Simple Estimators for Parametric Markovian Trend of Ergodic Processes Based on Sampled Data

Masuda, Hiroki
Faculty of Mathematics, Kyushu University

出版情報：Journal of the Japan Statistical Society. 35 (2), pp.147-170, 2005-12. 日本統計学会

MHF Preprint Series

Kyushu University
21st Century COE Program
Development of Dynamic Mathematics with High Functionality

Simple estimators for non-linear Markovian trend from sampled data: I. ergodic cases

H. Masuda

MHF 2005-7

(Received February 15, 2005)

Faculty of Mathematics
Kyushu University
Fukuoka, JAPAN

Simple Estimators for Non-linear Markovian Trend from Sampled Data: I. Ergodic Cases

Hiroki Masuda*

Graduate School of Mathematics, Kyushu University.
6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan.

Abstract

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 $\theta \in \Theta \subset \mathbf{R}^p$ is an unknown parameter. We consider estimation of θ when available data from X is discrete, say $(X_{ih_n})_{i=0}^n$ where $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$. Under some regularity conditions including the ergodicity of X, we obtain weak consistency and $\sqrt{nh_n}$-consistency of a trajectory-fitting estimate as well as a least-squares estimate, leaving Y general as much as possible. A Wiener-Poisson-driven setup is particularly discussed as an important special case.

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

1 Introduction

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

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

where $\theta \in \Theta \subset \mathbf{R}^p$, an open bounded convex domain, X_0 is an $\mathscr{F}_0$-measurable random element with $\mathscr{L}(X_0) = \eta$ possibly unknown, $b : \mathbf{R}^d \times \Theta \to \mathbf{R}^d$ is a measurable function, and $Y = (Y_t)_{t \in \mathbf{R}_+}$ is a d-dimensional zero-mean adapted process. Suppose that there exists a true parameter $\theta_0 \in \Theta$ which induces true data we observe, and that, instead of the full trajectory we have discretely sampled data $(X_{t_i^n})_{i=0}^n$, where $t_i^n = ih_n$ with positive bounded sequence (h_n) for which $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$. The purpose of this article is to derive a set of sufficient conditions for the $\sqrt{nh_n}$-consistency of the trajectory-fitting estimator (TFE) and the least-squares estimator (LSE) for θ_0 without any reference to the concrete structure of Y, apart from the ergodic assumption of X and certain moment conditions (Assumption 3 below); the estimation for parameters involved in Y is out of our scope, hence we do not express the Y's possible dependence on θ 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

*e-mail: hiroki@math.kyushu-u.ac.jp

differential equations (in such a case Y may be regarded as an exogenous randomness contaminating the skeleton dynamics $x = (x_t)_{t\in\mathbf{R}_+}$ described by the deterministic system $dx_t = b(x_t, \theta)dt$). Within this setup the estimates are not efficient in general, however, from the practical point of view it is often important to obtain an easy-to-use estimate. This point is the contribution of this article, more precisely, once the model (the structure of Y) is fully specified, the classical one-step improvement together with a "better" estimating function leads to a more efficient estimator.

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

We end this section with some 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 have need of formulating "statistical inference for stochastic differential equations from sampled data"; clearly this is a rather abstract matter because of diversity of the model. Such researches 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 an efficient result in a "smooth" case; in this direction, the reader can consult a great deal of the references cited in Prakasa Rao [14]. The TFE was studied by e.g. Dietz and Kutoyants [2] for continuously observed diffusion processes and by Kasonga [8] for a class of discretely observed diffusions; considering the ergodic Gaussian Ornstein-Uhlenbeck process, one can notice that the condition of Kasonga [8] is not suitable for ergodic cases, so the route we shall take in this paper is different from his whereas the same contrast function is used. The study of the LSE goes back to Dorogovcev [3] and Prakasa Rao [13] also in case of diffusions. Recently inference for processes with jumps based on sampled data draws the attention of statisticians, because of their practical importance for several kinds of realistic data exhibiting accidental big fluctuations. Nevertheless, much less than diffusions has been known so far: Shimizu and Yoshida [18] and Shimizu [17] studied asymptotic normality, the both of them dealing with the cases where the jump part of driving noise is of finite variation. The main result of Shimizu and Yoshida [18] implicitly includes an issue of LSE for pure-jump cases when allowed jumps are of compound Poisson type. Further, beyond the Markovian framework (but still with the Markovian-trend structure), yet no result concerning the discrete sampling has been formulated. Our present result provides a widely applicable $\sqrt{nh_n}$-consistent estimates for the drift coefficients of processes mentioned above having possibly infinitely many jumps on each compact time-interval.

2 Statement of the result

Let $(\Omega, \mathscr{F}, \mathbf{F} = (\mathscr{F}_t)_{t\geq 0}, P)$ be an underlying complete stochastic basis satisfying the usual hypothesis (e.g., Protter [15]), on which a d-dimensional zero-mean $\mathbf{F}$-adapted process $Y = (Y_t)_{t\in\mathbf{R}_+}$ is endowed. Let $\Theta \subset \mathbf{R}^p$ be an open convex domain with compact closure $\overline{\Theta}$, and consider the partly parametrized model X given by (1.1). As described in the Introduction, we have only sampled data $(X_{t_i^n})_{i=0}^n$. Let $\theta_0 \in \Theta$ denote the true value, which induce the true image measure of X associated with the initial distribution η, say P_0^η.

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

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

and then define $\Phi_n(\theta) = \Phi_n(\theta; (X_{t_i^n})_{i=1}^n)$ by

$$\Phi_n(\theta) = \sum_{k=1}^n |X_{t_i^n} - \bar{X}_{k,t_i^n}(\theta)|^2. \tag{2.2}$$

We define an estimate $\tilde{\theta}_n$ for θ_0 by $\Phi_n(\tilde{\theta}_n) = \inf_{\theta\in\overline{\Theta}} \Phi_n(\theta)$. We call $\tilde{\theta}_n$ trajectory-fitting estimate (TFE). We also consider $\Psi_n(\theta) = \Psi_n(\theta; (X_{t_i^n})_{k=1}^n)$ given by

$$\Psi_n(\theta) = \sum_{k=1}^n |X_{t_i^n} - X_{t_{i-1}^n} - h_n b(X_{t_{i-1}^n}, \theta)|^2, \tag{2.3}$$

and similarly define the least-squares estimate (LSE) $\hat{\theta}_n$ by $\Psi_n(\hat{\theta}_n) = \inf_{\theta\in\overline{\Theta}} \Psi_n(\theta)$.

The LSE is convenient when (2.1) cannot be explicitly solvable. There is an obvious connection between $\Phi_n(\theta)$ and $\Psi_n(\theta)$, that is, according to the usual Euler scheme for ordinary differential equations we have

$$\bar{X}_{k,t_i^n}(\theta) = X_{t_{i-1}^n} + h_n b(X_{t_{i-1}^n}, \theta) + O(h_n^2), \quad P_0^\eta\text{-a.s.}, \tag{2.4}$$

for all θ, under rather mild regularity of $(x, \theta) \mapsto b(x, \theta)$ as well as 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 a same asymptotic property up to the first order. In this sense one may bring redundancy of $\tilde{\theta}_n$ to his/her mind. We here just mention that our numerical experiments given in Section 4 later say that, $\tilde{\theta}_n$ may provide better 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 decreasing rates of h_n. So they may have different finite-sample properties, thus possibly affect corresponding one-step estimates when they are constructed.

2.2 Assumptions and main result

Throughout this article we shall use the following notation: $E_0^\eta[f] = \int f dP_0^\eta$ for any measurable function f; C stands for a universal constant independent of n, and $A_n \lesssim B_n$ implies $A_n \le CB_n$ with C possibly varying from line to line; ∇_a denotes the gradient operator with respect to a variable a; $\|F\|_I^{*,q} = \sup_{s\in I} |F_s|^q$ for any interval $I \subset \mathbf{R}_+$, constant $q > 0$, and process F; finally, we denote $\Delta b(t, s; \theta) = b(X_t, \theta) - b(X_s, \theta)$ 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 $C^{2,3}$, the possible derivatives fulfilling $\sup_{\theta\in\Theta} |\nabla_x^k \nabla_\theta^l b(x,\theta)| \lesssim (1+|x|)^C$ and especially $\sup_{x\in\mathbf{R}^d,\theta\in\Theta} |\nabla_x b(x,\theta)| < \infty$.*

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 $q>0$ [1]. Moreover, for every $q>0$ there exists a positive bounded sequence $\Delta_{q,n}=o(1)$ for which*

$$\sup_{1\le i\le n} E_0^\eta\Big[\|X - X_{t_{i-1}^n}\|^{*,q}_{(t_{i-1}^n,t_i^n]}\Big] \le \Delta_{q,n}. \tag{2.5}$$

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

$$E_0^\eta\left[\left|\frac{1}{nh_n}\sum_{i=1}^n \Delta b(X_{t_{i-1}^n};\theta_1,\theta_2)^\top (Y_{t_i^n} - Y_{t_{i-1}^n})\right|^{p'}\right] \le \epsilon_n |\theta_1-\theta_2|^{p''}.$$

Also, it holds that

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

Assumption 4. *It holds that $b(x,\theta)=b(x,\theta')$, π_0-a.e., if and only if $\theta=\theta'$.*

Assumption 5. *There exists an invariant probability measure π_0 (depending on θ_0, but not on η) for which*

$$T^{-1}\int_0^T F(X_t)dt \xrightarrow{P_0^\eta\text{-a.s.}} \pi_0(F) \tag{2.6}$$

as $T\to\infty$ for any π_0-integrable function F on $\mathbf{R}^d$.

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

$$\Gamma(\theta)^{ij} = \sum_{l=1}^d \int \nabla_{\theta_i} b^l(x,\theta)\nabla_{\theta_j} b^l(x,\theta)\pi_0(dx). \tag{2.7}$$

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

Theorem. *Fix any $\theta_0\in\Theta$, and suppose $h_n\to 0$ and $nh_n\to\infty$.*

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

(b) *Further suppose $nh_n^3 = O(1)$ [2], that $nh_n^\alpha\to\infty$ for some constant $\alpha\in(1,3)$, and that there exists $q'>1$ such that $nh_n(\Delta_{q',n})^{2/q'} = O(1)$. Moreover, suppose $\Gamma(\theta_0)$ is non-degenerate. Then $\sqrt{nh_n}(\hat\theta_n-\theta_0)$ and $\sqrt{nh_n}(\hat\theta_n-\theta_0)$ are P_0^η-tight.*

[1] Namely $\sup_{t\in\mathbf{R}_+} \|(X_t,Y_t)\|_{L^q(P_0^\eta)} < \infty$ for every $q>0$

[2] We here use the symbol $O(1)$ also for sequences possibly tending to 0.

2.3 Some remarks

Remark 2.1. Under the assumptions, $\pi_0(f) < \infty$ for every measurable f of at most polynomial growth.

Remark 2.2. From Gronwall's inequality we have

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

Therefore $E_0^\eta[\|Y - Y_{t_{i-1}^n}\|^{*,q}_{(t_{i-1}^n,t_i^n]}] = o(1)$ is sufficient for (2.5), while some further estimates should be made for Y depending on X, as in the case of (3.1) in Section 3.

Remark 2.3. Assumption 5 concerning a "stability" of X is essential in our framework. Such a property of a stochastic process is of independent interest, and closely related to the operator-based ergodic theory. Since we here consider naive estimates in the sense that the precise form of the driving process Y is uninvolved, our result may apply to, for example, the case where X is possibly non-Markovian, but (X, Y') with some process Y' is Markovian. Then we are able to utilize a well-developed stability theory for Markov processes (see the references cited in Masuda [12]) so as to get the ergodic theorem for (X, Y'); then we have the ergodic theorem for X as well through projection-type function $F(x, y') = F_1(x)$ in (2.6). See Example 3.2 and 3.3 in Section 3.

Remark 2.4. If $\Delta_{q,n}$ rapidly decreases, then mimicking the argument of Kasonga [7] one may strengthen the assertion (a) of Theorem to the strong consistency. However, such cases do not occur in the presence of jumps; e.g., if X is a diffusion process with jumps, then $\Delta_{q,n} = O(h_n)$ for every large $q > 0$. Nevertheless, diffusion type processes with Markovian drift coefficient and possibly non-Markovian diffusion coefficient are relevant; cf. Section 3.

Remark 2.5. If a set of additional regularity conditions are in force, we can of course prove the asymptotic normality of $\sqrt{nh_n}(\tilde{\theta}_n - \theta_0)$ and $\sqrt{nh_n}(\hat{\theta}_n - \theta_0)$ of the form

$$\sqrt{nh_n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathscr{L}(P_0^\eta)} N_p\big(0, \Gamma(\theta_0)^{-1}\Sigma_0\Gamma(\theta_0)^{-1,\top}\big) \tag{2.8}$$

for some $\Sigma_0 \in \mathbf{R}^{p\otimes p}$. Needless to say, in this case a more specified structure of Y is required for identifying Σ_0. (2.8) follows from the standard argument of the M-estimation theory. The details are not reported here.

3 Wiener-Poisson-driven case

As a special case let us consider the case where Y belongs to a class of martingales. Suppose that the underlying basis equips an r_w-dimensional standard Wiener process w and a Poisson random measure $\{\mu(I, E); I \subset \mathbf{R}_+, E \subset \mathbf{R}^{r_\mu}\backslash\{0\}\}$ with Lévy measure ν. Suppose that $\int_{|z|>1} |z|^q \nu(dz) < \infty$ for every $q > 0$, so that we can define an r_μ-dimensional zero-mean pure-jump Lévy process $J_t = \int_0^t \int z\tilde{\mu}(ds, dz)$, where $\tilde{\mu} = \mu - \nu$. Let $M = M^c + M^d$ be an $\mathbf{F}$-adapted d-dimensional martingale, where

$$M_t^c = \int_0^t \sigma_s dw_s \quad \text{and} \quad M_t^d = \int_0^t \zeta_s dJ_s$$

with predictable processes $\sigma = (\sigma^i_j)_{1\leq i\leq d, 1\leq j\leq r_w}$ and $\zeta = (\zeta^i_j)_{1\leq i\leq d, 1\leq j\leq r_\mu}$ possibly depending on the history of X as well as of (w,μ), so that the solution X may be a non-Markovian. Thus X considered is given by

$$X_t = X_0 + \int_0^t b(X_s,\theta)ds + M^c_t + M^d_t. \tag{3.1}$$

Here we set the Wiener-Poisson-driven setting for clarity of discussion; it is possible to formulate the result for general martingale Y.

The setting given above is still too general to go forward. Hence we now set the following *ad-hoc* assumption.

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

$$\kappa_t = \kappa^{(1)}_t \int_{(-t,0]} F_\kappa(X_{t+u}) r_\kappa(du) + \kappa^{(2)}_t, \tag{3.2}$$

where $\|\kappa^{(1)}\|_\infty < \infty$ [3], $\|\kappa^{(2)}\|^{*,q}_I < \infty$ for every $q > 0$ and every compact $I \subset \mathbf{R}_+$. Here, $\kappa^{(1)}, \kappa^{(2)}, F_\kappa$ and r_κ themselves do not depend on X, that is, κ depends on X through the function $F_\kappa(\cdot)$ only.

Indeed, we then obtain:

Lemma 3.1. *Suppose that $\sup_{t\in\mathbf{R}_+} \|X_t\|_{L^q(P^\eta_0)} < \infty$ for every $q > 0$, and that $nh_n^2 = O(1)$. Then Assumptions 2 and 3 are implied by Assumption WP.*

See Appendix D for the proof of Lemma 3.1. The reason why we presuppose the $L^q(P^\eta_0)$-boundedness of X is that the (functional-type) Lipschitz structure of the coefficients is not enough to induce the boundedness. A simple example of such a κ is of the form $\kappa_t = \sum_{l=0}^D F_\kappa(X_{t-l\delta})$ for some constant $\delta > 0$ and $D \in \mathbf{N}$, in which case it should be noted that we have to enlarge the underlying stochastic basis in order to equip it with the initial process $(X_t)_{t\in[-D\delta,0]}$.

Assumption 4 can be easily checked for each given $b(x,\theta)$. As for Assumption 5 we do not know the general answer, and we here go no further than mentioning how to verify it in some special cases.

Example 3.1. *Markovian case.* Masuda [12] studied ergodicity as well as exponential β-mixing bound for general diffusions with jumps. Let X be given by

$$dX_t = b(X_t)dt + \sigma(X_t)dw_s + \int \zeta(X_{t-}, z)\tilde{\mu}(dt, dz). \tag{3.3}$$

The appropriate regularity conditions on the coefficients (b, σ, ζ) as well as on the Lévy measure ν may lead to Assumption 5 and the $L^q(P^\eta_0)$-boundedness of X (see Masuda [12] for details). In cases where $\nu(\mathbf{R}^{r_\mu}) < \infty$ and small jumps occur with small probability in an appropriate sense, Shimizu and Yoshida [18] has provided the first order asymptotic behavior of an approximate maximum-likelihood type estimate.

[3] $\|\cdot\|_\infty$ stands for the sup-norm with respect to $\omega \in \Omega$ and $t \in \mathbf{R}_+$.

Example 3.2. *Non-Gaussian linear filtering model.* Here we mention a case where the drift coefficient is time-inhomogeneous [4] while Assumption 5 really holds true, and actually Theorem can apply.

Let $Z = (Z^{1\top}, Z^{2\top})^\top$ be any non-Gaussian $(d_1 + d_2)$-dimensional zero-mean Lévy process with finite moments of any order at time 1 (hence at any $t \in \mathbf{R}_+$). Consider the $(d_1 + d_2)$-dimensional non-Gaussian Ornstein-Uhlenbeck process $X = (X^{1\top}, X^{2\top})^\top$ given by

$$\begin{pmatrix} dX_t^1 \\ dX_t^2 \end{pmatrix} = \left\{ \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} - \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} X_t^1 \\ X_t^2 \end{pmatrix} \right\} dt + \begin{pmatrix} dZ_t^1 \\ dZ_t^2 \end{pmatrix}, \tag{3.4}$$

where the components of $a_i \in \mathbf{R}^{d_i}$ and $A_{ij} \in \mathbf{R}^{d_i \otimes d_j}$ are constants, where all the eigenvalues of $A = (A_{ij}) \in \mathbf{R}^{(d_1+d_2)\otimes(d_1+d_2)}$ have positive real parts. Then Assumptions WP is clearly met. Now suppose that *the available data is* $(X^1_{ih_n})_{i=0}^n$ *only*, and we want to get an estimate for $\theta = (a, A) \in \Theta$. This example corresponds to an estimation problem of a discretely observed continuous-time hidden Markov model with a latent time-inhomogeneous Markov process X^2; the solution X^2 admits a similar expression to (3.6) below. We know that Assumption 5 is fulfilled for every η, and that X is $L^q(P_0^\eta)$-bounded for every $q > 0$ (cf. Masuda [12]). Suppose the initial value $X_0 = (x_0^{1\top}, x_0^{2\top})^\top$ is known.

The point here is that one cannot apply our estimation result directly for X because of the lack of data $(X^2_{ih_n})_{i=1}^n$. Nevertheless, our result can apply in this special case, via a minor modification of Φ_n and Ψ_n as follows. Notice that we can obtain the ergodic theorem for X^1 by taking $F(x_1, x_2) = F_1(x_1)$ in (2.6) for any π_0-integrable F_1:

$$T^{-1} \int_0^T F_1(X_t^1) dt \xrightarrow{P_0^\eta\text{-a.s.}} \pi_0(F_1). \tag{3.5}$$

Write $\exp(-tA) = [K_t^{ij}]_{i,j=1}^2$, where $K_t^{ij} \in \mathbf{R}^{d_i \otimes d_j}$. Then, by using the expression $X_t = e^{-tA} X_0 + \int_0^t e^{-(t-s)A} dZ_s$ and Fubini's theorem (e.g. Protter [15, Theorem IV.45]), we see that

$$\begin{aligned} X_t^1 &= x_0^1 + \int_0^t \left\{ a_1 - A_{11} X_s^1 - A_{12} \left(\sum_{j=1}^2 K_s^{2j} x_0^j + \sum_{j=1}^2 \int_0^s K_{s-u}^{2j} du a_j \right) \right\} ds \\ &\quad + Z_t^1 + \sum_{j=1}^2 \int_0^t \int_u^t K_{s-u}^{2j} ds dZ_u^j. \end{aligned} \tag{3.6}$$

Thus X^1 itself forms a time-inhomogeneous Markov process: temporarily allowing the dependence of $b(x, \theta)$ of (1.1) on t, say $b(x, t, \theta)$, we have

$$b(X_t, t, \theta) = a_1 - A_{11} X_t^1 - A_{12} \left(\sum_{j=1}^2 K_t^{2j} x_0^j + \sum_{j=1}^2 \int_0^t K_{t-u}^{2j} du a_j \right),$$

which is smooth in θ. Since K^{ij} are deterministic, our result may apply with this minor change, providing a simple $\sqrt{nh_n}$-consistent estimate as long as $nh_n^2 = O(1)$ and Z admits moments of any order.

[4] That is, of the form $b(X_t, t, \theta)$ in place of $b(X_t, \theta)$.

See Masuda [10, Section 2] and references therein for fundamental facts of multidimensional Lévy-driven Ornstein-Uhlenbeck processes. See also Masuda [11, Example 2], in which the method of moment was considered for partially observed models in case of $h_n = h > 0$ for every $n \in \mathbf{N}$.

Example 3.3. *Jointly Markovian case.* As mentioned in Remark 2.3, Assumption 5 is met as soon as (X, Y') is ergodic for some process Y'. We here discuss such a case.

Let Z^1 and Z^2 be two d-dimensional zero-mean Lévy processes admitting moments of any order, and let $Y'_t = \int_0^t \zeta(Y'_{s-})dZ^2_s$ with a uniformly elliptic ζ. Suppose (X, Y') fulfills

$$\begin{pmatrix} dX_t \\ dY'_t \end{pmatrix} = \begin{pmatrix} b(X_t, \theta) \\ 0 \end{pmatrix} dt + \begin{pmatrix} 1 & a\zeta(Y'_{t-}) \\ 0 & \zeta(Y'_{t-}) \end{pmatrix} \begin{pmatrix} dZ^1_t \\ dZ^2_t \end{pmatrix}, \tag{3.7}$$

where a is a constant such that $|a| < 1$, which is put just for simplicity of verifying irreducibility of (X, Y'). Then the result of Masuda [12] may apply; for instance, it may be enough that ζ is additionally bounded and that $x^\top b(x, \theta) \le -\epsilon < 0$ outside a compact set with non-empty interior containing the origin. In this case the target X is $dX_t = b(X_t, \theta)dt + dY_t$ with $Y = Z^1 + \int_0^{\cdot} \zeta(Y'_{s-})dZ^2_s$, a d-dimensional martingale. Our result then readily applies. This example is a continuous-time version of autoregressive processes with autocorrelated error in the context of time-series analysis.

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 centred normal inverse Gaussian Lévy motion (NIGLM) [5] with $\mathscr{L}(Z_1) = NIG(\alpha, 0, \delta, 0)$, where α and δ are positive constants and we know that $E[|Z_t|^q] < \infty$ for every $t \in \mathbf{R}_+$ 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 that $\theta_0 > 0$ and that $nh_n^2 = O(1)$. Then all the required assumptions are fulfilled (see Masuda [12] for checking Assumption 5), and the estimates are explicitly given by

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

For simulation we set $\theta_0 = 3$, $(\alpha, \delta) = (3, 3)$, and $X_0 = 0$; in this case $Var[Z_t] = t$, whereas $\mathscr{L}(Z_t)$ has much heavier tails than $N(0, t)$, the case of Wiener process; specifically, the density behaves as $|x|^{-3/2} \exp(-3|x|)$. Also we set $h_n = \Delta n^{-\gamma}$ for $\gamma > 0$ and $\Delta > 0$. The sample paths of X were simulated via the Euler scheme (cf. Jacod and Protter [5, Section 6]) with generating mesh in each simulation being $h_n/100$. At each

[5] The NIGLM possesses no Gaussian component and divergent Lévy measure ν such that $\int_{|z|\le 1} |z|\nu(dz) = \infty$, implying that Z is of infinite variation on every compact time intervals. See Masuda [9] and references therein for the details of $NIG(\alpha, \beta, \delta, \mu)$-distribution.

stage we generate 1000 independent trajectories of X, and then the mean and standard deviation of the estimates were computed. Also computed were sample MSEs obtained from the 1000 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. Here we have $\Delta_{q,n} = O(h_n)$ (for every large $q > 0$), and note that we need $nh_n^2 = O(1)$ for the tightness (take $q' = 2$ in Theorem (b)). Therefore, for $\beta = 0.3$ we do not know whether the tightness holds true or not in our context, while the weak consistency is valid.

The results are given in Table 1, in which the mean, the standard deviation, and sample MSE are reported for TFE (resp. LSE) in the left (resp. right) side in each item; 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 there reported for the remaining independent 989 estimates. Nevertheless it is clear from the table that TFE remarkably dominates LSE for $\gamma = 0.3$, whereas LSE becomes better as γ increases. But, practically, it should be noted that TFE sometimes may not work for too large Δ and small n, as seen in the exceptional case above.

This numerical results says that TFE (resp. LSE) turns out to be workable when the observation times are relatively sparse (resp. dense), although there is no measure in practice for "sparsity" of sampling schemes, which should be carefully determined by taking the data characteristic in question into account. Also, TFE itself provides a good estimate in the linear case.

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

γ	Δ	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

Appendix: Proof of Theorem

Write $\theta = (\theta_a)_{a=1}^p$, and ξ^k for the kth component of any random vector ξ. 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 the form $(x(\theta),\theta) \mapsto B(x(\theta),\theta)$. Denote by $R_\theta(x) : \mathbf{R}^d \to \mathbf{R}^d$ any function indexed by θ such that $\sup_{\theta\in\Theta}|R_\theta(x)| \lesssim (1+|x|)^C$. Below we shall use Taylor's formula and Hölder and Gronwall-Bellman inequalities without notice.

A Elementary properties of the auxiliary function

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

Lemma A.1. *Under Assumption 1, we have*

$$\begin{aligned} \bar{X}_{i,t_i^n}(\theta) &= X_{t_{i-1}^n} + h_n b(X_{t_{i-1}^n},\theta) + h_n^2 R_\theta(X_{t_{i-1}^n}), && \text{(A.1)}\\ \nabla_{\theta_a}\bar{X}_{i,t_i^n}(\theta) &= h_n \nabla_{\theta_a} b(X_{t_{i-1}^n},\theta) + h_n^2 R_\theta(X_{t_{i-1}^n}), && \text{(A.2)}\\ \nabla^2_{\theta_b\theta_a}\bar{X}_{i,t_i^n}(\theta) &= h_n \nabla^2_{\theta_b\theta_a} b(X_{t_{i-1}^n},\theta) + h_n^2 R_\theta(X_{t_{i-1}^n}), && \text{(A.3)} \end{aligned}$$

P_0^η*-a.s., for every* $i \in \{1,\ldots,n\}$, $a,b \in \{1,\ldots,p\}$, *and* $\theta\in\Theta$. *If Assumption 2 is additionally met, the terms* $R_\theta(X_{t_{i-1}^n})$ *in (A.1) to (A.3) are* $L^q(P_0^\eta)$*-bounded for every* $q>0$.

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^η-a.s., it follows from the definition (2.1) and Assumption 1 that $\bar{X}_{i,t_i^n}(\theta) = X_{t_{i-1}^n} + h_n b(X_{t_{i-1}^n},\theta) + h_n^2 r_i(\theta)$, where

$$r_i(\theta) = \int_0^1\int_0^1 u\Big\{\nabla_{[1]}b(\bar{X}_{i,s}(\theta),\theta)b(\bar{X}_{i,s}(\theta),\theta)\Big\}\Big|_{s=uvh_n+t_{i-1}^n} du dv. \qquad \text{(A.4)}$$

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

$$\begin{aligned} \nabla_{\theta_a}\bar{X}^k_{i,t_i^n}(\theta) &= \int_{t_{i-1}^n}^{t_i^n}[\nabla_{[1]}b^k(\bar{X}_{i,s}(\theta),\theta)][\nabla_{\theta_a}\bar{X}_{i,s}(\theta)]ds \\ &\quad + \int_{t_{i-1}^n}^{t_i^n}\nabla_{[2,a]}b^k(\bar{X}_{i,s}(\theta),\theta)ds, && \text{(A.5)}\\ \nabla^2_{\theta_b\theta_a}\bar{X}^k_{i,t_i^n}(\theta) &= \int_{t_{i-1}^n}^{t_i^n}\Big\{[\nabla_{\theta_a}\bar{X}_{i,s}(\theta)]^\top[\nabla^2_{[1]}b^k(\bar{X}_{i,s}(\theta),\theta)][\nabla_{\theta_b}\bar{X}_{i,s}(\theta)] \\ &\quad +[\nabla_{[1]}b^k(\bar{X}_{i,s}(\theta),\theta)][\nabla^2_{\theta_b\theta_a}\bar{X}_{i,s}(\theta)] \\ &\quad +[\nabla_{[1]}\nabla_{[2,a]}b^k(\bar{X}_{i,s}(\theta),\theta)][\nabla_{\theta_b}\bar{X}_{i,s}(\theta)]\Big\}ds \\ &\quad + \int_{t_{i-1}^n}^{t_i^n}\nabla_{[2,b]}\nabla_{[2,a]}b^k(\bar{X}_{i,s}(\theta),\theta)ds && \text{(A.6)} \end{aligned}$$

for every $k\in\{1,2,\ldots,d\}$, from which we obtain $\sup_{\theta\in\Theta,t\in[t_{i-1}^n,t_i^n)}|\nabla_{\theta_a}\bar{X}_{i,t}(\theta)| \lesssim h_n(1+|X_{t_{i-1}^n}|)^C$ and $\sup_{\theta\in\Theta,t\in[t_{i-1}^n,t_i^n)}|\nabla^2_{\theta_b\theta_a}\bar{X}_{i,t}(\theta)| \lesssim h_n(1+|X_{t_{i-1}^n}|)^C$. From this the first terms of the right-hand sides of (A.5) and (A.6) can be bounded by a polynomial of $|X_{t_{i-1}^n}|$. 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_{i-1}^n yields (A.2) and (A.3). The second assertion is now obvious under the assumption. □

B Proof of (a)

Utilizing Lemma A.1, we shall prove the weak consistency of TFE and LSE simultaneously.

Define contrast functions associated with TFE and LSE by $(nh_n^2)^{-1}\{\Phi_n(\theta) - \Phi_n(\theta_0)\}$ and $(nh_n^2)^{-1}\{\Psi_n(\theta) - \Psi_n(\theta_0)\}$, respectively. We fix any $\theta_0 \in \Theta$ in what follows. Define $K_0 : \Theta \to \mathbf{R}_+$ by

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

By virtue of Assumption 4, $K_0(\theta) = 0$ if and only if $\theta = \theta_0$. Hence, according to the standard argument, the assertion (a) follows if we prove $\sup_{\theta\in\Theta} |\Xi_n^\Lambda(\theta)| \xrightarrow{P_0^\eta} 0$ as $n \to \infty$ for both $\Lambda_n = \Phi_n$ and Ψ_n, where

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

as $n \to \infty$. However, by the definitions of Φ_n and Ψ_n it is easy to see that

$$\sup_{\theta\in\Theta} |\Xi_n^\Phi(\theta)| \le \sup_{\theta\in\Theta} |\Xi_n^\Psi(\theta)| + I_n^1 + 2I_n^2, \tag{B.2}$$

where, writing $\chi_i^n(\theta) = X_{t_i^n} - X_{t_{i-1}^n} - h_n b(X_{t_{i-1}^n}, \theta)$,

$$\begin{aligned} I_n^1 &= \sup_{\theta\in\Theta} \left| \frac{h_n^2}{n} \sum_{i=1}^n \left\{|r_i(\theta)|^2 - |r_i(\theta_0)|^2\right\} \right|, \\ I_n^2 &= \sup_{\theta\in\Theta} \left| \frac{1}{n} \sum_{i=1}^n \left\{\chi_i^n(\theta_0)^\top r_i(\theta_0) - \chi_i^n(\theta)^\top r_i(\theta)\right\} \right|. \end{aligned}$$

Thus the proof of (a) is complete if we prove the following claims:

Claim B.1. $\max_{j=1,2} I_n^j \xrightarrow{P_0^\eta\text{-a.s.}} 0$ *as* $n \to \infty$.

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

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 $\sup_{\theta\in\Theta} |\Xi_n^\Psi(\theta)| \xrightarrow{P_0^\eta} 0$, and this uniform convergence is in turn ensured by Claim B.2 (see Ibragimov and Has'minskiĭ [4, Appendix I Theorem 20]). Though we shall prove almost sure convergences in Claim B.1, note that the convergences in P_0^η-probability are enough for our aim.

B.1 Proof of Claim B.1

We are going to show $I_n^j \xrightarrow{P_0^\eta-\text{a.s.}} 0$ for $j=1,2$. Fix a sufficiently large $q \in \mathbf{N}$.

For I_n^1, we readily get

$$E_0^\eta[|I_n^1|^q] \lesssim \frac{h_n^{2q}}{n} \sum_{i=1}^n \left\{ E_0^\eta\left[\left(\sup_{\theta\in\Theta} |R_\theta(X_{t_{i-1}^n})|\right)^{2q}\right]\right\}^{1/2} \lesssim h_n^{2q}$$

by Assumptions 1 and 2 together with what we have seen in the proof of Lemma A.1. Therefore Borel-Cantelli lemma yields $I_n^1 \xrightarrow{P_0^\eta-\text{a.s.}} 0$ because $h_n = o(1)$ and we may let q be arbitrarily large.

Next we consider I_n^2. Since

$$I_n^2 = \sup_{\theta\in\Theta} \left| \frac{1}{n} \sum_{i=1}^n \left\{ \chi_i^n(\theta_0)^\top (r_i(\theta_0) - r_i(\theta)) + h_n \Delta b(t_{i-1}^n; \theta, \theta_0)^\top r_i(\theta) \right\} \right|,$$

we get $E_0^\eta[|I_n^{(2)}|^q] \lesssim c_n^{2,1} + c_n^{2,2}$, where

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

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

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

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

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

Therefore we obtain $c_n^{2,1} \lesssim h_n^q$, so that $E_0^\eta[|I_n^2|^q] \lesssim h_n^q$, hence we obtain $I_n^2 \xrightarrow{P_0^\eta-\text{a.s.}} 0$ as before.

B.2 Proof of Claim B.2

Proofs of [U1] *and* [U2]. Fix any $\theta_1 \in \Theta$ and any $p' \in \mathbf{N}$ 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

$$\begin{aligned}
J_n^1(\theta_1) &= E_0^\eta\left[\left|\frac{1}{n}\sum_{i=1}^n |\Delta b(t_{i-1}^n;\theta_1,\theta_0)|^2 - \frac{1}{nh_n}\int_0^{nh_n} |\Delta b(s;\theta_1,\theta_0)|^2 ds\right|^{p'}\right],\\
J_n^2(\theta_1) &= E_0^\eta\left[\left|\frac{1}{nh_n}\int_0^{nh_n} |\Delta b(s;\theta_1,\theta_0)|^2 ds - K_0(\theta_1)\right|^{p'}\right],\\
J_n^3(\theta_1) &= E_0^\eta\left[\left|\frac{1}{nh_n}\sum_{i=1}^n \chi_i^n(\theta_1)^\top \Delta b(t_{i-1}^n;\theta_1,\theta_0) + K_0(\theta_1)\right|^{p'}\right].
\end{aligned}$$

Observe that

$$\begin{aligned}
J_n^1(\theta_1) &\lesssim \frac{1}{nh_n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} E_0^\eta\big[|\Delta b(s;\theta_1,\theta_0) - \Delta b(t_{i-1}^n;\theta_1,\theta_0)|^{p'}(1+|X_s|+|X_{t_{i-1}^n}|)^C\big]ds\\
&\lesssim \frac{1}{nh_n}\sum_{k=1}^n \int_{t_{i-1}^n}^{t_i^n} (E_0^\eta[|X_s - X_{t_{i-1}^n}|^{2p'}])^{1/2} ds \lesssim \sqrt{\Delta_{2p',n}} = o(1).
\end{aligned}$$

We see $(nh_n)^{-1}\int_0^{nh_n} |\Delta b(s;\theta_1,\theta_0)|^2 ds \xrightarrow{P_0^\eta-\text{a.s.}} K_0(\theta_1)$ under Assumption 5, hence

$$\sup_{n\in\mathbf{N}} J_n^2(\theta_1) \lesssim \sup_{n\in\mathbf{N}} \frac{1}{nh_n}\int_0^{nh_n} E_0^\eta[|\Delta b(s;\theta_1,\theta_0)|^{2p'}]ds + 1 \lesssim 1.$$

Therefore $(nh_n)^{-1}\int_0^{nh_n} |\Delta b(s;\theta_1,\theta_0)|^2 ds \xrightarrow{L^{p'}(P_0^\eta)} K_0(\theta_1)$, hence $J_n^2(\theta_1) \to 0$. As for $J_n^3(\theta_1)$, we have

$$\begin{aligned}
J_n^3(\theta_1) &\lesssim E_0^\eta\left[\left|\frac{1}{nh_n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \Delta b(s,t_{i-1}^n;\theta_0)^\top \Delta b(t_{i-1}^n;\theta_1,\theta_0)ds\right|^{p'}\right]\\
&\quad + E_0^\eta\left[\left|K_0(\theta_1) - \frac{1}{n}\sum_{i=1}^n |\Delta b(t_{i-1}^n;\theta_0,\theta_1)|^2\right|^{p'}\right]\\
&\quad + E_0^\eta\left[\left|\frac{1}{nh_n}\sum_{k=1}^n \Delta b(t_{i-1}^n;\theta_1,\theta_0)^\top (Y_{t_i^n} - Y_{t_{i-1}^n})\right|^{p'}\right]\\
&\lesssim \sqrt{\Delta_{2p',n}} + o(1) + o(1) = o(1),
\end{aligned}$$

where at the last inequality we used Assumption 3 for the third term. After all it follows that $E_0^\eta[|\Xi_n^\Psi(\theta_1)|^{p'}] \to 0$ and $\sup_{n\in\mathbf{N}} E_0^\eta[|\Xi_n^\Psi(\theta_1)|^{p'}] \lesssim 1$. Thus the proofs of [U1] and [U2] are complete.

Proof of [U3]. For any fixed $\theta_1, \theta_2 \in \Theta$, it follows from the assumptions that

$$\begin{aligned}
E_0^\eta\left[|\Xi_n^\Psi(\theta_1) - \Xi_n^\Psi(\theta_2)|^{p'}\right] &\leq E_0^\eta\left[\left|\frac{1}{nh_n}\sum_{i=1}^n \Delta b(t_{i-1}^n;\theta_1,\theta_2)^\top\left\{2(X_{t_i^n} - X_{t_{i-1}^n})\right.\right.\right. \\
&\qquad \left.\left.\left. - h_n\left(b(X_{t_{i-1}^n},\theta_1) + b(X_{t_{i-1}^n},\theta_2)\right)\right\}\right|^{p'}\right] \\
&\lesssim E_0^\eta\left[\left|\frac{1}{nh_n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \Delta b(t_{i-1}^n;\theta_1,\theta_2)^\top \Delta b(s, t_{i-1}^n;\theta_0) ds\right|^{p'}\right] \\
&\qquad + E_0^\eta\left[\left|\frac{1}{nh_n}\sum_{i=1}^n \Delta b(t_{i-1}^n;\theta_1,\theta_2)^\top (Y_{t_i^n} - Y_{t_{i-1}^n})\right|^{p'}\right] \\
&\qquad + E_0^\eta\left[\left|\frac{1}{n}\sum_{i=1}^n \Delta b(t_{i-1}^n;\theta_1,\theta_2)^\top \Delta b(t_{i-1}^n;\theta_0,\theta_1)\right|^{p'}\right] \\
&\qquad + E_0^\eta\left[\left|\frac{1}{n}\sum_{i=1}^n \Delta b(t_{i-1}^n;\theta_1,\theta_2)^\top \Delta b(t_{i-1}^n;\theta_0,\theta_2)\right|^{p'}\right] \\
&\lesssim |\theta_1 - \theta_2|^{p'},
\end{aligned}$$

hence we are done.

C Proof of Theorem (b)

We turn to the proof of $\sqrt{nh_n}$-consistency of TFE and LSE. Let $\bar{o}_{P_0^\eta}(\cdot)$ and $\bar{O}_{P_0^\eta}(\cdot)$ stand for the stochastic order symbols valid uniformly in $\theta \in \Theta$. First we prepare the following simple lemma, from which we can again unify the proof 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^η*-a.s. for every* $a, b \in \{1, \ldots, p\}$.

Proof. In view of Lemma A.1, it is easy to see that

$$\begin{aligned}
\nabla_{\theta_a}\Phi_n(\theta) &= \nabla_{\theta_a}\Psi_n(\theta) - 2h_n^2\sum_{i=1}^n \chi_i^n(\theta)^\top R_\theta(X_{t_{i-1}^n}) \\
&\qquad + 2h_n^3\sum_{i=1}^n R_\theta(X_{t_{i-1}^n})^\top \nabla_{\theta_a} b(X_{t_{i-1}^n},\theta) + 2h_n^4\sum_{i=1}^n R_\theta(X_{t_{i-1}^n})^\top R_\theta(X_{t_{i-1}^n}) \\
&= \nabla_{\theta_a}\Psi_n(\theta) + h_n^2\bar{O}_{P_0^\eta}(nh_n) + h_n^3\bar{O}_{P_0^\eta}(n) + h_n^4\bar{O}_{P_0^\eta}(n) \\
&= \nabla_{\theta_a}\Psi_n(\theta) + \bar{O}_{P_0^\eta}(nh_n^3),
\end{aligned}$$

where we used $c_n^{2,1} \lesssim h_n^q$ (with $q=1$) proved in Section B.1. Similarly,

$$\begin{aligned}
\nabla^2_{\theta_b\theta_a}\Phi_n(\theta) &= \nabla^2_{\theta_b\theta_a}\Psi_n(\theta) - 2h_n^2\sum_{i=1}\chi_i^n(\theta)^\top R_\theta(X_{t_{i-1}^n}) \\
&\quad + 2h_n^3\sum_{i=1}\big\{R_\theta(X_{t_{i-1}^n})^\top[\nabla_{\theta_a}b(X_{t_{i-1}^n},\theta)+\nabla_{\theta_b}b(X_{t_{i-1}^n},\theta)] \\
&\qquad + R_\theta(X_{t_{i-1}^n})^\top[\nabla^2_{\theta_b\theta_a}b(X_{t_{i-1}^n},\theta)]\big\} \\
&\quad + 2h_n^4\sum_{i=1}\big\{R_\theta(X_{t_{i-1}^n})^\top R_\theta(X_{t_{i-1}^n}) + R_\theta(X_{t_{i-1}^n})^\top R_\theta(X_{t_{i-1}^n})\big\} \\
&= \nabla^2_{\theta_b\theta_a}\Psi_n(\theta) + \bar{O}_{P_0^\eta}(nh_n^3),
\end{aligned}$$

hence the result. ☐

Recall that $\theta_0 \in \Theta$ is presupposed, hence by the weak consistency $\tilde{\theta}_n \in \Theta$ for every n large enough with P_0^η-probability tending to 1. Taking a subsequence $(\tilde{\theta}_{n_k})$ tending P_0^η-a.s. to θ_0 and then letting k sufficiently large, we may set $\nabla_\theta\Phi_n(\tilde{\theta}_n)=0$, P_0^η-a.s. for n large enough. Thus, from Lemma C.1 and the usual expansion we have

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

where θ_n^* is a point on the segment connecting $\tilde{\theta}_n$ and θ_0, and we now regard the gradients as column vectors. But Lemma C.1 implies that

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

under the condition $nh_n^3 = O(1)$. From Lemma C.1 we know that the term $\bar{o}_{P_0^\eta}(1)$ and $\bar{O}_{P_0^\eta}(1)$ in (C.3) are $L^q(P_0^\eta)$-bounded for every $q>0$ (both identically zero for LSE).

What is crucial is the following uniform WLLN, which will be proved later:

Claim C.1. $\sup_{\theta\in\Theta}\left|(nh_n^2)^{-1}\nabla_\theta^2\Psi_n(\theta)-\Gamma(\theta)\right| \xrightarrow{P_0^\eta} 0$ *with* $\Gamma(\theta)$ *defined by (2.7).*

Define some notation as follows: $\lambda_n = \sqrt{nh_n}(\tilde{\theta}_n-\theta_0)$, $\Gamma_n = (2nh_n^2)^{-1}\nabla_\theta^2\Psi_n(\theta_n^*)$, and $\Sigma_n = -(2\sqrt{nh_n^3})^{-1}\nabla_\theta\Psi_n(\theta_0)$. The tightness of (λ_n) is implied by existence of a tight subsubsequence of any subsequence of (λ_n); e.g. Kallenberg [6, Proposition 4.27]. Take any subsequence $(n')\subset\mathbf{N}$, 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 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^η-a.s. From (C.3) we then have $\lambda_m = \bar{\Gamma}_m^{-1}\{\Sigma_m + O(1)\}$, $m\in(n'')$, P_0^η-a.s., hence the proof is complete if we prove that the sequence $(\Sigma_n)_{n\in\mathbf{N}}$ is tight.

We have $\Sigma_n = \Sigma_n^1 + \Sigma_n^2$, where

$$\begin{aligned}
\Sigma_n^1 &= \sum_{i=1}^n \frac{1}{\sqrt{nh_n}} \int_{t_{i-1}^n}^{t_i^n} \nabla_\theta b(X_{t_{i-1}^n}, \theta_0)^\top \Delta b(s, t_{i-1}^n; \theta_0) ds, \\
\Sigma_n^2 &= \sum_{i=1}^n \frac{1}{\sqrt{nh_n}} \nabla_\theta b(X_{t_{i-1}^n}, \theta_0)^\top (Y_{t_i^n} - Y_{t_{i-1}^n}).
\end{aligned}$$

But, under Assumption 2, we can see that $E_0^\eta[|\Sigma_n^1|] \lesssim \sqrt{nh_n}(\Delta_{q',n})^{1/q'} = O(1)$, therefore $(\Sigma_n^1)_{n\in\mathbf{N}}$ is tight in view of Markov's inequality. This together with Assumption 3 implies the tightness of Σ_n. This completes the proofs for both of TFE and LSE.

C.1 Proof of Claim C.1

For every $a, b \in \{1, \dots, p\}$, we have

$$\begin{aligned}
\frac{1}{2nh_n^2} \nabla^2_{\theta_a\theta_b} \Psi_n(\theta_0) &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^d [\nabla_{\theta_a} b^j(X_{t_{i-1}^n}, \theta_0)][\nabla_{\theta_b} b^j(X_{t_{i-1}^n}, \theta_0)] \\
&\quad - \frac{1}{nh_n} \sum_{i=1}^n \sum_{j=1}^d \chi_i^{n,j}(\theta_0)[\nabla^2_{\theta_a\theta_b} b^j(X_{t_{i-1}^n}, \theta_0)] \\
&= H_n^{1,ab}(\theta_0) + H_n^{2,ab}(\theta_0), \quad \text{say.}
\end{aligned}$$

Apply Assumption 5 to conclude that $H_n^{1,ab}(\theta_0)$ tends in P_0^η-probability to $\Gamma(\theta_0)^{ab}$, as in the argument in Section B.2 (concerning $J_n^1(\theta_1)$ and $J_n^2(\theta_1)$). Also, mimicking the proof of $J_n^3(\theta_1) = o(1)$ in Section B.1, it is not difficult to show that $H_n^{2,ab}(\theta_0) = \bar{o}_{P_0^\eta}(1)$. Thus it remains to show the modulus of continuity of the random field $(H_n^{1,ab}(\theta) - \Gamma(\theta)^{ab})_{\theta\in\Theta}$. This can be seen as follows. Assumption 1 ensures that, for any $\theta_1, \theta_2 \in \Theta$ and $p' > p$, we have $\nabla_{\theta_a} b^j(x,\theta_1)\nabla_{\theta_b} b^j(x,\theta_1) - \nabla_{\theta_a} b^j(x,\theta_2)\nabla_{\theta_b} b^j(x,\theta_2) \le \bar{b}(x)|\theta_1 - \theta_2|^{p'}$ for some measurable function $\bar{b}$, independent of θ_1 and θ_2, of at most polynomial growth. Therefore we see that $E_0^\eta[|H_n^{1,ab}(\theta_1) - H_n^{1,ab}(\theta_2)|^{p'}] + E_0^\eta[|\Gamma(\theta_1)^{ab} - \Gamma(\theta_2)^{ab}|^{p'}] \lesssim |\theta_1 - \theta_2|^{p'}$, completing the proof.

D Proof of Lemma 3.1

D.1 On Assumption 2

Specifically writing $\kappa(X)$ instead of κ, we see that for every $Y', Y'' \in \mathbb{D}^d$ (the space of all càdlàg functions from $\mathbf{R}_+$ to $\mathbf{R}^d$),

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

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 (3.1) 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 $t \in \mathbf{R}_+$; see, e.g., Protter [15, Theorem V-7] for such fundamental facts concerning stochastic differential equations driven by a semimartingale.

Fix an arbitrary $q \geq 2$. Using the assumptions we get

$$\begin{aligned}\|\kappa\|^{*,q}_{\mathbf{R}_+} &\lesssim \sup_{t\in\mathbf{R}_+} E^\eta_0\left[\left|\int_{(-t,0]} F_\kappa(X_{u+t})r_\kappa(du)\right|^q\right] + 1 \\ &\lesssim \left(1 + \sup_{s\in\mathbf{R}_+} E^\eta_0[|X_s|^q]\right) r_\kappa(-\mathbf{R}_+) + 1 \lesssim 1,\end{aligned}$$

hence the $L^q(P^\eta_0)$-boundedness of M follows.

We now turn to the estimate (2.5) of Assumption 2. Put

$$g_{q,i}(t) = E^\eta_0\left[\|X_\cdot - X_{t^n_{i-1}}\|^{*,q}_{(t^n_{i-1},t]}\right], \quad t \in (t^n_{i-1}, t^n_i] \tag{D.1}$$

for $q \geq 2$. 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 [1, Lemma (A.14)]; we here rephrase it just to note the orders of the upper bounds in h_n, all of which are obvious from the original proofs.

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

$$E^\eta_0\left[\left\|\int_{t^n_{i-1}}^{\cdot} H_s ds\right\|^{*,q}_{(t^n_{i-1},t^n_i]}\right] \leq h_n^{q-1}\int_{t^n_{i-1}}^{t^n_i} E^\eta_0[|H_s|^q]ds \tag{D.2}$$

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

$$E^\eta_0\left[\left\|\int_{t^n_{i-1}}^{\cdot} G_s dw_s\right\|^{*,q}_{(t^n_{i-1},t^n_i]}\right] \lesssim h_n^{q/2-1}\int_{t^n_{i-1}}^{t^n_i} E^\eta_0[|G_s|^q]ds \tag{D.3}$$

for $i = 1, 2, \ldots, n$.
(c) For a d-dimensional $\mathscr{F} \otimes \mathcal{B}^{r_\mu}$-measurable process [6] $U(s,z) = U(\omega; s, z)$ defined on $\Omega \times \mathbf{R}_+ \times (\mathbf{R}^{r_\mu}\backslash\{0\})$ such that $|U(\omega; s, z)| \leq \bar{U}_s(\omega)\rho(z)$ with ξ predictable and $\rho \in L^2(\nu) \cap L^q(\nu)$, we have

$$E^\eta_0\left[\left\|\int_{t^n_{i-1}}^{\cdot}\int U(s,z)\tilde{\mu}(ds,dz)\right\|^{*,q}_{(t^n_{i-1},t^n_i]}\right] \lesssim \int_{t^n_{i-1}}^{t^n_i} E^\eta_0[|\bar{U}_s|^q]ds \tag{D.4}$$

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

Remark D.1. The inequalities (D.2) to (D.4) still hold true P^η_0-a.s. for $E^\eta_0[\cdot]$ replaced by the conditional expectation $E^\eta_0[\cdot|\mathscr{F}_{t^n_{i-1}}]$.

[6] $\mathcal{B}^{r_\mu}$ denotes the r_μ-dimensional Borel σ-field.

Now observe that Assumption 1 and Lemma D.1 yield

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

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

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

from which it follows that

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

Here note that the term "$\int_{t_{i-1}^n}^t g_{q,i}(s)ds$" in the upper bound of (D.6) appears only when F_κ is not identically null. Combine (D.5) and (D.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 $g_{q,i}(t) \lesssim h_n$ (here we exclude the trivial case $(\sigma, \zeta) \equiv 0$); this bound is same as diffusions with or without jumps. Hence the condition $\sqrt{nh_n}(\Delta_{q',n})^{1/q'} = O(1)$ is fulfilled with $q' = 2$, so we get (2.5).

D.2 On 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 and Lemma D.1 we have

$$\begin{aligned}
&E_0^\eta\left[\left|\frac{1}{nh_n}\sum_{i=1}^n \Delta b(X_{t_{i-1}^n};\theta_1,\theta_2)^\top (M_{t_i^n} - M_{t_{i-1}^n})\right|^{p'}\right] \\
&\quad\lesssim (nh_n)^{-p'}\left\{E_0^\eta\left[\left|\int_0^{nh_n}\sum_{i=1}^n \mathbf{1}_i(s)\Delta b(X_{t_{i-1}^n};\theta_1,\theta_2)^\top \sigma_s dw_s\right|^{p'}\right]\right. \\
&\qquad\left. + E_0^\eta\left[\left|\int_0^{nh_n}\sum_{i=1}^n \mathbf{1}_i(s)\Delta b(X_{t_{i-1}^n};\theta_1,\theta_2)^\top \zeta_s dJ_s\right|^{p'}\right]\right\} \\
&\quad\lesssim (nh_n)^{-p'}\left\{(nh_n)^{p'/2-1}\int_0^{nh_n} E_0^\eta\left[\left\{\sum_{i=1}^n \mathbf{1}_i(s)|\Delta b(X_{t_{i-1}^n};\theta_1,\theta_2)^\top \sigma_s|\right\}^{p'}\right]ds\right. \\
&\qquad\left. + \int_0^{nh_n} E_0^\eta\left[\left\{\sum_{i=1}^n \mathbf{1}_i(s)|\Delta b(X_{t_{i-1}^n};\theta_1,\theta_2)^\top \zeta_s|\right\}^{p'}\right]ds\right\} \\
&\quad\lesssim (nh_n)^{-p'}\left\{(nh_n)^{p'/2-1}\sum_{i=1}^n\int_{t_{i-1}^n}^{t_i^n} E_0^\eta\left[|\Delta b(X_{t_{i-1}^n};\theta_1,\theta_2)^\top \sigma_s|^{p'}\right]ds\right. \\
&\qquad\left. + \sum_{i=1}^n\int_{t_{i-1}^n}^{t_i^n} E_0^\eta\left[|\Delta b(X_{t_{i-1}^n};\theta_1,\theta_2)^\top \zeta_s|^{p'}\right]ds\right\} \\
&\quad\lesssim (nh_n)^{-p'}\left\{(nh_n)^{p'/2}|\theta_1-\theta_2|^{p'} + nh_n|\theta_1-\theta_2|^{p'}\right\} \\
&\quad\lesssim (nh_n)^{-p'/2}|\theta_1-\theta_2|^{p'},
\end{aligned}$$

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

$$E_0^\eta\left[\left|\frac{1}{\sqrt{nh_n}}\sum_{i=1}^n (M_{t_i^n} - M_{t_{i-1}^n})\nabla_\theta b(X_{t_{i-1}^n},\theta_0)\right|^{p'}\right] \lesssim (nh_n)^{-p'/2}\left\{(nh_n)^{p'/2} + nh_n\right\} \lesssim 1.$$

Thus the second statement of Assumption 3 is also fulfilled.

References

[1] 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.

[2] 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.

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

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

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

[6] Kallenberg, O. (1997), *Foundations of modern probability.* Springer-Verlag, New York.

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

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

[9] Masuda, H. (2002), Analytical properties of GIG and GH distributions (A survey; in Japanese). *Proceedings of the institute of statistical mathematics* **50**, 165–199.

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

[11] Masuda, H. (2004), Classical method of moments for partially and discretely observed ergodic models. *Stat. Inference Stoch. Process.* **8**, 25–50.

[12] Masuda, H. (2004), Exponential β-mixing bound for Lévy-driven stochastic differential equations. preprint.

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

[14] Prakasa Rao, B. L. S. (1999), *Statistical Inference for Diffusion type Processes.* Kendall's Library of Statistics, 8. Edward Arnold, London; Oxford University Press, New York.

[15] Protter, P. (1990), *Stochastic Integration and Differential Equations. A New Approach.* Applications of Mathematics, 21. Springer-Verlag, Berlin.

[16] Sato, K. (1999), *Lévy Processes and Infinitely Divisible Distributions.* Cambridge, Cambridge University Press.

[17] Shimizu, Y. (2004), M-estimation for discretely observed ergodic duffusion processes with infinite jumps. submitted.

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

List of MHF Preprint Series, Kyushu University

21st Century COE Program
Development of Dynamic Mathematics with High Functionality

MHF2004-1 Koji YONEMOTO & Takashi YANAGAWA
Estimating the Lyapunov exponent from chaotic time series with dynamic noise

MHF2004-2 Rui YAMAGUCHI, Eiko TSUCHIYA & Tomoyuki HIGUCHI
State space modeling approach to decompose daily sales of a restaurant into time-dependent multi-factors

MHF2004-3 Kenji KAJIWARA, Tetsu MASUDA, Masatoshi NOUMI, Yasuhiro OHTA & Yasuhiko YAMADA
Cubic pencils and Painlevé Hamiltonians

MHF2004-4 Atsushi KAWAGUCHI, Koji YONEMOTO & Takashi YANAGAWA
Estimating the correlation dimension from a chaotic system with dynamic noise

MHF2004-5 Atsushi KAWAGUCHI, Kentarou KITAMURA, Koji YONEMOTO, Takashi YANAGAWA & Kiyofumi YUMOTO
Detection of auroral breakups using the correlation dimension

MHF2004-6 Ryo IKOTA, Masayasu MIMURA & Tatsuyuki NAKAKI
A methodology for numerical simulations to a singular limit

MHF2004-7 Ryo IKOTA & Eiji YANAGIDA
Stability of stationary interfaces of binary-tree type

MHF2004-8 Yuko ARAKI, Sadanori KONISHI & Seiya IMOTO
Functional discriminant analysis for gene expression data via radial basis expansion

MHF2004-9 Kenji KAJIWARA, Tetsu MASUDA, Masatoshi NOUMI, Yasuhiro OHTA & Yasuhiko YAMADA
Hypergeometric solutions to the q - Painlevé equations

MHF2004-10 Raimundas VIDŪNAS
Expressions for values of the gamma function

MHF2004-11 Raimundas VIDŪNAS
Transformations of Gauss hypergeometric functions

MHF2004-12 Koji NAKAGAWA & Masakazu SUZUKI
Mathematical knowledge browser

MHF2004-13 Ken-ichi MARUNO, Wen-Xiu MA & Masayuki OIKAWA
Generalized Casorati determinant and Positon-Negaton-Type solutions of the Toda lattice equation

MHF2004-14 Nalini JOSHI, Kenji KAJIWARA & Marta MAZZOCCO
Generating function associated with the determinant formula for the solutions of the Painlevé II equation

MHF2004-15 Kouji HASHIMOTO, Ryohei ABE, Mitsuhiro T. NAKAO & Yoshitaka WATANABE
Numerical verification methods of solutions for nonlinear singularly perturbed problem

MHF2004-16 Ken-ichi MARUNO & Gino BIONDINI
Resonance and web structure in discrete soliton systems: the two-dimensional Toda lattice and its fully discrete and ultra-discrete versions

MHF2004-17 Ryuei NISHII & Shinto EGUCHI
Supervised image classification in Markov random field models with Jeffreys divergence

MHF2004-18 Kouji HASHIMOTO, Kenta KOBAYASHI & Mitsuhiro T. NAKAO
Numerical verification methods of solutions for the free boundary problem

MHF2004-19 Hiroki MASUDA
Ergodicity and exponential β-mixing bounds for a strong solution of Lévy-driven stochastic differential equations

MHF2004-20 Setsuo TANIGUCHI
The Brownian sheet and the reflectionless potentials

MHF2004-21 Ryuei NISHII & Shinto EGUCHI
Supervised image classification based on AdaBoost with contextual weak classifiers

MHF2004-22 Hideki KOSAKI
On intersections of domains of unbounded positive operators

MHF2004-23 Masahisa TABATA & Shoichi FUJIMA
Robustness of a characteristic finite element scheme of second order in time increment

MHF2004-24 Ken-ichi MARUNO, Adrian ANKIEWICZ & Nail AKHMEDIEV
Dissipative solitons of the discrete complex cubic-quintic Ginzburg-Landau equation

MHF2004-25 Raimundas VIDŪNAS
Degenerate Gauss hypergeometric functions

MHF2004-26 Ryo IKOTA
The boundedness of propagation speeds of disturbances for reaction-diffusion systems

MHF2004-27 Ryusuke KON
Convex dominates concave: an exclusion principle in discrete-time Kolmogorov systems

MHF2004-28 Ryusuke KON
Multiple attractors in host-parasitoid interactions: coexistence and extinction

MHF2004-29 Kentaro IHARA, Masanobu KANEKO & Don ZAGIER
Derivation and double shuffle relations for multiple zeta values

MHF2004-30 Shuichi INOKUCHI & Yoshihiro MIZOGUCHI
Generalized partitioned quantum cellular automata and quantization of classical CA

MHF2005-1 Hideki KOSAKI
Matrix trace inequalities related to uncertainty principle

MHF2005-2 Masahisa TABATA
Discrepancy between theory and real computation on the stability of some finite element schemes

MHF2005-3 Yuko ARAKI & Sadanori KONISHI
Functional regression modeling via regularized basis expansions and model selection

MHF2005-4 Yuko ARAKI & Sadanori KONISHI
Functional discriminant analysis via regularized basis expansions

MHF2005-5 Kenji KAJIWARA, Tetsu MASUDA, Masatoshi NOUMI, Yasuhiro OHTA & Yasuhiko YAMADA
Point configurations, Cremona transformations and the elliptic difference Painlevé equations

MHF2005-6 Kenji KAJIWARA, Tetsu MASUDA, Masatoshi NOUMI, Yasuhiro OHTA & Yasuhiko YAMADA
Construction of hypergeometric solutions to the q - Painlevé equations

MHF2005-7 Hiroki MASUDA
Simple estimators for non-linear Markovian trend from sampled data:
I. ergodic cases