jensen's and Burkholder–Davis–Gundy's inequality yields that

$$\sup_{\vartheta\in\Theta} E_{\vartheta_0}\left[\left|f(X_{t_i^n},\vartheta) - f(X_{t_{i-1}^n},\vartheta)\right|^p\right] \lesssim h_n^{p/2},$$

and

$$\sup_{\vartheta\in\Theta}\|\Delta_n(f;\vartheta)\|_{L_p(P_{\vartheta_0})} \lesssim nh_n^{1/2+1/p}\alpha_n^{-1}.$$

The last term tends to zero under the assumption. This completes the proof. □

Corollary 3 *Suppose that functions* U, V *and* C *are twice differentiable with bounded derivatives. Moreover, suppose that* $nh_n^{1/2+1/p}\gamma_{n,k}^{-1} \to 0$ *for some* $p \ge 1$. *Then*

$$E_{\vartheta_0}\left[\left|\frac{h_n}{\gamma_{n,k}}\sum_{i=1}^n \frac{1}{C_{i-1}^k} - \frac{1}{\gamma_{n,k}}\int_0^{T_n}\frac{1}{C^k(X_s)}\,\mathrm{d}s\right|^p\right] \to 0,$$

for any $k \ge 0$. *In particular, if* $X \in \mathcal{D}_k(\beta_{n,k},\gamma_{n,k})$, *then* (*LLN*1) *holds true.*

Proof Check Condition (48) with $f(x,\vartheta) = C^{-k}(x)$. □

Corollary 4 *Suppose that functions U, V and C are twice differentiable with bounded derivatives. Moreover, suppose that $nh_n^{1/2+1/p}\beta_{n,k}^{-1} \to 0$ for some $p \geq 1$, and that the following either (a) or (b) is satisfied:*

(a) $\|U^3 C^{-k}\|_{L_p}^* < \infty$.

(b) $\|U^2 C^{-k}\|_{L_p}^* < \infty$, *and V is constant.*

Then

$$E_{\vartheta_0}\left[\left|\frac{h_n}{\beta_{n,k}}\sum_{i=1}^{n}\frac{S_{i-1}}{C_{i-1}^k} - \frac{1}{\beta_{n,k}}\int_0^{T_n}\frac{S(X_s)}{C^k(X_s)}\,\mathrm{d}s\right|^p\right] \to 0.$$

Proof Put $f(x,\vartheta) = S(x)C^{-k}(x)$, and denote by $f'(x) = \partial_x f(x,\vartheta)$. Then, by the direct computation, we have the following estimates:

$$|f'| \lesssim \left|\frac{U}{C^k V^2}\right| + \left|\frac{U^2 V'}{C^k V^3}\right|;$$
$$|f''| \lesssim \left|\frac{U}{C^k V^2}\right| + \left|\frac{U^2 V''}{C^k V^3}\right| + \left|\frac{U^2 (V')^2}{C^k V^3}\right| + \left|\frac{U^2}{C^k V^4}\right|.$$

Use them to check Condition (48). □

Remark 8 Assume that $X \in \mathcal{D}_k(\beta_{n,k}, \gamma_{n,k})$, and that all the conditions in Corollary 4 are satisfied. Then (LLN2) holds true. Condition (a) or (b) is satisfied if $k \geq 3$, or $k \geq 2$ when V is a constant. Moreover, Condition (a) is satisfied for any $k \geq 0$ if U is bounded. In particular, (LLN1,2) hold for $\mathcal{D}_{\mathrm{NR}}$ if $h_n^{1/2}T_n^{1-\alpha} \to 0$.

Remark 9 For $\mathcal{D}_{\mathrm{OU}^+}$ and $\mathcal{D}_{\mathrm{Exp}}$, (LLN2) holds true with $k = 2$ and $\beta_{n,2} = T_n$ since

$$\mathcal{D}_{\mathrm{OU}^+},\ \mathcal{D}_{\mathrm{Exp}} \subset \mathcal{D}_2(T_n, 1);$$

see Examples 3 and 4. However, (LLN2) also holds if $k = 0$ and $\beta_{n,0} = e^{2\mu_0 T_n}$ although $\|X^2\|_{L_1}^*$ is not finite in these cases; see Lemmas 1 and 2 in Shimizu (2009a). It implies that (a) and (b) are not necessary, but sufficient.

Acknowledgments The author expresses many thanks to anonymous referees for their helpful comments and instructions. This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Young Scientists (B), 2009, no. 21740073, and JST, CREST.

References

Dietz, H. M., Kutoyants, Yu. A. (2003). Parameter estimation for some non-recurrent solutions of SDE. *Statistics & Decisions*, *21*(1), 29–45.

Feigin, P. D. (1976). Maximum likelihood estimation for continuous-time stochastic processes. *Advances in Applied Probability*, *8*(4), 712–736.

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

Gobet, E. (2002). LAN property for ergodic diffusions with discrete observations. *Annales de l'Institut Henri Poincaré Probabilités et Statistiques*, *38*(5), 711–737.

Höpfner, R., Kutoyants, Yu. A. (2003). On a problem of statistical inference in null recurrent diffusions. *Statistical Inference for Stochastic Processes*, *6*(1), 25–42.

Jacod, J. (2006). Parametric inference for discretely observed non-ergodic diffusions. *Bernoulli*, *12*(3), 383–401.

Kasonga, R. A. (1990). Parameter estimation by deterministic approximation of a solution of a stochastic differential equation. *Communications in Statistics. Stochastic Models*, *6*, 59–67.

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

Kutoyants, Y. A. (2004). *Statistical inference for ergodic diffusion processes*. Springer, London.

Shimizu, Y. (2009a). Notes on drift estimation for non-recurrent Ornstein-Uhlenbeck processes from sampled data. *Statistics & Probability Letters*, *79*(20), 2200–2207.

Shimizu, Y. (2009b). Quadratic type contrast functions for discretely observed non-ergodic diffusion processes, Research Report Series, 09-04. Department of Engineering Science, Osaka University, Osaka.

Shimizu, Y. (2009c). Local asymptotic mixed normality for discretely observed non-recurrent Ornstein-Uhlenbeck processes. *Annals of the Institute of Statistical Mathematics* (to appear). Available online at doi:10.1007/s10463-010-0307-4.