J. Japan Statist. Soc.
Vol. 35 No. 2 2005 147–170

# SIMPLE ESTIMATORS FOR PARAMETRIC MARKOVIAN TREND OF ERGODIC PROCESSES BASED ON SAMPLED DATA*

Hiroki Masuda**

Let $X$ be a stochastic process obeying a stochastic differential equation of the form $dX_t = b(X_t, \theta)dt + dY_t$, where $Y$ is an adapted driving process possibly depending on $X$'s past history, and $\theta \in \Theta \subset \boldsymbol{R}^p$ is an unknown parameter. We consider estimation of $\theta$ when $X$ is discretely observed at possibly non-equidistant time-points $(t_i^n)_{i=0}^n$. We suppose $h_n := \max_{1\le i\le n}(t_i^n - t_{i-1}^n) \to 0$ and $t_n^n \to \infty$ as $n \to \infty$: the data becomes more high-frequency as its size increases. Under some regularity conditions including the ergodicity of $X$, we obtain $\sqrt{nh_n}$-consistency of trajectory-fitting estimate as well as least-squares estimate, without identifying $Y$. Also shown is that some additional conditions, which requires $Y$'s structure to some extent, lead to asymptotic normality. In particular, a Wiener-Poisson-driven setup is discussed as an important special case.

*Key words and phrases*: Discrete sampling, parametric estimation, stochastic differential equation, trajectory-fitting.

## 1. Introduction

Consider the family of partly parametrized $d$-dimensional processes $X$ given by

$$X_t = X_0 + \int_0^t b(X_s, \theta)ds + Y_t, \tag{1.1}$$

where $\theta \in \Theta \subset \boldsymbol{R}^p$, an open bounded convex domain, $X_0$ is a random element with $\mathcal{L}(X_0) = \eta$ possibly unknown, $b : \boldsymbol{R}^d \times \Theta \to \boldsymbol{R}^d$ is a measurable function, and $Y = (Y_t)_{t\in\boldsymbol{R}_+}$ is a $d$-dimensional zero-mean adapted process. Suppose that there exists a true parameter $\theta_0 \in \Theta$ which induces true data that we observe, and that, instead of the full trajectory we have discretely sampled data $(X_{t_i^n})_{i=0}^n$, where $(t_i^n)$ is, for each $n \in \boldsymbol{N}$, a given positive bounded strictly increasing sequence such that $h_n := \max_{1\le i\le n}(t_i^n - t_{i-1}^n) = o(1)$ and $t_n^n \to \infty$ as $n \to \infty$; $(t_i^n)$ not have to be equally spaced.

The purpose of this article is to derive sets of sufficient conditions for weak consistency, $\sqrt{nh_n}$-consistency and asymptotic normality of the trajectory-fitting

Received February 16, 2005. Revised May 25, 2005. Accepted July 20, 2005.

*This work was partly supported by Grant-in-Aid for Scientific Research from the Ministry of Education, Japan, by Cooperative Research Program of the Institute of Statistical Mathematics, and by the 21st Century COE Program "Development of Dynamic Mathematics with High Functionality" at Kyushu University.

**Graduate School of Mathematics, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan. Email: hiroki@math.kyushu-u.ac.jp

estimator (TFE) and least-squares estimator (LSE) for $\theta_0$. Apart from the ergodicity of $X$ and some moment conditions, the proofs do not require any explicit structure of $Y$ up to the $\sqrt{nh_n}$-consistency, whereas the form of conditional covariance matrix of $Y$'s increments is needed for the asymptotic normality (see (2.7) below); the estimation for parameters possibly involved in $Y$ is out of our scope, so we do not express the $Y$'s possible dependence on $\theta$ in the notation. Existence of an "exogenous" processes contaminating $X$ is allowed; for example, our result may apply in cases where $Y$ obeys another stochastic differential equation (then $Y$ may be regarded as an exogenous randomness contaminating the skeleton dynamics $x = (x_t)_{t\in \boldsymbol{R}_+}$ described by $dx_t = b(x_t, \theta)dt$). Of course, within this setup the estimates are not efficient in general, however, from a practical point of view it is often important to obtain an easy-to-use estimate. This point is the primary contribution of this article. Once the model (the structure of $Y$) is fully specified, the classical one-step improvement together with a "better" estimating function often leads to a more efficient estimate with rate $\sqrt{nh_n}$ under the condition of our Theorem (b).

The rest of this article is organized as follows. The precise framework and the main result is described in Section 2. Section 3 presents a special important case where, given the initial element $X_0$, everything is realized on the Wiener-Poisson space. In Section 4 we consider a concrete model to observe the performance of the estimates for different decreasing rates of $h_n$. The proofs are given in the Appendix.

We end this section with some historical remarks and comments. The model in question is a fairly particular subclass of general "stochastic differential equations", which plays an important role for modelling a continuously time-varying phenomenon, as they are frequently used in many fields of application. However, quite often real data is sampled at discrete-time points, so that we need to formulate a "statistical inference for stochastic differential equations from sampled data". Clearly this is a rather abstract matter because of diversity of the model. Such studies date back to, at latest, the middle of the 1970s. In the light of history in this area, there exists an extensive literature on estimating both drift and diffusion coefficients for diffusion processes, including efficient results in "smooth" cases. Regarding these points, the reader can consult Prakasa Rao (1999) and references therein.

The TFE was studied by e.g. Dietz and Kutoyants (1997) for continuously observed diffusion processes and by Kasonga (1990) for a class of discretely observed diffusions; considering, for example, the ergodic Markovian Ornstein-Uhlenbeck process driven by a Wiener process, we can notice that the condition of Kasonga (1990) is not suitable for ergodic cases, so the route we shall take in this article is different from his, although the same contrast function is used. The study of the LSE for discrete sampling goes back to Dorogovcev (1976) and Prakasa Rao (1983), also in case of diffusions.

Recently, inference for processes with jumps based on sampled data has drawn the attention of statisticians because of the applicability to several kinds

of realistic data seeming to have a discontinuous path. Nevertheless, much less than diffusions has been known so far: Shimizu and Yoshida (2005) and Shimizu (2005) studied asymptotic normality, both dealing with cases where the jump part of driving noise process is of finite variation. Beyond the Markovian framework (but still with the Markovian-trend structure), no result concerning the discrete sampling has yet been established. Our present result provides a widely applicable $\sqrt{nh_n}$-consistent estimates for the drift coefficient of the process in question possibly having infinitely many jumps on every compact time-interval yielding unbounded variation of the driving noise process.

## 2. Statement of the result

Let $(\Omega, \mathscr{F}, \boldsymbol{F} = (\mathscr{F}_t)_{t\geq 0}, P)$ be an underlying complete stochastic basis satisfying the usual hypothesis (e.g., Protter (1990)), on which a $d$-dimensional zero-mean $\boldsymbol{F}$-adapted process $Y = (Y_t)_{t\in \boldsymbol{R}_+}$ is endowed. Let $\Theta \subset \boldsymbol{R}^p$ be an open convex domain with compact closure, and consider the partly parametrized model $X$ given by (1.1) with $\mathscr{F}_0$-measurable initial element $X_0$. As was mentioned at the beginning, we have only sampled data $(X_{t_i^n})_{i=0}^n$. Hereafter we write $\Delta_i^n t = t_i^n - t_{i-1}^n$. Let $\theta_0 \in \Theta$ denote the true value, which induce the true image measure $P_0^\eta$ of $X$ associated with initial distribution $\eta$. Throughout this article, any order symbol is used for $n \to \infty$.

### 2.1. Two contrast functions

We introduce the set of auxiliary processes $\{\bar{X}_{i,t}(\theta) : t \in [t_{i-1}^n, t_i^n)\}_{i=1}^n$ defined by

$$\begin{cases} d\bar{X}_{i,t}(\theta) = b(\bar{X}_{i,t}(\theta), \theta)dt, & t \in [t_{i-1}^n, t_i^n), \\ \bar{X}_{i,t_{i-1}^n}(\theta) = X_{t_{i-1}^n}, \end{cases} \tag{2.1}$$

and then consider the function $\Phi_n(\theta) = \Phi_n(\theta; (X_{t_i^n})_{i=0}^n)$ given by

$$\Phi_n(\theta) = \sum_{i=1}^n |X_{t_i^n} - \bar{X}_{i,t_i^n}(\theta)|^2.$$

Let $(\tilde{\theta}_n)$ be any sequence such that $\Phi_n(\tilde{\theta}_n) \leq \Phi_n(\theta_0) + o_{P_0^\eta}(nh_n^2)$. We call $\tilde{\theta}_n$ the trajectory-fitting estimate (TFE) for $\theta_0$. We also consider $\Psi_n(\theta) = \Psi_n(\theta; (X_{t_i^n})_{i=0}^n)$ given by

$$\Psi_n(\theta) = \sum_{i=1}^n |X_{t_i^n} - X_{t_{i-1}^n} - b(X_{t_{i-1}^n}, \theta)\Delta_i^n t|^2, \tag{2.2}$$

and similarly define the least-squares estimate (LSE) $\hat{\theta}_n$ by any sequence satisfying $\Psi_n(\hat{\theta}_n) \leq \Psi_n(\theta_0) + o_{P_0^\eta}(nh_n^2)$. The LSE is convenient when (2.1) is not explicitly solvable. There is an obvious simple relation between $\Phi_n(\theta)$ and $\Psi_n(\theta)$,

that is, according to the usual Euler scheme for ordinary differential equations, we have

$$\bar{X}_{i,t_i^n}(\theta) = X_{t_{i-1}^n} + b(X_{t_{i-1}^n}, \theta)\Delta_i^n t + O(h_n^2), \qquad P_0^\eta\text{-a.s.}, \tag{2.3}$$

for all $\theta$, under rather mild regularity of $(x, \theta) \mapsto b(x, \theta)$ and non-explosivity of $X$.

In the Appendix an asymptotic equivalence between $\Phi_n(\theta)$ and $\Psi_n(\theta)$ as well as between their derivatives will be given. Especially, (C.3) in the Appendix says that, as soon as $(nh_n^3)^{-1/2}\nabla_\theta \Psi_n(\theta_0)$ weakly tends to some limit and the rate condition $nh_n^3 = o(1)$ holds, $\tilde{\theta}_n$ and $\hat{\theta}_n$ have the same asymptotic property up to the first order. In this sense one may bring redundancy of $\tilde{\theta}_n$ to his/her mind. But there is no "absolute" order of superiority between TFE and LSE. Our numerical experiments given in Section 4 show that $\tilde{\theta}_n$ may provide a better finite-sample performance than $\hat{\theta}_n$, and vice versa: roughly speaking, $\tilde{\theta}_n$ (resp. $\hat{\theta}_n$) provides a better performance than $\hat{\theta}_n$ (resp. $\tilde{\theta}_n$) for slower (resp. faster) decreasing rates of $h_n$.

### 2.2. Assumptions and main result

We shall use the following notation: $E_0^\eta[f] = \int f dP_0^\eta$ for any $P_0^\eta$-integrable function $f$; $C$ stands for a positive constant independent of $n$, $A_n \lesssim B_n$ implies $A_n \le CB_n$ with $C$ possibly varying from line to line, and $A_n \asymp B_n$ means that there exists $C$ such that $C^{-1} \le A_n/B_n \le C$ for each $n$; $\nabla_a$ denotes the gradient operator with respect to a variable $a$; $\boldsymbol{R}^{k\otimes l}$ stands for the set of all $k \times l$ matrices with real entries; $\|F\|_I^{*,q} = \sup_{s\in I} |F_s|^q$ for any interval $I \subset \boldsymbol{R}_+$, constant $q > 0$, and process $F$; finally, $\Delta b(t, s; \theta) = b(X_t, \theta) - b(X_s, \theta)$ and $\Delta b(t; \theta, \theta') = b(X_t, \theta) - b(X_t, \theta')$.

Assumption 1. The function $(x, \theta) \mapsto b(x, \theta)$ is of class $C^{2,2}$, the possible derivatives fulfilling $\sup_{\theta\in\Theta} |\nabla_x^k \nabla_\theta^l b(x, \theta)| \lesssim (1 + |x|)^C$, and moreover $\sup_{x\in\boldsymbol{R}^d,\theta\in\Theta} |\nabla_x b(x, \theta)| < \infty$.

Assumption 2. The stochastic integral equation (1.1) admits a unique solution $X$, and the process $(X, Y)$ is $L^q(P_0^\eta)$-bounded for every $q \ge 2$, that is, $\sup_{t\in\boldsymbol{R}_+} \|(X_t, Y_t)\|_{L^q(P_0^\eta)} < \infty$ for every $q \ge 2$. Moreover, for every $q \ge 2$ there exists a positive bounded sequence $\Delta_{q,n} = O(h_n)$ such that $h_n^q/\Delta_{q,n} = o(1)$, and that

$$\sup_{1\le i\le n} E_0^\eta[\|Y - Y_{t_{i-1}^n}\|_{(t_{i-1}^n, t_i^n]}^{*,q}] \lesssim \Delta_{q,n}. \tag{2.4}$$

Assumption 3. There exist numbers $p' \ge p'' > p$ and a positive bounded sequence $\epsilon_n = o(1)$ such that, for every $\theta_1, \theta_2 \in \Theta$,

$$E_0^\eta \left[ \left| \frac{1}{nh_n^2} \sum_{i=1}^n (\Delta_i^n t)\Delta b(X_{t_{i-1}^n}; \theta_1, \theta_2)^\top (Y_{t_i^n} - Y_{t_{i-1}^n}) \right|^{p'} \right] \lesssim \epsilon_n |\theta_1 - \theta_2|^{p''}.$$

Also, it holds that

$$\frac{1}{\sqrt{nh_n^3}}\sum_{i=1}^{n}(\Delta_i^n t)\nabla_\theta b(X_{t_{i-1}^n},\theta_0)^\top(Y_{t_i^n}-Y_{t_{i-1}^n})=O_{P_0^\eta}(1).$$

ASSUMPTION 4. $X$ admits an invariant probability measure $\pi_0$ (generally depending on $\theta_0$, but not on $\eta$), for which

$$T^{-1}\int_0^T F(X_t)dt \xrightarrow{P_0^\eta} \pi_0(F) \tag{2.5}$$

as $T\to\infty$ for any $\pi_0$-integrable function $F$ on $\boldsymbol{R}^d$.

ASSUMPTION 5. It holds that $b(x,\theta)=b(x,\theta')$, $\pi_0$-a.e., if and only if $\theta=\theta'$.

ASSUMPTION 6. Let the random variables $m_i^n$ and $v_i^n$ respectively denote the conditional mean vector and covariance matrix of $Y_{t_i^n}-Y_{t_{i-1}^n}$ given $\mathscr{F}_{t_{i-1}^n}$. Then

$$\frac{1}{\sqrt{nh_n^3}}\sum_{i=1}^{n}(\Delta_i^n t)\nabla_\theta b(X_{t_{i-1}^n},\theta_0)^\top m_{i-1}^n=o_{P_0^\eta}(1), \tag{2.6}$$

$$\frac{1}{nh_n^3}\sum_{i=1}^{n}(\Delta_i^n t)^2\nabla_\theta b(X_{t_{i-1}^n},\theta_0)^\top v_i^n\nabla_\theta b(X_{t_{i-1}^n},\theta_0)\xrightarrow{P_0^\eta}\Sigma_0, \tag{2.7}$$

where $\Sigma_0\in\boldsymbol{R}^{p\otimes p}$ is nonrandom.

Define $\Gamma(\theta)=[\Gamma(\theta)^{ij}]_{i,j=1}^p:\Theta\to\boldsymbol{R}^{p\otimes p}$ by

$$\Gamma(\theta)^{ij}=\sum_{l=1}^{d}\int\nabla_{\theta_i}b^l(x,\theta)\nabla_{\theta_j}b^l(x,\theta)\pi_0(dx). \tag{2.8}$$

Our main result is the following, whose proof is postponed to the Appendix.

THEOREM. *Fix any $\theta_0\in\Theta$, and suppose $h_n\to 0$, $t_n^n\to\infty$,*

$$\max_{1\le i\le n}|\Delta_i^n t-h_n|=o(h_n)\qquad \textit{and}\qquad t_n^n\asymp nh_n. \tag{2.9}$$

*Then we have the following.*

(a) *Under Assumptions 1 to 5, $\tilde\theta_n$ and $\hat\theta_n$ are weakly consistent under $P_0^\eta$.*

(b) *Further suppose that $nh_n^3=O(1)$, that $nh_n^\alpha\to\infty$ for some constant $\alpha\in(1,3)$, and that there exists $q'\ge 2$ such that $nh_n(\Delta_{q',n})^{2/q'}=O(1)$. Moreover, suppose $\Gamma(\theta_0)$ is nonsingular. Then the sequences $\{\sqrt{nh_n}(\tilde\theta_n-\theta_0)\}_{n\in\boldsymbol{N}}$ and $\{\sqrt{nh_n}(\hat\theta_n-\theta_0)\}_{n\in\boldsymbol{N}}$ are $P_0^\eta$-tight.*

(c) *Further suppose Assumption 6, that $\Gamma(\theta_0)$ is non-singular, and $nh_n\Delta_{2,n}=o(1)$. Then*

$$\sqrt{nh_n}(\tilde{\theta}_n - \theta_0) \xrightarrow{\mathscr{L}(P_0^\eta)} \mathcal{N}_p(0, \Gamma(\theta_0)^{-1}\Sigma_0\Gamma(\theta_0)^{-1\top}), \tag{2.10}$$

*and* $\sqrt{nh_n}(\hat{\theta}_n - \theta_0)$ *has the same weak limit.*

*Remark* 1. The rate condition (2.9) is set just for technical reasons. The second one implies that $nh_n \to \infty$ under $t_n^n \to \infty$. Of course, the equidistant sampling (i.e. $h_n = \Delta_i^n t$ for every $i$) is the case.

*Remark* 2. Under the assumptions, $\pi_0(f) < \infty$ for every measurable $f$ of at most polynomial growth. To the knowledge of the author, how to identify $\pi_0$ in general is a difficult matter and has not been solved even for diffusion with jumps except for Lévy-driven Ornstein-Uhlenbeck processes. This is an important issue for direct use of Theorem (c), although the $\sqrt{nh_n}$-consistency can be verified without identifying $\pi_0$.

*Remark* 3. Assumption 4 concerning the "ergodicity" of $X$ is essential in our framework. Such a property of stochastic processes is of independent interest, and closely related to the operator-based ergodic theory. Our result may apply to, for example, cases where we do not know whether $X$ itself is ergodic or not, but instead we know that $(X, Y')$ with some process $Y'$ is ergodic. In particular, if $(X, Y')$ is Markovian, then we can utilize a well-developed stability theory for Markov processes (see the references cited in Masuda (2004)) to get the ergodic theorem for $(X, Y')$, and then we have the ergodic theorem for $X$ as well through a projection-type function $F(x, y') = F_1(x)$ in (2.5). See Example 3.2 below for such an example.

*Remark* 4. Under Assumption 2, Gronwall's inequality yields

$$\begin{aligned} E_0^\eta[\|X - X_{t_{i-1}^n}\|_{(t_{i-1}^n, t_i^n]}^{*,q}] &\lesssim h_n^q + E_0^\eta[\|Y - Y_{t_{i-1}^n}\|_{(t_{i-1}^n, t_i^n]}^{*,q}] \\ &\lesssim h_n^q + \Delta_{q,n} \lesssim \Delta_{q,n}. \end{aligned}$$

If $\Delta_{q,n}$ rapidly decreases and the convergence (2.5) takes place $P_0^\eta$-a.s., then we may strengthen the assertion Theorem (a) to a strong consistency, mimicking the argument of Kasonga (1988). However, such cases do not occur in the presence of jumps: e.g., if $X$ is a diffusion process with jumps, then $\Delta_{q,n} = O(h_n)$ whatever $q$ is large, see Appendix E. When $X$ is a diffusion-type process with Markovian drift coefficient and possibly non-Markovian diffusion coefficient, $\Delta_{q,n}$ may decrease rapidly by taking $q$ large.

*Remark* 5. In order to utilize the asymptotic normality (2.10), $\Gamma(\theta_0)$ and $\Sigma_0$ should be estimated. Claim C.1 in the Appendix (a uniform law of large numbers) says that

$$\hat{\Gamma}_n := \frac{1}{2nh_n^2}\nabla_\theta^2\Psi(\hat{\theta}_n) \xrightarrow{P_0^\eta} \Gamma(\theta_0). \tag{2.11}$$

Hence $\hat{\Gamma}_n$ can serve as an explicit estimator of $\Gamma(\theta_0)$, without any knowledge of $\pi_0$ (recall Remark 2). Moreover, when $\Sigma_0$ is of the form

$$\Sigma_0 = \int H(x, \theta_0)\pi_0(dx)$$

for some measurable $H : \boldsymbol{R}^d \times \Theta \to \boldsymbol{R}^{p\otimes p}$ like (3.5) in Example 3.1 below, we can utilize the uniform law of large numbers in order to ensure that

$$\hat{\Sigma}_n := \frac{1}{n}\sum_{i=1}^{n} H(X_{t^n_{i-1}}, \hat{\theta}_n) \xrightarrow{P_0^\eta} \Sigma_0 \tag{2.12}$$

as is the case with (2.11). Of course, $H$ here must be smooth enough to fulfill the conditions corresponding to [U1] to [U3] in Claim B.2 in Appendix B. So, in this case $\Sigma_0$ can be also estimated without specifying $\pi_0$. Unfortunately, it is not clear how to construct an estimator of $\Sigma_0$ in our general setup. For example, $\Sigma_0$ of (3.9) in Example 3.2 below involves integration with respect to $y'$ corresponding to the unobservable variables $(Y'_{t^n_i})_{i=0}^n$, so that we cannot follow the same line as (2.11) and (2.12). We shall not pursue this in this article.

## 3. Wiener-Poisson-driven case

In this section we shall consider $Y$ belonging to a class of martingales. Let the underlying basis equip an $r_w$-dimensional standard Wiener process $w$ and a Poisson random measure $\{\mu(I, E); I \subset \boldsymbol{R}_+, E \subset \boldsymbol{R}^{r_\mu}\backslash\{0\}\}$ with corresponding Lévy measure $\nu$. Suppose that

$$\int_{|z|>1} |z|^q \nu(dz) < \infty \qquad \text{for every } \ q > 0, \tag{3.1}$$

so that the process $J_t = \int_0^t \int z\tilde{\mu}(ds, dz)$, where $\tilde{\mu} = \mu - \nu$, is an $r_\mu$-dimensional zero-mean Lévy process of pure-jump type. Let $M = M^c + M^d$ be an $\boldsymbol{F}$-adapted $d$-dimensional martingale, where

$$M^c_t = \int_0^t \sigma_s dw_s \qquad \text{and} \qquad M^d_t = \int_0^t \zeta_{s-} dJ_s$$

with predictable processes $\sigma = (\sigma^i_j)_{1\le i\le d, 1\le j\le r_w}$ and $\zeta = (\zeta^i_j)_{1\le i\le d, 1\le j\le r_\mu}$. The $X$ considered is given by

$$X_t = X_0 + \int_0^t b(X_s, \theta)ds + M^c_t + M^d_t, \tag{3.2}$$

where $\sigma$ and $\zeta$ may depend on the history of $X$ as well as of $(w, \mu)$, so that the solution to (3.2) may be non-Markovian. We shall implicitly exclude the trivial case, $(\sigma, \zeta) \equiv 0$.

The setting given above is still too general to go forward, and we now set the following *ad-hoc* assumption. Denote by $\|\cdot\|_\infty$ the sup-norm with respect to $\omega \in \Omega$ and $t \in \boldsymbol{R}_+$.

Assumption WP. For $\kappa = \sigma$ and $\zeta$, there exist a finite signed measure $r_\kappa$ on $(-\infty, 0]$, finite $\boldsymbol{F}$-adapted processes $\kappa^{(1)}$ and $\kappa^{(2)}$, and a globally Lipschitz measurable function $F_\kappa : \boldsymbol{R}^d \to \boldsymbol{R}^d$, for which $\kappa$ is represented as

$$\kappa_t = \kappa_t^{(1)} \int_{(-t,0]} F_\kappa(X_{t+u}) r_\kappa(du) + \kappa_t^{(2)}, \tag{3.3}$$

where $\|\kappa^{(1)}\|_\infty < \infty$, $\|\kappa^{(2)}\|_I^{*,q} < \infty$ for every $q > 0$ and compact $I \subset \boldsymbol{R}_+$. Here, $\kappa^{(1)}, \kappa^{(2)}, F_\kappa$ and $r_\kappa$ themselves do not depend on $X$, that is, $\kappa$ depends on $X$ through the function $F_\kappa(\cdot)$ only.

The underlying stochastic basis here is the enlarged Wiener-Poisson space on which the initial process $(X_t)_{t\in[-D\delta,0]}$ is attached. We then obtain:

Lemma 3.1. *Suppose that* $\sup_{t\in\boldsymbol{R}_+} \|X_t\|_{L^q(P_0^\eta)} < \infty$ *for every* $q \geq 2$. *Then Assumptions* 2 *with* $\Delta_{q,n} = h_n$ $(q \geq 2)$ *and* 3 *are implied by Assumption WP.*

See Appendix E for the proof of Lemma 3.1. Given $h_n$, it is easy to check the remaining conditions on the decreasing rate of $h_n$ in Theorem (b). The reason why we set the $L^q(P_0^\eta)$-boundedness of $X$ is that the (functional-type) Lipschitz structure of the coefficients and (3.1) are not sufficient to obtain the boundedness. A simple example of $\kappa$ of (3.3) is $\kappa_t = \sum_{l=0}^D F_\kappa(X_{t-l\delta})$ for some constant $\delta > 0$ and $D \in \boldsymbol{N}$.

Assumption 5 can be easily checked for each given $b(x,\theta)$. As for Assumption 4, we do not know any general answer, and we here go no further than mentioning how to verify it in two examples (Examples 3.1 and 3.2 below). Turning to Assumption 6, (2.6) is automatic whereas (2.7) is left to be considered. The isometry of Itô's stochastic integrals and the orthogonality between $M^c$ and $M^d$ yield

$$v_i^n = \int_{t_{i-1}^n}^{t_i^n} E_0^\eta[\sigma_s^{\otimes 2} | \mathscr{F}_{t_{i-1}^n}] ds + \int_{t_{i-1}^n}^{t_i^n} E_0^\eta \left[ \zeta_s \left( \int z^{\otimes 2} \nu(dz) \right) \zeta_s^\top \,\middle|\, \mathscr{F}_{t_{i-1}^n} \right] ds, \tag{3.4}$$

where $v^{\otimes 2} = vv^\top$. The expression (3.4) implies that, in order to identify $\Sigma_0$ we should furthermore get knowledge about $\kappa^{(l)}$, $F_\kappa$, and $r_\kappa$ to some extent. So it is not clear whether Assumption WP and Assumption 4 are sufficient for (2.7) in general. Nevertheless, in Markovian cases we have the following (see Appendix E for the proof):

Lemma 3.2. *Suppose that $X$ is a diffusion with jumps, say $\sigma_s = \sigma(X_s)$ and $\zeta_s = \zeta(X_s)$, where $\sigma$ and $\zeta$ are $C^1$-functions on $\boldsymbol{R}^d$ with bounded partial derivatives. If Assumption* 4 *is additionally met, Assumption* 6 *holds true with*

$$\begin{aligned} \Sigma_0 = &\int \nabla_\theta b(x,\theta_0)^\top \sigma(x)^{\otimes 2} \nabla_\theta b(x,\theta_0) \pi_0(dx) \\ &+ \int \nabla_\theta b(x,\theta_0)^\top \zeta(x) \left( \int z^{\otimes 2} \nu(dz) \right) \zeta(x)^\top \nabla_\theta b(x,\theta_0) \pi_0(dx). \end{aligned} \tag{3.5}$$

We end this section with two examples.

*Example* 3.1. (Markovian case) Let $X$ be given by

$$dX_t = b(X_t)dt + \sigma(X_t)dw_s + \int \zeta(X_{t-}, z)\tilde{\mu}(dt, dz). \tag{3.6}$$

The appropriate regularity conditions on the coefficients $(b, \sigma, \zeta)$ as well as on the Lévy measure $\nu$ (here possibly $\int_{|z|\le 1} |z|\nu(dz) = \infty$) may lead to Assumption 4 and the $L^q(P_0^\eta)$-boundedness of $X$ (see Masuda (2004b, 2005) for details). If $\zeta(x, z) = \zeta(x)z$ (though this is not essential), $\Sigma_0$ is given by (3.5). When $X$ is a Lévy-driven Ornstein-Uhlenbeck process (i.e., $b(x, \theta) = \theta_1 - \theta_1 x$, $\sigma(x)$ is a constant non-negative definite matrix, and $\zeta(x, z) = z$), we can identify $\pi_0$, which is necessarily operator-selfdecomposable, and it is also possible to give $\pi_0$ beforehand; we know that all the Assumptions are then fulfilled as soon as $\int |x|^p \eta(dx) < \infty$ for every $p > 0$ (see Masuda (2004a) and references therein). For model (3.6), Shimizu and Yoshida (2005) proved the first order efficiency of an approximate maximum-likelihood type estimate in case where $\nu(\boldsymbol{R}^{r_\mu}) < \infty$ and the probability of occurrence of small jumps is small in a certain sense.

*Example* 3.2. (Filtering model; a non-Markovian case) Here is an example in which $(X, Y')$ is ergodic for some process $Y'$. This example is a continuous-time analogue of autoregressive processes with autocorrelated error in the context of time-series analysis. As mentioned in Remark 3, in this case we can get the ergodic theorem for $X$: taking $F(x, y') = F_1(x)$ in (2.5) for any $\pi_0$-integrable $F_1$ and denoting by $\pi_0^X(dx) = \pi_0(dx, \boldsymbol{R}^{d_2})$ the invariant probability measure of $X$, where $d_2$ is the dimension of $Y'$ and $\pi_0$ the invariant probability measure of the process $(X, Y')$, we have

$$T^{-1} \int_0^T F_1(X_t^1)dt \xrightarrow{P_0^\eta\text{-a.s.}} \pi_0^X(F_1). \tag{3.7}$$

Now let $Z^1$ and $Z^2$ be two zero-mean Lévy processes taking values respectively in $\boldsymbol{R}^{r_1}$ and $\boldsymbol{R}^{r_2}$ and admitting moments of any order at time 1 (hence at any $t \in \boldsymbol{R}_+$). Let $Y'_t = \int_0^t \zeta(Y'_{s-})dZ_s^2$ with some $\zeta : \boldsymbol{R}^{d_2} \to \boldsymbol{R}^{d_2 \otimes r_2}$. Suppose then $(X, Y')$ satisfies

$$\begin{pmatrix} dX_t \\ dY'_t \end{pmatrix} = \begin{pmatrix} b(X_t, \theta) \\ 0 \end{pmatrix} dt + \begin{pmatrix} \xi(X_{t-}, Y'_{t-}) & B\zeta(Y'_{t-}) \\ 0 & \zeta(Y'_{t-}) \end{pmatrix} \begin{pmatrix} dZ_t^1 \\ dZ_t^2 \end{pmatrix} \tag{3.8}$$

for some $b : \boldsymbol{R}^{d_1} \times \Theta \to \boldsymbol{R}^{d_1}$, $\xi : \boldsymbol{R}^{d_1} \times \boldsymbol{R}^{d_2} \to \boldsymbol{R}^{d_1 \otimes r_1}$, and constant matrix $B \in \boldsymbol{R}^{d_1 \otimes d_2}$. We suppose that the available data is not $(X_{t_i^n}, Y'_{t_i^n})_{i=0}^n$, but $(X_{t_i^n})_{i=0}^n$ only. In this case the target $X$ is $dX_t = b(X_t, \theta)dt + dY_t$ with $d_1$-dimensional martingale $Y = \int_0^\cdot \xi(X_{s-}, Y'_{s-})dZ_s^1 + BY'$; $X$ is not necessarily Markovian, while $(X, Y')$ is. With good property of the coefficient of $(dZ^1, dZ^2)$ that induces the irreducibility of $(X, Y')$, the result of Masuda (2004b, 2005) may be used to

verify Assumption 4, and Theorem (b) may apply. Especially, if the coefficient of $(dZ^1, dZ^2)$ is smooth enough to yield an expansion of the form

$$E_0^\eta[(Y_{t_i^n} - Y_{t_{i-1}^n})^{\otimes 2} \mid \mathscr{F}_{t_{i-1}^n}] = h_n \Upsilon(X_{t_{i-1}^n}, Y'_{t_{i-1}^n}) + \bar{o}_{P_0^\eta}(h_n)$$

for some $\Upsilon : \boldsymbol{R}^{d_1} \times \boldsymbol{R}^{d_2} \to \boldsymbol{R}^{d_1 \otimes d_1}$ (see the first paragraph of the Appendix for the definition of $\bar{o}_{P_0^\eta}(\cdot)$), then Theorem (c) may also apply with

$$\Sigma_0 = \iint \nabla_\theta b(x, \theta_0)^\top \Upsilon(x, y') \nabla_\theta b(x, \theta_0) \pi_0(dx, dy'). \tag{3.9}$$

## 4. The effect of data frequency: a numerical example

In this section we look at finite-sample behaviors of TFE and LSE for different decreasing rates of $h_n$ in a one-dimensional Markovian case, which belongs to the Wiener-Poisson-driven case discussed in the previous section.

Let $Z$ be a non-skewed and centered normal inverse Gaussian Lévy motion with $\mathscr{L}(Z_1) = NIG(\alpha, 0, \delta, 0)$, where $\alpha$ and $\delta$ are positive constants; see e.g. Masuda (2002) and references therein for the details of $NIG(\alpha, \beta, \delta, \mu)$-distribution. Here $Z$ is of pure-jump type with divergent symmetric Lévy measure $\nu$ such that $\int_{|z| \le 1} |z| \nu(dz) = \infty$, implying that $Z$ is of infinite variation on every compact time intervals. We know that $E[|Z_t|^q] < \infty$ for every $t \in \boldsymbol{R}_+$ and $q > 0$. Then consider $X$ given by

$$dX_t = -\theta X_t dt + \left(1 + \frac{1}{1 + X_t^2}\right) dZ_t. \tag{4.1}$$

Suppose $\theta_0 > 0$. If $nh_n^2 = o(1)$, all the required assumptions are fulfilled with $\eta$ satisfying $\int |x|^q \eta(dx) < \infty$ for every $q > 0$, $\Delta_{q,n} = h_n$ (for every $q \ge 2$); also we have

$$\Sigma_0 = \left(\int z^2 \nu(dz)\right) \left\{\int x^2 (x^2 + 2)^2 (1 + x^2)^{-2} \pi_0(dx)\right\}, \qquad \Gamma(\theta_0) = \int x^2 \pi_0(dx).$$

The estimates are explicitly given by

$$\begin{aligned} \tilde{\theta}_n &= -\frac{1}{h_n} \log\left(\frac{\sum_{i=1}^n X_{t_i^n} X_{t_{i-1}^n}}{\sum_{i=1}^n X_{t_{i-1}^n}^2}\right) \qquad \text{and} \\ \hat{\theta}_n &= -\frac{1}{h_n}\left(\frac{\sum_{i=1}^n X_{t_i^n} X_{t_{i-1}^n}}{\sum_{i=1}^n X_{t_{i-1}^n}^2} - 1\right). \end{aligned} \tag{4.2}$$

Note that $nh_n^2 = O(1)$ is sufficient for the tightness of the estimates; take $q' = 2$ in Theorem (b).

For simulation we set $\theta_0 = 3$, $(\alpha, \delta) = (3, 3)$, and $X_0 = 0$, in which case $\mathrm{Var}[Z_t] = t$ though $\mathscr{L}(Z_t)$ has much heavier tails than $N(0, t)$, the case of Wiener process; specifically, the density behaves as $|x|^{-3/2} \exp(-3|x|)$ for $|x| \to \infty$. Also

we set $h_n = \Delta_i^n t = \Delta n^{-\gamma}$ for every $i$ and some $\gamma > 0$ and $\Delta > 0$. The sample paths of $X$ were simulated via the Euler scheme (cf. Jacod and Protter (1998, Section 6)) with generating mesh in each simulation being $h_n/100$. At each stage we simulated 1000 independent trajectories of $X$, and then computed the mean, standard deviation, and sample MSE of the estimates; it seems quite hard to give theoretical expressions for the bias and MSE of the estimates. We chose $\gamma = 0.3, 0.5$ and $0.8$ for the decreasing rate of $h_n$. Note that for $\gamma = 0.3$ we do not know whether the tightness holds true or not in our context, while the weak consistency is valid as soon as $h_n \to 0$. Also, only $\gamma = 0.8$ leads to the asymptotic normality.

The results are given in Table 1. In all trials except for the starred case, the quantity $\sum_{i=1}^n X_{t_i^n} X_{t_{i-1}^n} / \sum_{i=1}^n X_{t_{i-1}^n}^2$ was positive so that the corresponding TFE were indeed well-defined; in the case of $(\gamma, \Delta, n) = (0.3, 5, 500)$, there was 11 exceptions, so we reported the remaining independent 989 estimates. It is clear from the table that TFE remarkably dominates LSE for $\gamma = 0.3$, whereas LSE becomes better as $\gamma$ increases. TFE (resp. LSE) seems convenient when the observation times are moderately sparse (resp. dense), although there is no

Table 1. TFE $\tilde{\theta}_n$ (left) and LSE $\hat{\theta}_n$ (right) for (4.1). The true value is $\theta_0 = -3$.

| $\gamma$ | $\Delta$ | $n$ | Simulated mean | | S.D. | | Sample MSE | |
|---|---|---|---|---|---|---|---|---|
| 0.3 | 0.5 | 500 | −3.0410 | −2.7042 | 0.6383 | 0.5661 | 0.1677 | 0.1902 |
| | | 1000 | −3.0358 | −2.7611 | 0.5477 | 0.4973 | 0.0913 | 0.1183 |
| | 1.0 | 500 | −3.0584 | −2.4302 | 0.5801 | 0.4565 | 0.1167 | 0.3680 |
| | | 1000 | −3.0180 | −2.5084 | 0.4882 | 0.4030 | 0.0572 | 0.2680 |
| | 3.0 | 500 | −3.0801 | −1.6287 | 0.6283 | 0.3012 | 0.1622 | 1.8886 |
| | | 1000 | −3.0305 | −1.8014 | 0.4897 | 0.2755 | 0.0584 | 1.4425 |
| | 5.0 | 500* | −3.1872 | −1.1671 | 0.8722 | 0.2412 | 0.6136 | 3.3628 |
| | | 1000 | −3.0789 | −1.3545 | 0.5937 | 0.2214 | 0.1305 | 2.7102 |
| 0.5 | 0.5 | 500 | −3.1300 | −3.0183 | 0.8203 | 0.7901 | 0.4697 | 0.3900 |
| | | 1000 | −3.0923 | −3.0154 | 0.7579 | 0.7386 | 0.3384 | 0.2978 |
| | 1.0 | 500 | −3.0600 | −2.8550 | 0.7092 | 0.6606 | 0.2565 | 0.2114 |
| | | 1000 | −3.0579 | −2.9125 | 0.6257 | 0.5955 | 0.1566 | 0.1334 |
| | 3.0 | 500 | −3.0398 | −2.4914 | 0.5757 | 0.4681 | 0.1114 | 0.3067 |
| | | 1000 | −3.0176 | −2.6219 | 0.5028 | 0.3781 | 0.0642 | 0.1788 |
| | 5.0 | 500 | −3.0340 | −2.1970 | 0.5704 | 0.4038 | 0.1070 | 0.6715 |
| | | 1000 | −3.0135 | −2.3946 | 0.4798 | 0.3774 | 0.0532 | 0.3868 |
| 0.8 | 0.5 | 500 | −3.7232 | −3.6916 | 1.4578 | 1.4441 | 5.0394 | 4.8270 |
| | | 1000 | −3.5976 | −3.5809 | 1.4094 | 1.4022 | 4.3026 | 4.2035 |
| | 1.0 | 500 | −3.3340 | −3.2896 | 1.1629 | 1.1452 | 1.9404 | 1.8039 |
| | | 1000 | −3.3314 | −3.3064 | 1.1193 | 1.1101 | 1.6793 | 1.6256 |
| | 3.0 | 500 | −3.1134 | −3.0098 | 0.8446 | 0.8160 | 0.5217 | 0.4436 |
| | | 1000 | −3.1296 | −3.0695 | 0.8038 | 0.7879 | 0.4343 | 0.3902 |
| | 5.0 | 500 | −3.0885 | −2.9170 | 0.7465 | 0.7054 | 0.3170 | 0.2544 |
| | | 1000 | −3.0832 | −2.9883 | 0.6947 | 0.6730 | 0.2398 | 0.2052 |

measure in practice for "sparsity" of sampling schemes which should be carefully determined by taking the data characteristic in question into account.

## Appendix: Proof of Theorem

Write $\theta = (\theta_a)_{a=1}^p$, and $\xi^k$ for the $k$-th component of any random vector $\xi$. In order to avoid possible misreading of gradient operators, we specifically write $\nabla_{[1]} B(x(\theta),\theta) = \nabla_x B(x,\theta)|_{x=x(\theta)}$ and $\nabla_{[2,a]} B(x(\theta),\theta) = \nabla_{\theta_a} B(x,\theta)|_{x=x(\theta)}$ for any function of $\theta$ of the form $(x(\theta),\theta) \mapsto B(x(\theta),\theta)$. For a sequence of random variables $(H_n(\theta))$ depending on $\theta$ and for a nonrandom positive sequence $(a_n)$, we write $H_n(\theta) = \bar{o}_{P_0^\eta}(a_n)$ and $H_n(\theta) = \bar{O}_{P_0^\eta}(a_n)$ if $\sup_{\theta\in\Theta} E_0^\eta[|H_n(\theta)/a_n|^q] = o(1)$ and $\sup_{\theta\in\Theta} E_0^\eta[|H_n(\theta)/a_n|^q] = O(1)$ for every $q > 0$, respectively. In what follows we shall use Taylor's formula and Hölder and Gronwall-Bellman inequalities without notice.

**Appendix** A: "Elementary properties of the auxiliary function"

We begin with preparing simple almost sure expansions concerning auxiliary sequence $(\bar{X}_{i,t_i^n}(\theta))_{i=1}^n$, which later enable us to unify the proofs for TFE and LSE.

Lemma A.1. *Under Assumptions* 1 *and* 2, *we have*

$$\bar{X}_{i,t_i^n}(\theta) = X_{t_{i-1}^n} + b(X_{t_{i-1}^n},\theta)\Delta_i^n t + \bar{O}_{P_0^\eta}(h_n^2), \tag{A.1}$$

$$\nabla_{\theta_a}\bar{X}_{i,t_i^n}(\theta) = \nabla_{\theta_a} b(X_{t_{i-1}^n},\theta)\Delta_i^n t + \bar{O}_{P_0^\eta}(h_n^2), \tag{A.2}$$

$$\nabla^2_{\theta_b\theta_a}\bar{X}_{i,t_i^n}(\theta) = \nabla^2_{\theta_b\theta_a} b(X_{t_{i-1}^n},\theta)\Delta_i^n t + \bar{O}_{P_0^\eta}(h_n^2), \tag{A.3}$$

*$P_0^\eta$-a.s., for every $i \in \{1,\dots,n\}$, $a,b \in \{1,\dots,p\}$, and $\theta \in \Theta$.*

Proof. Since $\sup_{\theta\in\Theta, t\in[t_{i-1}^n,t_i^n)} |\bar{X}_{i,t}(\theta)| \lesssim (1+|X_{t_{i-1}^n}|)$, $P_0^\eta$-a.s., it follows from the definition (2.1) and Assumption 1 that $\bar{X}_{i,t_i^n}(\theta) = X_{t_{i-1}^n} + b(X_{t_{i-1}^n},\theta)\Delta_i^n t + r_i^n(\theta)(\Delta_i^n t)^2$, where

$$r_i^n(\theta) = \int_0^1\int_0^1 u\{\nabla_{[1]} b(\bar{X}_{i,s}(\theta),\theta) b(\bar{X}_{i,s}(\theta),\theta)\}\,|_{s=uvh_n+t_{i-1}^n}\, dudv.$$

Clearly $\sup_{\theta\in\Theta}|r_i^n(\theta)| \lesssim (1+|X_{t_{i-1}^n}|)$, hence we get (A.1). Also we have

$$\begin{aligned}\nabla_{\theta_a}\bar{X}^k_{i,t_i^n}(\theta) &= \int_{t_{i-1}^n}^{t_i^n} [\nabla_{[1]} b^k(\bar{X}_{i,s}(\theta),\theta)][\nabla_{\theta_a}\bar{X}_{i,s}(\theta)]ds \\ &\quad + \int_{t_{i-1}^n}^{t_i^n} \nabla_{[2,a]} b^k(\bar{X}_{i,s}(\theta),\theta)ds, \\ \nabla^2_{\theta_b\theta_a}\bar{X}^k_{i,t_i^n}(\theta) &= \int_{t_{i-1}^n}^{t_i^n} \{[\nabla_{\theta_a}\bar{X}_{i,s}(\theta)]^\top[\nabla^2_{[1]} b^k(\bar{X}_{i,s}(\theta),\theta)][\nabla_{\theta_b}\bar{X}_{i,s}(\theta)]\end{aligned} \tag{A.4}$$

$$
\begin{aligned}
&\quad + [\nabla_{[1]} b^k(\bar{X}_{i,s}(\theta),\theta)][\nabla^2_{\theta_b\theta_a}\bar{X}_{i,s}(\theta)] \\
&\quad + [\nabla_{[1]}\nabla_{[2,a]} b^k(\bar{X}_{i,s}(\theta),\theta)][\nabla_{\theta_b}\bar{X}_{i,s}(\theta)]\}ds \\
&+ \int_{t^n_{i-1}}^{t^n_i} \nabla_{[2,b]}\nabla_{[2,a]} b^k(\bar{X}_{i,s}(\theta),\theta)ds
\end{aligned} \tag{A.5}
$$

for every $k \in \{1,2,\dots,d\}$, from which we obtain $\sup_{\theta\in\Theta, t\in[t^n_{i-1},t^n_i)} |\nabla_{\theta_a}\bar{X}_{i,t}(\theta)| \lesssim h_n(1+|X_{t^n_{i-1}}|)^C$ and $\sup_{\theta\in\Theta, t\in[t^n_{i-1},t^n_i)} |\nabla^2_{\theta_b\theta_a}\bar{X}_{i,t}(\theta)| \lesssim h_n(1+|X_{t^n_{i-1}}|)^C$. From this, the first terms of the right-hand sides of (A.4) and (A.5) can be bounded by a polynomial of $|X_{t^n_{i-1}}|$, uniformly in $\theta \in \Theta$. Then expanding $s \mapsto \nabla_{[2,a]} b^k(\bar{X}_{k,s}(\theta),\theta)$ and $s \mapsto \nabla_{[2,b]}\nabla_{[2,a]} b^k(\bar{X}_{k,s}(\theta),\theta)$ around $t^n_{i-1}$ yields (A.2) and (A.3). □

**Appendix** B: "Proof of (a)"

Define contrast functions associated with TFE and LSE by

$$
K^\Phi_n(\theta) = \frac{1}{nh_n^2}(\Phi_n(\theta) - \Phi_n(\theta_0)) \quad \text{and} \quad K^\Psi_n(\theta) = \frac{1}{nh_n^2}(\Psi_n(\theta) - \Psi_n(\theta_0)),
$$

respectively. We fix $\theta_0 \in \Theta$ arbitrarily in what follows. Define $K_0 : \Theta \to \boldsymbol{R}_+$ by

$$
K_0(\theta) = \int |b(x,\theta) - b(x,\theta_0)|^2 \pi_0(dx). \tag{B.1}
$$

Assumption 5 implies that $K_0(\theta) = 0$ if and only if $\theta = \theta_0$. According to the standard argument (see, e.g., van der Vaart (1998, Chapter 5)), the assertion (a) then follows if we prove $\sup_{\theta\in\Theta} |\Xi^\Lambda_n(\theta)| \xrightarrow{P^\eta_0} 0$ as $n \to \infty$ for $\Lambda = \Phi$ and $\Psi$, where

$$
\Xi^\Lambda_n(\theta) = \frac{1}{nh_n^2}(\Lambda_n(\theta) - \Lambda_n(\theta_0)) - K_0(\theta).
$$

(Note that we have $K^\Phi_n(\tilde{\theta}_n) \le K^\Phi_n(\theta_0) + o_{P^\eta_0}(1)$ and $K^\Psi_n(\hat{\theta}_n) \le K^\Psi_n(\theta_0) + o_{P^\eta_0}(1)$.) However, by the definitions of $\Phi_n$ and $\Psi_n$ we have

$$
\sup_{\theta\in\Theta} |\Xi^\Phi_n(\theta)| \lesssim \sup_{\theta\in\Theta} |\Xi^\Psi_n(\theta)| + I^1_n + I^2_n, \tag{B.2}
$$

where, writing $\chi^n_i(\theta) = X_{t^n_i} - X_{t^n_{i-1}} - b(X_{t^n_{i-1}},\theta)\Delta^n_i t$,

$$
\begin{aligned}
I^1_n &= \sup_{\theta\in\Theta}\left|\frac{1}{nh_n^2}\sum_{i=1}^n (\Delta^n_i t)^4\{|r^n_i(\theta)|^2 - |r^n_i(\theta_0)|^2\}\right|, \\
I^2_n &= \sup_{\theta\in\Theta}\left|\frac{1}{nh_n^2}\sum_{i=1}^n (\Delta^n_i t)^2\{\chi^n_i(\theta_0)^\top r^n_i(\theta_0) - \chi^n_i(\theta)^\top r^n_i(\theta)\}\right|.
\end{aligned}
$$

Now it suffices to show the following claims:

Claim B.1. $\max_{j=1,2} I^j_n \xrightarrow{P^\eta_0} 0$ as $n \to \infty$.

Claim B.2. There exists $p' \geq p'' > p$ such that for every $\theta_1, \theta_2 \in \Theta$: [U1] $\Xi_n^\Psi(\theta_1) \xrightarrow{P_0^\eta} 0$; [U2] $\sup_{n\in\boldsymbol{N}} E_0^\eta[|\Xi_n^\Psi(\theta_1)|^{p'}] \lesssim 1$; and [U3] $\sup_{n\in\boldsymbol{N}} E_0^\eta[|\Xi_n^\Psi(\theta_1) - \Xi_n^\Psi(\theta_2)|^{p'}] \lesssim |\theta_1 - \theta_2|^{p''}$.

Claim B.1 means that it suffices to prove $\sup_{\theta\in\Theta} |\Xi_n^\Psi(\theta)| \xrightarrow{P_0^\eta} 0$, which is slightly simpler to handle than $\sup_{\theta\in\Theta} |\Xi_n^\Phi(\theta)| \xrightarrow{P_0^\eta} 0$, and this uniform convergence is in turn ensured by Claim B.2 (see Ibragimov and Has'minskiĭ (1981, Appendix I. Theorem 20)).

Proof of Claim B.1. We are going to show $E_0^\eta[I_n^j] \to 0$ for $j = 1, 2$. For $I_n^1$, we readily get

$$E_0^\eta[I_n^1] \lesssim \frac{h_n^2}{n} \sum_{i=1}^n E_0^\eta[(1 + |X_{t_{i-1}^n}|)^C] \lesssim h_n^2 = o(1)$$

by Assumptions 1 and 2 together with what we have seen in the proof of Lemma A.1. Next we consider $I_n^2$. Notice that $E_0^\eta[I_n^2] \lesssim c_n^{2,1} + c_n^{2,2}$, where

$$c_n^{2,1} = \frac{1}{n} \sum_{i=1}^n E_0^\eta \left[ |\chi_i^n(\theta_0)| \left( \sup_{\theta\in\Theta} |r_i^n(\theta_0) - r_i^n(\theta)| \right) \right],$$
$$c_n^{2,2} = \frac{h_n}{n} \sum_{i=1}^n E_0^\eta \left[ \left( \sup_{\theta\in\Theta} |\Delta b(t_{i-1}^n; \theta, \theta_0)| \right) \left( \sup_{\theta\in\Theta} |r_i^n(\theta)| \right) \right].$$

It is clear that $c_n^{2,2} \lesssim h_n$. As for $c_n^{2,1}$, we first estimate as

$$c_n^{2,1} \lesssim \frac{1}{n} \sum_{i=1}^n \{E_0^\eta[|\chi_i^n(\theta_0)|^2]\}^{1/2}.$$

On the other hand we have $|\chi_i^n(\theta_0)| \lesssim \int_{t_{i-1}^n}^{t_i^n} |X_s - X_{t_{i-1}^n}| ds + Y_{t_i^n} - Y_{t_{i-1}^n}$, so that

$$\begin{aligned} &\{E_0^\eta[|\chi_k^n(\theta_0)|^2]\}^{1/2} \\ &\quad \lesssim \left\{ h_n \int_{t_{i-1}^n}^{t_i^n} E_0^\eta[|X_s|^2 + |X_{t_{i-1}^n}|^2] ds \right\}^{1/2} + \Delta_{2,n}^{1/2} \lesssim h_n + \Delta_{2,n}^{1/2} \lesssim \Delta_{2,n}^{1/2} \end{aligned}$$

on account of Assumption 2. Therefore we get $c_n^{2,1} \lesssim \Delta_{2,n}^{1/2}$, and this together with $c_n^{2,2} \lesssim h_n$ yields $E_0^\eta[I_n^2] \lesssim \Delta_{2,n}^{1/2} = o(1)$. The proof is complete. □

*Remark* 6. It is possible to show $\max_{j=1,2} I_n^j \xrightarrow{P_0^\eta\text{-a.s.}} 0$ when $\sum_{n=1}^\infty (\Delta_{\alpha q,n})^{1/\alpha} < \infty$ (for some $\alpha > 1$ and sufficiently large $q$). Indeed, in a manner similar to the above we can see that $E_0^\eta[(I_n^1)^q] \lesssim h_n^{2q}$ and $E_0^\eta[(I_n^2)^q] \lesssim (\Delta_{\alpha q,n})^{1/\alpha}$ for every $\alpha > 1$. Therefore Borel-Cantelli lemma yields the claim.

Proof of Claim B.2. We begin with [U1] and [U2]. Fix any $\theta_1 \in \Theta$ and any $p' \in \boldsymbol{N}$ greater than $p$. According to the definition of $\Xi_n^\Psi(\theta)$, simple computations yield $E_0^\eta[|\Xi_n^\Psi(\theta_1)|^{p'}] \lesssim J_n^1(\theta_1) + J_n^2(\theta_1) + J_n^3(\theta_1)$, where

$$J_n^1(\theta_1) = E_0^\eta\left[\left|\frac{1}{t_n^n}\int_0^{t_n^n}|\Delta b(s;\theta_1,\theta_0)|^2 ds - K_0(\theta_1)\right|^{p'}\right],$$

$$J_n^2(\theta_1) = E_0^\eta\left[\left|\frac{1}{nh_n^2}\sum_{i=1}^n(\Delta_i^n t)\chi_i^n(\theta_0)^\top \Delta b(t_{i-1}^n;\theta_1,\theta_0)\right|^{p'}\right],$$

$$J_n^3(\theta_1) = E_0^\eta\left[\left|\frac{1}{nh_n^2}\sum_{i=1}^n(\Delta_i^n t)^2|\Delta b(t_{i-1}^n;\theta_1,\theta_0)|^2 - \frac{1}{t_n^n}\int_0^{t_n^n}|\Delta b(s;\theta_1,\theta_0)|^2 ds\right|^{p'}\right].$$

From Assumption 4 we know that

$$\frac{1}{t_n^n}\int_0^{t_n^n}|\Delta b(s;\theta_1,\theta_0)|^2 ds \xrightarrow{P_0^\eta} K_0(\theta_1).$$

Also, we have $\sup_{n\in\boldsymbol{N}} J_n^1(\theta_1) \lesssim \sup_{n\in\boldsymbol{N}}(t_n^n)^{-1}\int_0^{t_n^n} E_0^\eta[|\Delta b(s;\theta_1,\theta_0)|^{2p'}]ds + 1 \lesssim 1$. Therefore $(t_n^n)^{-1}\int_0^{t_n^n}|\Delta b(s;\theta_1,\theta_0)|^2 ds \xrightarrow{L^{p'}(P_0^\eta)} K_0(\theta_1)$, implying $J_n^1(\theta_1) \to 0$. Next, from the assumptions it follows that

$$\begin{aligned}
J_n^2(\theta_1) &\lesssim E_0^\eta\left[\left|\frac{1}{nh_n^2}\sum_{i=1}^n(\Delta_i^n t)\int_{t_{i-1}^n}^{t_i^n}\Delta b(s,t_{i-1}^n;\theta_0)^\top \Delta b(t_{i-1}^n;\theta_1,\theta_0)ds\right|^{p'}\right] \\
&\quad + E_0^\eta\left[\left|\frac{1}{nh_n^2}\sum_{i=1}^n(\Delta_i^n t)(Y_{t_i^n} - Y_{t_{i-1}^n})^\top \Delta b(t_{i-1}^n;\theta_1,\theta_0)\right|^{p'}\right] \\
&\lesssim \frac{1}{nh_n}\sum_{i=1}^n\int_{t_{i-1}^n}^{t_i^n}(E_0^\eta[|X_s - X_{t_{i-1}^n}|^{\alpha p'}])^{1/\alpha}ds + \epsilon_n \\
&\lesssim (\Delta_{\alpha p',n})^{1/\alpha} + \epsilon_n = o(1).
\end{aligned}$$

In particular, $\sup_{n\in\boldsymbol{N}} J_n^2(\theta_1) \lesssim 1$. As for $J_n^3(\theta_1)$, we first estimate as follows:

$$\begin{aligned}
J_n^3(\theta_1) &\lesssim E_0^\eta\left[\left|\frac{1}{t_n^n}\sum_{i=1}^n|\Delta b(t_{i-1}^n;\theta_1,\theta_0)|^2\Delta_i^n t - \frac{1}{t_n^n}\int_0^{t_n^n}|\Delta b(s;\theta_1,\theta_0)|^2 ds\right|^{p'}\right] \\
&\quad + E_0^\eta\left[\left|\frac{1}{t_n^n}\sum_{i=1}^n\left(\frac{t_n^n\Delta_i^n t}{nh_n^2} - 1\right)|\Delta b(t_{i-1}^n;\theta_1,\theta_0)|^2\Delta_i^n t\right|^{p'}\right] \\
&\lesssim \frac{(nh_n)^{p'-1}}{(t_n^n)^{p'}}\sum_{i=1}^n\int_{t_{i-1}^n}^{t_i^n}E_0^\eta[|X_s - X_{t_{i-1}^n}|^{p'}(1+|X_{t_{i-1}^n}|)^C]ds
\end{aligned}$$

$$+ \frac{n^{p'-1}}{(t_n^n)^{p'}} \left( \max_{1\le i\le n} a_i^n \right)^{p'} \sum_{i=1}^n E_0^\eta[(1+|X_{t_{i-1}^n}|)^C]$$

$$\lesssim \sqrt{\Delta_{2p',n}} + \left( \frac{n}{t_n^n} \max_{1\le i\le n} a_i^n \right)^{p'}, \tag{B.3}$$

where we wrote $a_i^n = t_n^n \Delta_i^n t/(nh_n^2) - 1$ and partly used the second one of (2.9). Observe that

$$\frac{n}{t_n^n} \max_{1\le i\le n} a_i^n \lesssim \frac{\max_{1\le i\le n} |\Delta_i^n t - h_n|}{h_n} + \left| \frac{t_n^n - nh_n}{t_n^n} \right| = o(1) + \left| \frac{t_n^n - nh_n}{t_n^n} \right|.$$

Moreover, we have

$$\begin{aligned} \left| \frac{t_n^n - nh_n}{t_n^n} \right| &\le \frac{nh_n}{t_n^n} \frac{|h_n - \min_{1\le i\le n} \Delta_i^n t|}{h_n} \\ &\lesssim O(1) \cdot \frac{\max_{1\le i\le n} |\Delta_i^n t - h_n|}{h_n} = O(1)o(1) = o(1). \end{aligned}$$

Thus, from (B.3) we get $J_n^3(\theta_1) \to 0$, and in particular $\sup_{n\in\boldsymbol{N}} J_n^3(\theta_1) \lesssim 1$. After all we get $E_0^\eta[|\Xi_n^\Psi(\theta_1)|^{p'}] \to 0$ and $\sup_{n\in\boldsymbol{N}} E_0^\eta[|\Xi_n^\Psi(\theta_1)|^{p'}] \lesssim 1$, so the proofs of [U1] and [U2] are complete.

Now we turn to [U3]. For any $\theta_1, \theta_2 \in \Theta$, we have

$$\begin{aligned} &E_0^\eta[|\Xi_n^\Psi(\theta_1) - \Xi_n^\Psi(\theta_2)|^{p'}] \\ &\lesssim E_0^\eta \Bigg[ \Bigg| \frac{1}{nh_n^2} \sum_{i=1}^n \Delta b(t_{i-1}^n; \theta_2, \theta_1)^\top (\Delta_i^n t) \{ 2(X_{t_i^n} - X_{t_{i-1}^n}) \\ &\qquad - (b(X_{t_{i-1}^n}, \theta_1) + b(X_{t_{i-1}^n}, \theta_2)) \} \Bigg|^{p'} \Bigg] + |K_0(\theta_1) - K_0(\theta_2)|^{p'} \\ &= A_n^1(\theta_1, \theta_2) + A_n^2(\theta_1, \theta_2), \qquad \text{say.} \end{aligned}$$

Obviously $A_n^1(\theta_1, \theta_2) \lesssim \{|\theta_1 - \theta_2|^{p'} + |\theta_1 - \theta_2|^{2p'}\} \int (1+|x|)^C \pi_0(dx) \lesssim |\theta_1 - \theta_2|^{p'} + |\theta_1 - \theta_2|^{2p'}$ under the assumptions. At the same time, it follows from (1.1) and the assumptions that

$$\begin{aligned} A_n^2(\theta_1, \theta_2) &\lesssim E_0^\eta \Bigg[ \Bigg| \frac{1}{nh_n^2} \sum_{i=1}^n \Delta b(t_{i-1}^n; \theta_2, \theta_1)^\top (\Delta_i^n t) \Bigg\{ 2 \int_{t_{i-1}^n}^{t_i^n} \Delta b(s, t_{i-1}^n; \theta_0) ds \\ &\qquad + 2(Y_{t_i^n} - Y_{t_{i-1}^n}) \\ &\qquad + (\Delta_i^n t)(\Delta b(t_{i-1}^n; \theta_0, \theta_1) + \Delta b(t_{i-1}^n; \theta_0, \theta_2)) \Bigg\} \Bigg|^{p'} \Bigg] \\ &\lesssim E_0^\eta \left[ \left\{ \frac{1}{nh_n} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} |\theta_1 - \theta_2| (1 + |X_{t_{i-1}^n}|)^C |X_s - X_{t_{i-1}^n}| ds \right\}^{p'} \right] \end{aligned}$$

$$
\begin{aligned}
&+ E_0^\eta\left[\left|\frac{1}{nh_n^2}\sum_{i=1}^n(\Delta_i^n t)\Delta b(t_{i-1}^n;\theta_2,\theta_1)^\top(Y_{t_i^n}-Y_{t_{i-1}^n})\right|^{p'}\right]\\
&+ E_0^\eta\left[\left\{\frac{1}{n}\sum_{i=1}^n|\theta_1-\theta_2|(1+|X_{t_{i-1}^n}|)^C\right\}^{p'}\right]\\
&\lesssim |\theta_1-\theta_2|^{p'}(1+\epsilon_n+1)\lesssim|\theta_1-\theta_2|^{p'},
\end{aligned}
$$

hence we are done. □

**Appendix** C: "Proof of Theorem (b)"

For the proof of $\sqrt{nh_n}$-consistency of TFE and LSE, we prepare the following simple lemma, from which we can again unify the proofs for TFE and LSE.

LEMMA C.1. *Under Assumptions* 1 *and* 2, *we have*

$$\nabla_{\theta_a}\Phi_n(\theta)=\nabla_{\theta_a}\Psi_n(\theta)+\bar{O}_{P_0^\eta}(nh_n^3), \tag{C.1}$$

$$\nabla^2_{\theta_b\theta_a}\Phi_n(\theta)=\nabla^2_{\theta_b\theta_a}\Psi_n(\theta)+\bar{O}_{P_0^\eta}(nh_n^3), \tag{C.2}$$

$P_0^\eta$*-a.s. for every* $a,b\in\{1,\dots,p\}$.

PROOF. For convenience, we denote by $R_\theta(x):\boldsymbol{R}^d\to\boldsymbol{R}^d$ any function indexed by $\theta$ fulfilling $\sup_{\theta\in\Theta}|R_\theta(x)|\lesssim(1+|x|)^C$; under the assumptions, clearly $R_\theta(X_t)=\bar{O}_{P_0^\eta}(1)$ for any $t\in\boldsymbol{R}_+$. Recall that $\bar{X}_{i,t_i^n}(\theta)=X_{t_{i-1}^n}+b(X_{t_{i-1}^n},\theta)\Delta_i^n t+r_i^n(\theta)(\Delta_i^n t)^2$, where all of $\nabla_\theta^j r_i^n(\theta)$, $j=0,1,2$, are $R_\theta(X_{t_{i-1}^n})$. From this and Lemma A.1, we get

$$
\begin{aligned}
\nabla_{\theta_a}\Phi_n(\theta)&=\nabla_{\theta_a}\Psi_n(\theta)-2\sum_{i=1}^n(\Delta_i^n t)^2\nabla_{\theta_a}\chi_i^n(\theta)^\top r_i^n(\theta)+\sum_{i=1}^n(\Delta_i^n t)^4\nabla_{\theta_a}|r_i^n(\theta)|^2\\
&=\nabla_{\theta_a}\Psi_n(\theta)+\bar{O}_{P_0^\eta}(nh_n^3)+\bar{O}_{P_0^\eta}(nh_n^4),
\end{aligned}
$$

where the same way as showing $c_n^{2,1}\lesssim h_n$ in Claim B.1 was used for the middle term. Hence (C.1) follows, and the proof of (C.2) is similar. □

Recall that $\theta_0\in\Theta$ is presupposed, so that by weak consistency we have $\tilde{\theta}_n\in\Theta$ for every $n$ large enough with $P_0^\eta$-probability tending to 1. Taking a subsequence $(\tilde{\theta}_{n_k})$ tending $P_0^\eta$-a.s. to $\theta_0$ and then letting $k$ be sufficiently large, we may set $\nabla_\theta\Phi_n(\tilde{\theta}_n)=0$ $P_0^\eta$-a.s. for $n$ large enough. Thus the usual expansion yields

$$\frac{1}{2nh_n^2}\nabla_\theta^2\Phi_n(\theta_n^*)\sqrt{nh_n}(\tilde{\theta}_n-\theta_0)=-\frac{1}{2\sqrt{nh_n^3}}\nabla_\theta\Phi_n(\theta_0)^\top,$$

where $\theta_n^*$ is a point on the segment connecting $\tilde{\theta}_n$ and $\theta_0$. But Lemma C.1 implies

that

$$\left(\frac{1}{2nh_n^2}\nabla_\theta^2\Psi_n(\theta_n^*)+\bar{o}_{P_0^\eta}(1)\right)\sqrt{nh_n}(\tilde{\theta}_n-\theta_0)$$
$$=-\frac{1}{2\sqrt{nh_n^3}}\nabla_\theta\Psi_n(\theta_0)^\top+\bar{O}_{P_0^\eta}(\sqrt{nh_n^3}). \tag{C.3}$$

We have $\bar{O}_{P_0^\eta}(\sqrt{nh_n^3})=\bar{O}_{P_0^\eta}(1)$ because of the condition $nh_n^3=O(1)$. Also, both of the terms $\bar{o}_{P_0^\eta}(1)$ and $\bar{O}_{P_0^\eta}(1)$ in (C.3) are identically zero in case of LSE.

What is crucial is the following uniform weak law of large numbers:

Claim C.1. $\sup_{\theta\in\Theta}|(2nh_n^2)^{-1}\nabla_\theta^2\Psi_n(\theta)-\Gamma(\theta)|\xrightarrow{P_0^\eta}0$ with $\Gamma(\theta)$ defined by (2.8).

Proof. Fix $\theta\in\Theta$ arbitrarily. For every $a,b\in\{1,\dots,p\}$, we have

$$\begin{aligned}\frac{1}{2nh_n^2}\nabla_{\theta_a\theta_b}^2\Psi_n(\theta)&=\frac{1}{nh_n^2}\sum_{i=1}^n\sum_{j=1}^d(\Delta_i^n t)^2[\nabla_{\theta_a}b^j(X_{t_{i-1}^n},\theta)][\nabla_{\theta_b}b^j(X_{t_{i-1}^n},\theta)]\\&\quad-\frac{1}{nh_n^2}\sum_{i=1}^n\sum_{j=1}^d(\Delta_i^n t)\chi_i^{n,j}(\theta)[\nabla_{\theta_a\theta_b}^2b^j(X_{t_{i-1}^n},\theta)]\\&=H_n^{1,ab}(\theta)+H_n^{2,ab}(\theta),\qquad\text{say.}\end{aligned}$$

Apply Assumption 4 to conclude that $H_n^{1,ab}(\theta)$ tends in $P_0^\eta$-probability to $\Gamma(\theta)^{ab}$, as in the argument concerning $J_n^1(\theta_1)$ in the proof of Claim B.2. Also, by mimicking the proof of $J_n^3(\theta_1)=o(1)$, it is not difficult to show that $H_n^{2,ab}(\theta)=\bar{o}_{P_0^\eta}(1)$. Thus it remains to consider the modulus of continuity of the random field $(H_n^{1,ab}(\theta)-\Gamma(\theta)^{ab})_{\theta\in\Theta}$. Assumption 1 ensures $|\nabla_{\theta_a}b^j(x,\theta_1)\nabla_{\theta_b}b^j(x,\theta_1)-\nabla_{\theta_a}b^j(x,\theta_2)\nabla_{\theta_b}b^j(x,\theta_2)|^{p'}\le(1+|x|)^C|\theta_1-\theta_2|^{p'}$ for any $\theta_1,\theta_2\in\Theta$ and $p'>p$. Therefore we have $E_0^\eta[|H_n^{1,ab}(\theta_1)-H_n^{1,ab}(\theta_2)|^{p'}]+E_0^\eta[|\Gamma(\theta_1)^{ab}-\Gamma(\theta_2)^{ab}|^{p'}]\lesssim|\theta_1-\theta_2|^{p'}$, as desired. □

Now we introduce some notation as follows:

$$\begin{aligned}\lambda_n&=\sqrt{nh_n}(\tilde{\theta}_n-\theta_0),\\\Gamma_n&=\frac{1}{2nh_n^2}\nabla_\theta^2\Psi_n(\theta_n^*),\\\Sigma_n&=-\frac{1}{2\sqrt{nh_n^3}}\nabla_\theta\Psi_n(\theta_0)^\top.\end{aligned}$$

The tightness of $(\lambda_n)$ is implied by existence of a tight sub-subsequence of any subsequence of $(\lambda_n)$; e.g. Kallenberg (1997, Proposition 4.27). Take any subsequence $(n')\subset\boldsymbol{N}$, then the consistency of $\tilde{\theta}_n$ and the continuity of $\theta\mapsto\Gamma(\theta)$ together with Claim C.1 imply that we can find a further subsequence

$(n'') \subset (n')$ along which $\Gamma_{n''} \xrightarrow{P_0^\eta-\text{a.s.}} \Gamma(\theta_0)$. Without loss of generality we may take $(n'')$ as an increasing sequence. By the presupposed non-degeneracy of $\Gamma(\theta_0)$, we may suppose that $\bar{\Gamma}_m := \Gamma_m + \bar{o}_{P_0^\eta}(1)$, which corresponds to the term $(2nh_n^2)^{-1}\nabla_\theta^2\Psi_n(\theta_n^*) + \bar{o}_{P_0^\eta}(1)$ in the left-hand side of (C.3), is bounded and non-degenerate for every $m \in (n'')$ large enough. Taking a tail of the increasing sequence $(n'')$, we may suppose that $(\bar{\Gamma}_m)$ is bounded and non-degenerate uniformly in $m \in (n'')$, so that $\sup_{m\in(n'')} |\bar{\Gamma}_m^{-1}| \lesssim 1$, $P_0^\eta$-a.s. From (C.3) we then have $\lambda_m = \bar{\Gamma}_m^{-1}\{\Sigma_m + O(1)\}$, $m \in (n'')$, $P_0^\eta$-a.s., hence the proof is complete if we prove that the sequence $(\Sigma_n)_{n\in\boldsymbol{N}}$ is tight. This can be seen as follows. We have $\Sigma_n = \Sigma_n^1 + \Sigma_n^2$, where

$$\Sigma_n^1 = \sum_{i=1}^n \frac{\Delta_i^n t}{\sqrt{nh_n^3}} \int_{t_{i-1}^n}^{t_i^n} \nabla_\theta b(X_{t_{i-1}^n}, \theta_0)^\top \Delta b(s, t_{i-1}^n; \theta_0) ds,$$

$$\Sigma_n^2 = \sum_{i=1}^n \frac{\Delta_i^n t}{\sqrt{nh_n^3}} \nabla_\theta b(X_{t_{i-1}^n}, \theta_0)^\top (Y_{t_i^n} - Y_{t_{i-1}^n}).$$

But, under Assumption 2 we see that $E_0^\eta[|\Sigma_n^1|] \lesssim \sqrt{nh_n}(\Delta_{q',n})^{1/q'} = O(1)$ $(q' \geq 2)$, therefore $(\Sigma_n^1)_{n\in\boldsymbol{N}}$ is tight in view of Markov's inequality. This together with Assumption 3 implies the tightness of $\Sigma_n$, completing the proof.

**Appendix** D: "Proof of Theorem (c)"

According to (C.3) and Claim C.1, it suffices to show that $\Sigma_n$ weakly tends to $\mathcal{N}_p(0, \Sigma_0)$ under $P_0^\eta$. First, under the condition of Theorem (c) we notice that

$$E_0^\eta[|\Sigma_n^1|] \lesssim \frac{1}{\sqrt{nh_n}} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \sqrt{\Delta_{2,n}} ds = O(\sqrt{nh_n\Delta_{2,n}}) = o(1),$$

so that $\Sigma_n^1 = o_{P_0^\eta}(1)$. Thus it remains to obtain the central limit theorem for $\Sigma_n^2$. Put $\Sigma_n^2 = \sum_{i=1}^n B_i^n(\theta_0)$. By (2.6), we have $\sum_{i=1}^n B_i^n(\theta_0) = o_{P_0^\eta}(1) + \sum_{i=1}^n C_i^n(\theta_0)$, where

$$C_i^n(\theta_0) = \frac{\Delta_i^n t}{\sqrt{nh_n^3}} \nabla_\theta b(X_{t_{i-1}^n}, \theta_0)^\top (Y_{t_i^n} - Y_{t_{i-1}^n} - m_i^n).$$

By virtue of the central limit theorem for multidimensional martingale-difference triangular arrays (e.g. Shiryaev (1996), p. 543, Theorem 4) combined with the Cramér-Wald device, it suffices to prove the following: (i) $\sum_{i=1}^n E_0^\eta[|C_i^n(\theta_0)|^{2+\delta} \mid \mathscr{F}_{t_{i-1}^n}] \xrightarrow{P_0^\eta} 0$ for some $\delta > 0$; and (ii) $\sum_{i=1}^n E_0^\eta[C_i^n(\theta_0)^{\otimes 2} \mid \mathscr{F}_{t_{i-1}^n}] \xrightarrow{P_0^\eta} \Sigma_0$. But (ii) is ensured by (2.7). On the other hand, under the condition $nh_n^\alpha \to \infty$ for some $\alpha \in (1, 3)$, it is not difficult to see that, for $\delta = 2/(\alpha - 1) > 0$,

$$E_0^\eta\left[\left|\sum_{i=1}^n E_0^\eta[|C_i^n(\theta_0)|^{2+\delta} \mid \mathscr{F}_{t_{i-1}^n}]\right|\right] \lesssim (nh_n^{1+2/\delta})^{-\delta/2} = o(1).$$

Hence we get (i), and the proof is thus achieved.

**Appendix** E: "Proofs of Lemmas 3.1 and 3.2"

Proof of Lemma 3.1. First we consider Assumption 2. Specifically writing $\kappa(X)$ instead of $\kappa$, we see that for every $Y', Y'' \in \mathbb{D}^d$ (the space of all càdlàg functions from $\boldsymbol{R}_+$ to $\boldsymbol{R}^d$),

$$\begin{aligned}|\kappa(Y')_t - \kappa(Y'')_t| &\le \|\kappa^{(1)}\|_\infty \left| \int_{(-t,0]} (F_\kappa(Y'_{t+u}) - F_\kappa(Y''_{t+u}))r_\kappa(du) \right| \\ &\lesssim \int_{(-t,0]} |Y'_{t+u} - Y''_{t+u}||r|(du) \\ &\le \|Y' - Y''\|^*_t, \qquad P_0^\eta\text{-a.s.},\end{aligned}$$

where $\|F\|^*_t := \sup_{s\le t}|F_s|$. Therefore Assumptions 1 and WP imply the existence and uniqueness of the solution process $X$ to (3.2) for every $\theta \in \Theta$, and moreover we know that $X_t$ is $\mathscr{F}_0 \vee \sigma(w_u - w_v, J_u - J_v; u, v \in [0,t])$-measurable for each $t \in \boldsymbol{R}_+$; see, e.g., Protter (1990, Theorem V-7).

Fix a $q \ge 2$ arbitrarily in the sequel. Using the assumptions we get

$$\begin{aligned}\|\kappa\|^{*,q}_{\boldsymbol{R}_+} &\lesssim \sup_{t\in\boldsymbol{R}_+} E_0^\eta \left[ \left| \int_{(-t,0]} F_\kappa(X_{u+t}) r_\kappa(du) \right|^q \right] + 1 \\ &\lesssim \left( 1 + \sup_{s\in\boldsymbol{R}_+} E_0^\eta[|X_s|^q] \right) r_\kappa(-\boldsymbol{R}_+) + 1 \lesssim 1,\end{aligned}$$

hence the $L^q(P_0^\eta)$-boundedness of $M = M^c + M^d$ follows.

We now turn to the estimate (2.4) of Assumption 2. Put

$$g_{q,i}(t) = E_0^\eta[\|X_\cdot - X_{t_{i-1}^n}\|^{*,q}_{(t_{i-1}^n,t]}], \qquad t \in (t_{i-1}^n, t_i^n]. \tag{E.1}$$

For diffusions with jumps such an estimate is rather classical and well known, however, not so straightforward to obtain in our setup.

We shall utilize the following lemma, which is essentially due to Bichteler and Jacod (1983, Lemma (A.14)); we here rephrase it just to note the orders of the upper bounds with respect to $h_n$, all of which are obvious from the original proofs.

Lemma E.1. *Let $q \ge 2$.*
(*a*) *For a $d$-dimensional measurable process $H$, we have*

$$E_0^\eta \left[ \left\| \int_{t_{i-1}^n}^{\cdot} H_s ds \right\|^{*,q}_{(t_{i-1}^n, t_i^n]} \right] \le h_n^{q-1} \int_{t_{i-1}^n}^{t_i^n} E_0^\eta[|H_s|^q] ds \tag{E.2}$$

*for $i = 1, 2, \ldots, n$.*
(*b*) *For a $\boldsymbol{R}^{d\otimes r_w}$-valued predictable process $G$, we have*

$$E_0^\eta \left[ \left\| \int_{t_{i-1}^n}^{\cdot} G_s dw_s \right\|^{*,q}_{(t_{i-1}^n, t_i^n]} \right] \lesssim h_n^{q/2-1} \int_{t_{i-1}^n}^{t_i^n} E_0^\eta[|G_s|^q] ds \tag{E.3}$$

*for* $i = 1, 2, \ldots, n$.
(*c*) *Let* $\mathcal{B}^{r_\mu}$ *denote the* $r_\mu$*-dimensional Borel* $\sigma$*-field. For a* $d$*-dimensional* $\mathcal{F}\otimes\mathcal{B}^{r_\mu}$*-measurable process* $U(s,z) = U(\omega; s, z)$ *defined on* $\Omega \times \boldsymbol{R}_+ \times (\boldsymbol{R}^{r_\mu}\backslash\{0\})$ *such that* $|U(\omega; s, z)| \le \bar{U}_s(\omega)\rho(z)$ *with* $\xi$ *predictable and* $\rho \in L^2(\nu) \cap L^q(\nu)$, *we have*

$$E_0^\eta\left[\left\|\int_{t_{i-1}^n}^{\cdot}\int U(s,z)\tilde{\mu}(ds,dz)\right\|_{(t_{i-1}^n,t_i^n]}^{*,q}\right] \lesssim \int_{t_{i-1}^n}^{t_i^n} E_0^\eta[|\bar{U}_s|^q]ds \tag{E.4}$$

*for* $i = 1, 2, \ldots, n$.

*Remark* 7. The inequalities (E.2) to (E.4) still hold true $P_0^\eta$-a.s. for $E_0^\eta[\cdot]$ replaced by the conditional expectation $E_0^\eta[\cdot \mid \mathcal{F}_{t_{i-1}^n}]$.

Now observe that Assumption 1 and Lemma E.1 yield

$$\begin{aligned}
g_{q,i}(t) &\lesssim E_0^\eta\left[\left\|\int_{t_{i-1}^n}^{\cdot}\Delta b(s, t_{i-1}^n; \theta_0)ds\right\|_{(t_{i-1}^n,t]}^{*,q} + h_n^q|b(X_{t_{i-1}^n}, \theta_0)|^q\right.\\
&\qquad\left. + \left\|\int_{t_{i-1}^n}^{\cdot}\sigma_s dw_s\right\|_{(t_{i-1}^n,t]}^{*,q} + \left\|\int_{t_{i-1}^n}^{\cdot}\zeta_s dJ_s\right\|_{(t_{i-1}^n,t]}^{*,q}\right]\\
&\lesssim h_n^{q-1}\int_{t_{i-1}^n}^{t} g_{q,i}(s)ds + h_n^q\\
&\qquad + h_n^{q/2-1}\int_{t_{i-1}^n}^{t} E_0^\eta[|\sigma_s|^q]ds + \int_{t_{i-1}^n}^{t} E_0^\eta[|\zeta_s|^q]ds.
\end{aligned} \tag{E.5}$$

In view of Assumption WP we see that, for $\kappa = \sigma$ and $\zeta$, and for $s \ge t_{i-1}^n$,

$$\begin{aligned}
|\kappa_s|^q &\lesssim \left\{\int_{-s}^{0}|F_\kappa(X_{s+u})||r_\kappa|(du)\right\}^q + |\kappa_s^{(2)}|^q\\
&\lesssim \int_{-s}^{0}|F_\kappa(X_{s+u})|^q|r_\kappa|(du) + |\kappa_s^{(2)}|^q\\
&\lesssim \int_{-(s-t_{i-1}^n)}^{0}|F_\kappa(X_{s+u}) - F_\kappa(X_{t_{i-1}^n})|^q|r_\kappa|(du) + |F_\kappa(X_{t_{i-1}^n})|^q\\
&\qquad + \int_{-s}^{-(s-t_{i-1}^n)}|F_\kappa(X_{s+u})|^q|r_\kappa|(du) + |\kappa_s^{(2)}|^q\\
&\lesssim \|X - X_{t_{i-1}^n}\|_{(t_{i-1}^n,s]}^{*,q} + |F_\kappa(X_{t_{i-1}^n})|^q\\
&\qquad + \int_{-s}^{-(s-t_{i-1}^n)}|F_\kappa(X_{s+u})|^q|r_\kappa|(du) + |\kappa_s^{(2)}|^q,
\end{aligned}$$

which in turn yields

$$\int_{t_{i-1}^n}^{t} E_0^\eta[|\kappa_s|^q]ds \lesssim \int_{t_{i-1}^n}^{t} g_{q,i}(s)ds + h_n E_0^\eta[|F_\kappa(X_{t_{i-1}^n})|^q]$$

$$
\begin{aligned}
&+ \int_{t_{i-1}^n}^{t_i^n} \int_{-s}^{-(s-t_{i-1}^n)} E_0^\eta[|F_\kappa(X_{s+u})|^q]|r_\kappa|(du)ds \\
&+ \int_{t_{i-1}^n}^{t_i^n} E_0^\eta[|\kappa_s^{(2)}|^q]ds \\
&\lesssim \int_{t_{i-1}^n}^{t} g_{q,i}(s)ds + h_n. \qquad \text{(E.6)}
\end{aligned}
$$

Here note that the term "$\int_{t_{i-1}^n}^t g_{q,i}(s)ds$" in the upper bound of (E.6) appears only when $F_\kappa$ is not identically null. Combine (E.5) and (E.6) to conclude that, for each $i = 1, \dots, n$, $g_{q,i}(t) \lesssim h_n^{q/2}$ if $\zeta \equiv 0$, and otherwise $g_{q,i}(t) \lesssim h_n$. Using (E.3) and (E.4) together with (E.6), we get (2.4) with $\Delta_{q,n} = h_n$ for any $q \geq 2$.

As soon as $nh_n^2 = O(1)$, the condition $\sqrt{nh_n}(\Delta_{q',n})^{1/q'} = O(1)$ in Theorem (b) is met with $q' = 2$.

*Remark* 8. As evidenced by the above, if $\zeta^{(j)} \equiv 0$, $\sigma^{(2)} \equiv 0$ and $\sigma$ is not identically null, then it is possible to obtain a more rapidly decreasing rate of $\Delta_{q,n}$ (indeed depending on $q$ in this case).

Next we consider Assumption 3. Fix any integer $p' > p$ such that $p' \geq 2$, and $\theta_1, \theta_2 \in \Theta$. Write $\mathbf{1}_i(s) = \mathbf{1}_{(t_{i-1}^n, t_i^n]}(s)$. Then, using Assumption 1, $t_n^n \asymp nh_n$, and Lemma E.1, we have

$$
\begin{aligned}
&E_0^\eta \left[ \left| \frac{1}{nh_n^2} \sum_{i=1}^n (\Delta_i^n t) \Delta b(X_{t_{i-1}^n}; \theta_1, \theta_2)^\top (M_{t_i^n} - M_{t_{i-1}^n}) \right|^{p'} \right] \\
&\lesssim (nh_n^2)^{-p'} \left\{ E_0^\eta \left[ \left| \int_0^{t_n^n} \sum_{i=1}^n (\Delta_i^n t) \mathbf{1}_i(s) \Delta b(X_{t_{i-1}^n}; \theta_1, \theta_2)^\top \sigma_s dw_s \right|^{p'} \right] \right. \\
&\qquad \left. + E_0^\eta \left[ \left| \int_0^{t_n^n} \sum_{i=1}^n (\Delta_i^n t) \mathbf{1}_i(s) \Delta b(X_{t_{i-1}^n}; \theta_1, \theta_2)^\top \zeta_s dJ_s \right|^{p'} \right] \right\} \\
&\lesssim (nh_n^2)^{-p'} \left\{ (t_n^n)^{p'/2-1} \right. \\
&\qquad \times \int_0^{t_n^n} E_0^\eta \left[ \left\{ \sum_{i=1}^n \mathbf{1}_i(s) |\Delta_i^n t| |\Delta b(X_{t_{i-1}^n}; \theta_1, \theta_2)^\top \sigma_s| \right\}^{p'} \right] ds \\
&\qquad \left. + \int_0^{t_n^n} E_0^\eta \left[ \left\{ \sum_{i=1}^n \mathbf{1}_i(s) |\Delta_i^n t| |\Delta b(X_{t_{i-1}^n}; \theta_1, \theta_2)^\top \zeta_s| \right\}^{p'} \right] ds \right\} \\
&\lesssim (nh_n)^{-p'} \left\{ (nh_n)^{p'/2-1} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} E_0^\eta[|\Delta b(X_{t_{i-1}^n}; \theta_1, \theta_2)^\top \sigma_s|^{p'}] ds \right.
\end{aligned}
$$

$$
\begin{aligned}
&\qquad\qquad + \sum_{i=1}^{n} \int_{t_{i-1}^n}^{t_i^n} E_0^\eta[|\Delta b(X_{t_{i-1}^n};\theta_1,\theta_2)^\top \zeta_s|^{p'}]ds \Bigg\} \\
&\lesssim (nh_n)^{-p'}\{(nh_n)^{p'/2}|\theta_1-\theta_2|^{p'} + nh_n|\theta_1-\theta_2|^{p'}\} \\
&\lesssim (nh_n)^{-p'/2}|\theta_1-\theta_2|^{p'},
\end{aligned}
$$

hence the first statement of 3 is fulfilled with $\epsilon_n = (nh_n)^{-p'/2}$ and $p'' = p'$. All without distinction, we can get

$$
E_0^\eta\left[\left|\frac{1}{\sqrt{nh_n^3}}\sum_{i=1}^{n}(\Delta_i^n t)(M_{t_i^n} - M_{t_{i-1}^n})\nabla_\theta b(X_{t_{i-1}^n},\theta_0)\right|^{p'}\right] \lesssim 1,
$$

so that the second statement of Assumption 3 also follows. □

Proof of Lemma 3.2. It is enough to consider component-by-component, so that without loss of generality we may set $d = r_w = r_\mu = p = 1$. The boundedness of the derivatives readily yields

$$
\begin{aligned}
\text{(E.7)} \qquad \int_{t_{i-1}^n}^{t_i^n} E_0^\eta[\kappa(X_s)^2 \mid \mathscr{F}_{t_{i-1}^n}]ds &= \kappa(X_{t_{i-1}^n})^2\Delta_i^n t \\
&\quad + 2\int_{t_{i-1}^n}^{t_i^n} E_0^\eta\Bigg[\Bigg\{\int_0^1 [\kappa\kappa'](X_{t_{i-1}^n} + u(X_s - X_{t_{i-1}^n}))du\Bigg\} \\
&\qquad\qquad \times (X_s - X_{t_{i-1}^n}) \;\Bigg|\; \mathscr{F}_{t_{i-1}^n}\Bigg]ds
\end{aligned}
$$

for $\kappa \in \{\sigma, \zeta\}$ with $\kappa'$ denoting the derivative. Hence, writing the last term of (E.7) as $r_i^{\kappa,n}$, we get

$$
\begin{aligned}
&E_0^\eta\left[\left|\frac{1}{nh_n^3}\sum_{i=1}^{n}(\Delta_i^n t)^2\{\nabla_\theta b(X_{t_{i-1}^n},\theta_0)\}^2 r_i^{\kappa,n}\right|^q\right] \\
&\quad\lesssim \frac{1}{nh_n}\sum_{i=1}^{n}\int_{t_{i-1}^n}^{t_i^n} E_0^\eta[|\nabla_\theta b(X_{t_{i-1}^n},\theta_0)|^q \\
&\qquad\qquad \cdot E_0^\eta[\{1 + |X_{t_{i-1}^n}| + |X_s - X_{t_{i-1}^n}|\}^q |X_s - X_{t_{i-1}^n}|^q \mid \mathscr{F}_{t_{i-1}^n}]]ds \\
&\quad\lesssim (\Delta_{\alpha q,n})^{1/\alpha} = o(1)
\end{aligned}
$$

for every $q \geq 2$ and $\alpha > 1$. Therefore the proof is complete if we show that

$$
\begin{aligned}
&\frac{1}{nh_n^3}\sum_{i=1}^{n}(\Delta_i^n t)^3\{\nabla_\theta b(X_{t_{i-1}^n},\theta_0)\}^2\{\sigma(X_{t_{i-1}^n})^2 + K\zeta(X_{t_{i-1}^n})^2\} \\
&\qquad \xrightarrow{P_0^\eta} \int \{\nabla_\theta b(x,\theta_0)\}^2(\sigma(x)^2 + K\zeta(x)^2)\pi_0(dx),
\end{aligned}
$$

where $K = \int |z|^2\nu(dz)$. But this convergence can be verified by virtue of Assumption 4 as before (the proof of Claim B.2), hence the desired claim. □

**Acknowledgements**

The author thanks the anonymous referees for careful reading and valuable comments.

## References

Bichteler, K. and Jacod, J. (1983). Calcul de Malliavin pour les diffusions avec sauts: existence d'une densite dans le cas unidimensionnel. Seminar on probability, XVII, 132–157, Lecture Notes in Math., **986**, Springer, Berlin.

Dietz, H. M. and Kutoyants, Y. A. (1997). A class of minimum-distance estimators for diffusion processes with ergodic properties, *Statist. Decisions*, **15**, 211–227.

Dorogovcev, A. Ja. (1976). The consistency of an estimate of a parameter of a stochastic differential equation, *Theory Probab. Math. Stat.*, **10**, 73–82.

Ibragimov, I. A. and Has'minskiĭ, R. Z. (1981). *Statistical Estimation. Asymptotic Theory*, Springer-Verlag, New York-Berlin.

Jacod, J. and Protter, P. (1998). Asymptotic error distributions for the Euler method for stochastic differential equations, *Ann. Probab.*, **26**, 267–307.

Kallenberg, O. (1997). *Foundations of Modern Probability*, Springer-Verlag, New York.

Kasonga, R. A. (1988). The consistency of a nonlinear least squares estimator from diffusion processes, *Stochastic Process. Appl.*, **30**, 263–275.

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

Masuda, H. (2002). Analytical properties of GIG and GH distributions, *Proc. Inst. Statist. Math.*, **50**, 165–199 (in Japanese).

Masuda, H. (2004a). On multi-dimensional Ornstein-Uhlenbeck processes driven by a general Lévy process, *Bernoulli*, **10**, 1–24.

Masuda, H. (2004b). Ergodicity and exponential $\beta$-mixing bounds for a strong solution of Lévy-driven stochastic differential equations, MHF Preprint Series 2004-19, Kyushu University.

Masuda, H. (2005). Ergodicity and exponential $\beta$-mixing bounds for multidimensional diffusions with jumps (submitted).

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

Prakasa Rao, B. L. S. (1999). *Statistical Inference for Diffusion type Processes*, Arnold, London.

Protter, P. (1990). *Stochastic Integration and Differential Equations. A New Approach*, Springer-Verlag, Berlin.

Shimizu, Y. (2005). M-estimation for discretely observed ergodic diffusion processes with infinite jumps, *Stat. Inference Stoch. Process.* (to appear).

Shimizu, Y. and Yoshida, N. (2005). Estimation of parameters for diffusion processes with jumps from discrete observations, *Stat. Inference Stoch. Process.* (to appear).

Shiryaev, A. N. (1996). *Probability. Second edition*, Springer-Verlag, New York.

van der Vaart, A. W. (1998). *Asymptotic Statistics*, Cambridge University Press, Cambridge.