Least squares estimators for stochastic differential equations driven by small Lévy noises

Hongwei Long\textsuperscript{a,*}, Chunhua Ma\textsuperscript{b}, Yasutaka Shimizu\textsuperscript{c}

\textsuperscript{a}Department of Mathematical Sciences, Florida Atlantic University, Boca Raton, FL 33431-0991, USA
\textsuperscript{b}School of Mathematical Sciences and LPMC, Nankai University, Tianjin 300071, PR China
\textsuperscript{c}Department of Applied Mathematics, Waseda University, 3-4-1 Okubo, Shinjuku, Tokyo 169-8555, Japan

Received 5 October 2015; received in revised form 26 July 2016; accepted 15 August 2016
Available online 24 August 2016

Highlights

\begin{itemize}
\item We consider parameter estimation for stochastic processes driven by Lévy noises.
\item We propose least squares estimator for the drift parameters.
\item Consistency and rate of convergence of the estimator are established.
\item A simulation study illustrates the asymptotic behavior of the estimator.
\end{itemize}

Abstract

We study parameter estimation for discretely observed stochastic differential equations driven by small Lévy noises. We do not impose Lipschitz condition on the dispersion coefficient function $\sigma$ and any moment condition on the driving Lévy process, which greatly enhances the applicability of our results to many practical models. Under certain regularity conditions on the drift and dispersion functions, we obtain consistency and rate of convergence of the least squares estimator (LSE) of parameter when $\varepsilon \to 0$ and $n \to \infty$ simultaneously. We present some simulation study on a two-factor financial model driven by stable noises.

MSC: primary 62F12; 62M05; secondary 60G52; 60J75

Keywords: Asymptotic distribution; Consistency; Discrete observations; Least squares method; Stochastic differential equations; Parameter estimation

\cite{http://dx.doi.org/10.1016/j.spa.2016.08.006}

\copyright 2016 Elsevier B.V. All rights reserved.

* Corresponding author. Fax: +1 561 2972436.
E-mail addresses: hlong@fau.edu (H. Long), mach@nankai.edu.cn (C. Ma), shimizu@waseda.jp (Y. Shimizu).

http://dx.doi.org/10.1016/j.spa.2016.08.006

0304-4149/\copyright 2016 Elsevier B.V. All rights reserved.
1. Introduction

Let $\mathbb{R}_0^r = \mathbb{R}^r \setminus 0$ and let $\mu(du)$ be a $\sigma$-finite measure on $\mathbb{R}_0^r$ satisfying $\int_{\mathbb{R}_0^r}(|u|^2 \wedge 1) \mu(du) < \infty$ with $|u| = (\sum_{i=1}^r u_i^2)^{1/2}$. Let $\{B_t = (B_1^t, \ldots, B_r^t) : t \geq 0\}$ be an $r$-dimensional standard Brownian motion and let $N(dt, du)$ be a Poisson random measure on $(0, \infty) \times \mathbb{R}_0^r$ with intensity measure $dt \mu(du)$. Suppose that $\{B_t\}$ and $\{N(dt, du)\}$ are independent of each other. Then an $r$-dimensional Lévy process $\{L_t\}$ can be given as

$$L_t = B_t + \int_0^t \int_{|u|>1} uN(ds, du) + \int_0^t \int_{|u|\leq 1} u\tilde{N}(ds, du), \quad (1.1)$$

where $\tilde{N}(ds, du) = N(ds, du) - ds \mu(du)$. Let us consider a family of $d$-dimensional jump–diffusion processes defined as the solution of

$$dX_t^e = b(X_t^{e\theta}, \theta)dt + \varepsilon \sigma(X_t^{e\theta})dL_t, \quad t \in [0, 1], \quad X_0^e = x,$$

where $\theta \in \tilde{\Theta}$, the closure of an open convex bounded subset $\Theta$ of $\mathbb{R}^p$. The function $b(x, \theta) = (b^k(x, \theta))$ is $\mathbb{R}^d$-valued and defined on $\mathbb{R}^d \times \tilde{\Theta}$; the function $\sigma(x) = (\sigma^k(x))$ is defined on $\mathbb{R}^d$ and takes values on the space of matrices $\mathbb{R}^d \otimes \mathbb{R}^r$; the initial value $x \in \mathbb{R}^d$, and $\varepsilon > 0$ are known constants. A stochastic process of form (1.2) has long been used in the financial world and has been the fundamental tool in financial modeling. We refer to Sundaresan [38] and Fan [7] for overviews, Barndorff-Nielsen, Mikosch and Resnick [2] for recent developments on Lévy-driven processes, and Sørensen [34], Shimizu and Yoshida [33], Shimizu [30], Ogihara and Yoshida [26] and Masuda [25] for statistical inference. Examples of (1.2) include (i) the multivariate diffusion process defined by

$$dX_t^e = b(X_t^e, \theta)dt + \varepsilon \sigma(X_t^e)dB_t,$$

see Stroock and Varadhan [37]; (ii) the Vasicek model with jumps or the Lévy driven Ornstein–Uhlenbeck process defined by

$$dX_t^e = \kappa(\beta - X_t^e)dt + \varepsilon dL_t,$$

where $\kappa$ and $\beta$ are positive constants, and $L_t$ can be chosen to be a standard symmetric $\alpha$-stable Lévy process (see Hu and Long [12] and Fasen [8]) or a (positive) Lévy subordinator (see Barndorff-Nielsen and Shephard [3]); (iii) the Cox–Ingersoll–Ross (CIR) model driven by $\alpha$-stable Lévy processes defined by

$$dX_t^e = \kappa(\beta - X_t^e)dt + \varepsilon \sqrt{X_t^e} dL_t, \quad (1.3)$$

where $\{L_t\}$ is a spectrally positive $\alpha$-stable process with $1 < \alpha < 2$; see Fu and Li [9] and Li and Ma [18].

Assume that the only unknown quantity in (1.2) is the parameter $\theta$. We denote the true value of the parameter by $\theta_0$ and assume that $\theta_0 \in \tilde{\Theta}$. Suppose that this process is observed at regularly spaced time points $\{t_k = k/n, \ k = 1, 2, \ldots, n\}$. The purpose of this paper is to study the least squares estimator for the true value $\theta_0$ based on the sampling data $(X_{tk}^e)_{k=1}^n$ with small dispersion $\varepsilon$ and large sample size $n$. There are several practical advantages in small noise asymptotics: (i) we can get the drift parameter estimation by samples from a fixed finite time interval under relatively
mild conditions so no ergodicity and moment conditions are required, which is important for the case of Lévy processes; (ii) the multi-dimensional models can be treated as easily as one-dimensional models.

The small diffusion asymptotic \( \varepsilon \to 0 \) has been extensively studied with wide applications to real world problems; see Uchida and Yoshida [44] and Takahashi [39] for the applications to contingent claim pricing. Several papers have been devoted to small diffusion asymptotics for parameter estimators in diffusion models. For the problem of parameter estimation based on continuous observations, we refer to Kunitomo and Takahashi [14], Kutoyants [15,16], Yoshida [46–48], Takahashi and Yoshida [40], and Uchida and Yoshida [43,44]. From a practical point of view in parametric inference, it is more realistic and interesting to consider asymptotic estimation for diffusion processes with small noise based on discrete observations. Substantial progress has been made in this direction. Genon-Catalot [10] and Laredo [17] studied the efficient estimation for diffusion processes with small noise based on discrete observations. Substantial progress has been made in this direction. Genon-Catalot [10] and Laredo [17] studied the efficient estimation for diffusion processes with small noise based on discrete observations. Substantial progress has been made in this direction. Genon-Catalot [10] and Laredo [17] studied the efficient estimation for diffusion processes with small noise based on discrete observations. Substantial progress has been made in this direction. Genon-Catalot [10] and Laredo [17] studied the efficient estimation for diffusion processes with small noise based on discrete observations. Substantial progress has been made in this direction. Genon-Catalot [10] and Laredo [17] studied the efficient estimation for diffusion processes with small noise based on discrete observations. Substantial progress has been made in this direction. Genon-Catalot [10] and Laredo [17] studied the efficient estimation for diffusion processes with small noise based on discrete observations.

Sørensen [35] used martingale estimating functions to establish consistency and asymptotic normality of the estimators of drift and diffusion coefficient parameters when \( \varepsilon \to 0 \) and \( n \to \infty \). Sørensen and Uchida [36] and Gloter and Sørensen [11] used a contrast function to study the efficient estimation for unknown parameters in both drift and diffusion coefficient functions. Uchida [41,42] used the martingale estimating function approach to study estimation of drift parameters for small diffusions under weaker conditions. Thus, in the cases of small diffusions, the asymptotic distributions of the estimators are normal under suitable conditions on \( \varepsilon \) and \( n \).

Long [19] studied the parameter estimation problem for discretely observed one-dimensional Ornstein–Uhlenbeck processes with small Lévy noises. In that paper, the drift function is linear in both \( x \) and \( \theta \) \( (b(x, \theta) = -\theta x \) and \( \sigma(x) = 1) \), the driving Lévy process is \( L_t = a B_t + b Z_t \), where \( a \) and \( b \) are known constants, \( (B_t, t \geq 0) \) is the standard Brownian motion and \( Z_t \) is an \( \alpha \)-stable Lévy motion independent of \( (B_t, t \geq 0) \). The consistency and rate of convergence of the least squares estimator are established. The asymptotic distribution of the LSE is shown to be the convolution of a normal distribution and a stable distribution. Ma [22] extended the results of Long [19] to the case when the driving noise is a general Lévy process. Long [20] discussed the drift parameter estimation when \( b(x, \theta) = \theta b(x) \), \( \sigma(x) \) is bounded, and \( L_t \) is an \( \alpha \)-stable Lévy process with \( 1 < \alpha < 2 \). Ma and Yang [23] also studied the parameter estimation problem for discretely observed CIR model when \( b(x, \theta) = \beta - cx \), \( \sigma(x) = (x^+)^{1/q} \) for some \( q > 1 \), and \( L_t \) is a spectrally positive \( \alpha \)-stable Lévy process with \( 1 < \alpha < 2 \). Recently Long et al. [21] discussed the statistical estimation of the drift parameter for a class of SDEs driven by small Lévy noise with general drift function \( b(x, \theta) \) and \( \sigma(x) = 1 \).

In this paper, we consider the stochastic model (1.2) driven by small Lévy noises with drift function \( b(x, \theta) \) and a non-constant dispersion function \( \sigma(x) \). We are interested in estimating the drift parameter \( \theta \) based on discrete observations \( \{X_t\}_{t=1}^n \) when \( \varepsilon \to 0 \) and \( n \to \infty \). We shall use the least squares method to obtain an asymptotically consistent estimator. The novelty of this paper compared with Long et al. [21] is that there is a non-constant function \( \sigma(x) \) and compared with Long [20] is that \( \sigma(x) \) is unbounded and the Lévy process is not only \( \alpha \)-stable. As mentioned earlier, the least squares contrast has been used for estimating drift parameters for diffusions without jumps (see Sørensen and Uchida [36], Uchida [41,42], and Gloter and Sørensen [11]). The originality of the present paper lies in the fact that the driving process is Lévy instead of Brownian.

The paper is organized as follows. In Section 2, we propose our least squares estimator in our general framework and state the main results, which provide consistency and asymptotic
behavior of the LSE. In Section 3, we present a two-factor financial model as an example and its numerical simulation study. All the proofs are given in Section 4.

2. Main results

We start this section by presenting the necessary conditions and the construction of the estimator. Let $X^0 = (X^0_t, t \geq 0)$ be the solution to the underlying ordinary differential equation (ODE) under the true value of the drift parameter:

$$dX^0_t = b(X^0_t, \theta_0)dt, \quad X^0_0 = x_0.$$ 

For a multi-index $m = (m_1, \ldots, m_k)$, we define a derivative operator in $z \in \mathbb{R}^k$ as $\partial^m_z := \partial_{z_1}^{m_1} \cdots \partial_{z_k}^{m_k}$, where $\partial_{z_i}^{m_i} := \partial^{m_i}/\partial z_i^{m_i}$. Let $C^{k,l}(\mathbb{R}^d \times \Theta; \mathbb{P})$ be the space of all functions $f : \mathbb{R}^d \times \Theta \to \mathbb{R}^q$ which is $k$ and $l$ times continuously differentiable with respect to $x$ and $\theta$, respectively. Moreover $C_k^l(\mathbb{R}^d \times \Theta; \mathbb{P})$ is a class of $f \in C^{k,l}(\mathbb{R}^d \times \Theta; \mathbb{P})$ satisfying that $\sup_{\theta \in \Theta} |\partial^\alpha_\theta \partial^\beta_x f(x, \theta)| \leq C(1 + |x|)^{\lambda}$ for universal positive constants $C$ and $\lambda$, where $\alpha = (\alpha_1, \ldots, \alpha_p)$ and $\beta = (\beta_1, \ldots, \beta_d)$ are multi-indices with $0 \leq \sum_{j=1}^p \alpha_j \leq l$ and $0 \leq \sum_{i=1}^d \beta_i \leq k$, respectively. We use the notation $P_{\theta_0}$ to stand for convergence in probability under $P_{\theta_0}$.

Now let us introduce the following set of assumptions.

(A1) For all $\varepsilon > 0$, the SDE (1.2) admits a unique strong solution $X^\varepsilon$ on some probability space $(\Omega, \mathcal{F}_t, \mathbb{P})$.

(A2) There exists a constant $K > 0$ such that

$$|b(x, \theta) - b(y, \theta)| \leq K|x - y|; \quad |b(x, \theta)| + |\sigma(x)| \leq K(1 + |x|)$$

for each $x, y \in \mathbb{R}^d$ and $\theta \in \Theta$.

(A3) $b(\cdot, \cdot) \in C^{0,3}(\mathbb{R}^d \times \Theta; \mathbb{R}^d)$.

(A4) $\sigma$ is continuous and there exists some open convex subset $\mathcal{U}$ of $\mathbb{R}^d$ such that $X^0_t \in \mathcal{U}$ for all $t \in [0, 1]$, and $\sigma$ is smooth on $\mathcal{U}$. Moreover $\sigma\sigma^*(x)$ is invertible on $\mathcal{U}$.

(A5) $\theta \neq \theta_0 \iff b(X^0_t, \theta) \neq b(X^0_t, \theta_0)$ for at least one value of $t \in [0, 1]$.

(B) $\varepsilon = \varepsilon_n \to 0$ and $n\varepsilon \to \infty$ as $n \to \infty$.

Consider the following contrast function

$$\Psi_{n, \varepsilon}(\theta) = \left(\sum_{k=1}^n \varepsilon^{-2} \frac{p_k^*(\theta)}{\Lambda_k} A_{k-1}^{-1} p_k(\theta)\right)1_{\{Z > 0\}}, \quad (2.4)$$

where

$$p_k(\theta) = X_{t_k} - X_{t_{k-1}} - \frac{1}{n} b(X_{t_{k-1}}, \theta),$$

$$\Lambda_k = [\sigma\sigma^*](X_{t_k})$$

and the random variable $Z = \inf_{k=0, \ldots, n-1} \det \Lambda_k$ is introduced to insure that $\Psi_{n, \varepsilon}$ is well defined. Let $\hat{\theta}_{n, \varepsilon}$ be a minimum contrast estimator, i.e. a family of random variables satisfying

$$\hat{\theta}_{n, \varepsilon} := \arg \min_{\theta \in \Theta} \Psi_{n, \varepsilon}(\theta).$$

Since minimizing $\Psi_{n, \varepsilon}(\theta)$ is equivalent to minimizing

$$\Phi_{n, \varepsilon}(\theta) := \varepsilon^2 (\Psi_{n, \varepsilon}(\theta) - \Psi_{n, \varepsilon}(\theta_0)),$$
we may write the LSE as

\[
\hat{\theta}_{n, \varepsilon} = \arg \min_{\theta \in \Theta} \Phi_{n, \varepsilon}(\theta).
\]

(2.5)

Note that in the case when \( b(x, \theta) \) is nonlinear in both \( x \) and \( \theta \), it is generally very difficult or impossible to obtain an explicit formula for the least squares estimator \( \hat{\theta}_{n, \varepsilon} \). However, we can use some nice criteria in statistical inference (see Chapter 5 of Van der Vaart [45] and Shimizu [31] for a more general criterion) to establish the consistency of the LSE as well as its asymptotic behaviors (asymptotic distribution and rate of convergence). The main result of this paper is the following asymptotics of the LSE \( \hat{\theta}_{n, \varepsilon} \) with high frequency \( (n \to \infty) \) and small dispersion \( (\varepsilon \to 0) \).

**Theorem 2.1.** Under the conditions \((A1)\)–\((A5)\), we have

\[
\hat{\theta}_{n, \varepsilon} \xrightarrow{P_{\theta_0}} \theta_0,
\]

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

Introduce the matrix \( I(\theta_0) = \left( I_{ij}(\theta_0) \right)_{1 \leq i, j \leq p} \), where

\[
I_{ij}(\theta) = \int_0^1 (\partial_\theta b)^* (X_s^0, \theta) [\sigma \sigma^*]^{-1}(X_s^0) \partial_\theta b(X_s^0, \theta) ds.
\]

(2.6)

**Theorem 2.2.** Assume \((A1)\)–\((A5)\), \((B)\) and that \( \theta_0 \in \Theta \) with the matrix \( I(\theta_0) \) given in \((2.6)\) being positive definite. Then

\[
\varepsilon^{-1}(\hat{\theta}_{n, \varepsilon} - \theta_0) \xrightarrow{P_{\theta_0}} \left( \int_0^1 (\partial_\theta b)^* (X_t^0, \theta_0) [\sigma \sigma^*]^{-1}(X_t^0) \sigma(X_t^0) dL_t \right)^*_{1 \leq i \leq p},
\]

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

**Remark 2.3.** The scheme of the proofs for Theorems 2.1 and 2.2 is not new. The originality lies in the proofs is that we have to deal with the new terms coming from the jump part of the Lévy processes and the unboundedness of \( \sigma(x) \).

**Remark 2.4.** We do some comparisons of our results with the case of diffusions here. The rate of convergence in Theorem 2.2 is the same as the case of diffusions (see Theorem 1 in Uchida [41]). For the constraint \( n\varepsilon \to \infty \) in condition \((B)\), this condition is compatible with the condition \( (\varepsilon n^l)^{-1} \to 0 \) for a positive integer \( l \) in Uchida [41]. When \( l = 1 \), these two conditions coincide and the approximate martingale estimating function in [41] reduces to our least squares contrast function. This condition \((B)\) is coming from Lemma 4.7 (see the proof of Lemma 3.6 in Long et al. [21]), which is necessary for the rate of convergence in Theorem 2.2. There is an extra condition \( n\varepsilon^{\alpha/(\alpha-1)} \to 0 \) in Theorem 3.1 of Long [20]. But this extra condition is in fact not necessary and the proof of Theorem 3.1 in Long [20] used different techniques based on moment inequalities for stable stochastic integrals (see Lemma 2.4 of Long [20]). So, if we use the techniques in the present paper or Long et al. [21], this extra condition can be removed.

**Remark 2.5.** The limit distribution in Theorem 2.2 depends on the driving Lévy noises, which is quite intractable. However, this is the nature of the model \((1.2)\) since many of the Lévy
processes have intractable distributions. In general, an approximated limit distribution of the estimator is available from an empirical distribution (or histogram) by using Monte Carlo samples of \( \epsilon^{-1}(\hat{\theta}_{n,\epsilon} - \theta_0) \). We refer to Brouste et al. [4] for discussion on Monte Carlo method and asymptotic expansion method related to SDEs, and Robert and Casella [29] for general discussion on Monte Carlo methods.

**Remark 2.6.** The results in Theorems 2.1 and 2.2 still hold if we replace \( \varepsilon \sigma(X_{t-}^k)\) by a more general noise term \( \varepsilon \sum_{k=1}^{K} \alpha_k(X_{t-}^k)\) in the stochastic model (1.2), where the \( L^{(k)} \)'s are independent \( r \)-dimensional Lévy processes.

3. An example and its simulation: A two-factor model driven by stable noises

3.1. Two-factor model

Real rates and asset prices do not evolve continuously in time. Typical stylized facts of financial time series as asset returns, exchange rates and interest rates are jumps and a heavy tailed distribution in the sense that the second moment is infinite. These characteristics were already noticed in the 60s by the influential works of Mandelbrot [24] and Fama [6]. Thus, \( \alpha \)-stable distributions as generalization of a Gaussian distribution have often been discussed as more realistic models for asset returns than the usual normal distribution; see Rachev et al. [28]. Let \( Y \) be the log price of some asset and \( R \) represents the short term interest rate. We consider the following affine two-factor model defined by

\[
dY_t = (R_t + \mu_1)dt + \varepsilon dL^{1}_t, \quad Y_0 = y_0 \in \mathbb{R},
\]

\[
dR_t = \mu_2(m - R_t)dt + \varepsilon R^{1/\alpha}_t (\rho dL^{1}_t + (1 - \rho^\alpha)^{1/\alpha} dL^{2}_t), \quad R_0 = r_0 > 0,
\]

where \( \{L^{1}_t\} \) and \( \{L^{2}_t\} \) are independent spectrally positive \( \alpha \)-stable process with \( 1 < \alpha < 2 \), the unknown parameter \( \theta = (\mu_1, \mu_2, m) \in \mathbb{R} \times (0, \infty)^2 \) and the known constant \( \rho \in (0, 1) \). More precisely, the corresponding characteristic function of \( L^{k}_t \) is given by

\[
E[e^{iuL^{k}_t}] = \exp \left( -\sigma |u|^{\alpha} \left( 1 - i\beta \text{sign}(u) \tan \frac{\alpha \pi}{2} \right) + i\gamma u \right) \quad (k = 1, 2).
\]

Such a stable process is often written as \( L^{k}_t \sim S_{\alpha}(\sigma, \beta, \gamma) \). Note that the scale parameter \( \sigma \) affects the variation of the process, and that \( L^{k}_t \) is spectrally positive if the skew parameter \( \beta = 1 \), which is the case we will use in simulations later. In the above model, the second component is called the stable CIR model. We refer to Fu and Li [9] for the pathwise uniqueness of the positive strong solution of SDEs. The asymptotic estimation problem was studied by Li and Ma [18]. When \( \alpha = 2 \), and \( \{L^{1}_t, L^{2}_t\} = \{B^{1}_t, B^{2}_t\} \) is a two-dimensional Brownian motion, the parameter estimation of the above two-factor model was studied by Gloter and Sørensen [11].

For simplicity of computation, we set \( m = 1 \) in the SDE (3.8) for \( R_t \). We can easily get the explicit formulas for the LSE of the parameters \( \mu_1 \) and \( \mu_2 \) (the detail is omitted).

3.2. Numerical results

We set values of parameters as

\[
(y_0, r_0) = (1.0, 1.5), \quad \mu_1 = \mu_2 = 1, \quad \rho = 0.3, \quad (\alpha, \beta, \gamma) = (1.7, 1.0, 0.0),
\]

and try to estimate \( (\mu_1, \mu_2) \) when the volatility of the processes is relatively larger or smaller. For that purpose, we will change the values of \( \sigma \) and \( \varepsilon \) as \( \sigma = 1.0, 0.3, 0.1 \), and \( \varepsilon = 0.3, 0.1, 0.01 \), and try to estimate \( (\mu_1, \mu_2) \).
respectively. In each simulation, we generate 10000 paths and compute the mean and the standard deviation of $\hat{\mu}_1$ and $\hat{\mu}_2$.

We generate a sample path of $(Y, R)$ by discretizing the SDEs (3.7) and (3.8) via the Euler scheme. For generating stable random variables, we used an implemented function rstable in YUIMA package of R; see [49], which is a package for simulating SDEs on R; see Brouste et al. [4] or [49] for details. The results of the experiments are presented in Tables 1–3. For references, we show sample paths of the process $(Y, R)$ in Figs. 1 and 2 in cases where $(\sigma, \epsilon) = (1.0, 0.3)$ and $(0.3, 0.3)$.

When the scale parameter $\sigma = 1.0$, the performance of the LSE is not good due to the large fluctuation of the process. However, when $\sigma$ is relatively small such as $\sigma = 0.3$ or 0.1, the parameters are relatively well estimated even in the small sample such as $n = 10$, and we can observe that the consistency when $n$ increases. However the speed of convergence is not so rapid; see the case where $\epsilon = 0.01$ and $n = 200$. Such a phenomenon is also well known in estimation of the drift parameter in diffusion processes (see Gloter and Sørensen [11]), and the case of stable noise is more apparent than that. This indicates that there may remain some estimation bias in the drift estimation due to the high volatility of the stable noise process. The normal QQ-plots presented in Fig. 3 indicate the heavy tailedness of the asymptotic distribution of our estimator, which supports the results in Theorem 2.2.

We see a possibility to eliminate this bias in a Threshold Estimation proposed in Shimizu [32], where only ‘small’ increments of the data are available to estimate the drift. In this method, the asymptotic distribution of the estimator seems to be Gaussian, which is tractable for testing hypothesis problem. Although it is not clear that we can apply the method directly in the model discussed here since the method requires some restrictive condition on the index parameter $\alpha$ and others. The bias correction and constructing an estimator that is asymptotically normal are important issues in the future.


### Table 2
Case $\sigma = 0.3$: Mean (upper) and standard deviation (s.d., lower) of the estimators $\hat{\mu}_1$ and $\hat{\mu}_2$ through 10000 iterations.

<table>
<thead>
<tr>
<th>Case: $\sigma = 0.3$</th>
<th>$\epsilon = 0.3$</th>
<th>$n = 10$</th>
<th>$n = 50$</th>
<th>$n = 200$</th>
<th>True</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9743</td>
<td>0.9839</td>
<td>0.9880</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.0737)</td>
<td>(0.05519)</td>
<td>(0.0675)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9599</td>
<td>1.0120</td>
<td>1.0057</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.1589)</td>
<td>(0.1625)</td>
<td>(0.1775)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case: $\sigma = 0.3$</th>
<th>$\epsilon = 0.1$</th>
<th>$n = 10$</th>
<th>$n = 50$</th>
<th>$n = 200$</th>
<th>True</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9710</td>
<td>0.9821</td>
<td>0.9861</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.0570)</td>
<td>(0.0523)</td>
<td>(0.0568)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9603</td>
<td>1.0068</td>
<td>1.0021</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.1486)</td>
<td>(0.1664)</td>
<td>(0.1656)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case: $\sigma = 0.3$</th>
<th>$\epsilon = 0.01$</th>
<th>$n = 10$</th>
<th>$n = 50$</th>
<th>$n = 200$</th>
<th>True</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9806</td>
<td>0.9825</td>
<td>0.9875</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.2636)</td>
<td>(0.0508)</td>
<td>(0.0503)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9632</td>
<td>1.0156</td>
<td>1.0079</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.1785)</td>
<td>(0.1541)</td>
<td>(0.1646)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 3
Case $\sigma = 0.1$: Mean (upper) and standard deviation (s.d., lower) of the estimators $\hat{\mu}_1$ and $\hat{\mu}_2$ through 10000 iterations.

<table>
<thead>
<tr>
<th>Case: $\sigma = 0.1$</th>
<th>$\epsilon = 0.3$</th>
<th>$n = 10$</th>
<th>$n = 50$</th>
<th>$n = 200$</th>
<th>True</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9584</td>
<td>0.9731</td>
<td>0.9750</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.0961)</td>
<td>(0.0878)</td>
<td>(0.0960)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9674</td>
<td>1.0078</td>
<td>1.0106</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.1673)</td>
<td>(0.1759)</td>
<td>(0.1756)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case: $\sigma = 0.1$</th>
<th>$\epsilon = 0.1$</th>
<th>$n = 10$</th>
<th>$n = 50$</th>
<th>$n = 200$</th>
<th>True</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9604</td>
<td>0.9729</td>
<td>0.9800</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.0245)</td>
<td>(0.0205)</td>
<td>(0.0201)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9454</td>
<td>0.9849</td>
<td>0.9916</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.0629)</td>
<td>(0.0671)</td>
<td>(0.0701)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case: $\sigma = 0.1$</th>
<th>$\epsilon = 0.01$</th>
<th>$n = 10$</th>
<th>$n = 50$</th>
<th>$n = 200$</th>
<th>True</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9612</td>
<td>0.9737</td>
<td>0.9860</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.0152)</td>
<td>(0.0031)</td>
<td>(0.0025)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9415</td>
<td>0.9800</td>
<td>0.9981</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.0086)</td>
<td>(0.0086)</td>
<td>(0.0090)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 4. Proofs of the main results

Let

$$Y_t^{n,\epsilon} := X_{[nt]/n}^\epsilon.$$
Fig. 1. A sample path of \((Y, R)\) when \(\epsilon = 0.3\) and \(\sigma = 1.0\). Due to large jumps and fluctuations, it seems hard to estimate the drift parameters \(\mu_1\) and \(\mu_2\).

Fig. 2. A sample path of \((Y, R)\) when \(\epsilon = 0.3\) and \(\sigma = 0.3\). The jumps and fluctuations of \(Y\) and \(R\) are smaller than those in Fig. 1 so that the estimation of \(\mu_1\) and \(\mu_2\) is easier.

**Lemma 4.1.** Suppose that Conditions (A1) and (A2) hold. Then the sequence \(\{Y_t^{n, \epsilon}\}\) converges uniformly on compacts in probability to the deterministic process \(\{X_t^0\}\) as \(\epsilon \to 0\) and \(n \to \infty\).
We see that both tails of $\hat{\mu}_1$ and $\hat{\mu}_2$ are heavier than a Gaussian distribution.

**Proof.** Let $J_t := N([0, t] \times \{|u| > 1\})$. It is easy to see that $(J_t : t \geq 0)$ is a Poisson process with intensity $\lambda = \mu(\{|u| > 1\})$. Let $\tau_1 < \tau_2 < \cdots < \tau_n < \cdots$ be the jump times of $\{J_t\}$. Obviously
\[
\lim_{n \to \infty} \tau_n = \infty \text{ a.s. Then}
\]
\[
\int_0^t \int_{|u| > 1} uN(ds, du) = \sum_{i=1}^{J_t} \xi_i,
\]
where \(\{\xi_i : i = 1, 2, \ldots\}\) are i.i.d. \(\mathbb{R}^r\)-valued random variables with common probability distribution \(\mu_0(\cdot)/\lambda\). Here \(\mu_0(\cdot) = \mu(\cdot \cap \{|u| > 1\})\). For the convenience of statement, we also write \(\tau_0 = 0\) and \(\xi_0 = 0 \in \mathbb{R}^r\). Here we shall use the standard interlacing procedure developed in Applebaum [1] to construct the solution \(X^\varepsilon\) of (1.2). Let \(d\)-dimensional process \(\{Z^\varepsilon_t(i) : t \geq 0\}\) be the unique strong solution of the following SDE:
\[
Z^\varepsilon_t(i) = Z^\varepsilon_0(i) + \int_0^t b(Z^\varepsilon_s(i), \theta)ds + \varepsilon \int_0^t \sigma(Z^\varepsilon_s(i))dB_s(i)
\]
\[+ \varepsilon \int_0^t \int_{|u| \leq 1} \sigma(Z^\varepsilon_s(i))u\tilde{N}_i(ds, du), \tag{4.9}\]
where \(B_t(i) = B_{\tau_{i-1}+t} - B_{\tau_{i-1}}\) and \(N_i([0, t] \times A) = N([\tau_{i-1}, \tau_{i-1}+t] \times A)\) for any \(A \in \mathcal{B}(\mathbb{R}^r)\), and the initial value \(Z^\varepsilon_0(i)\) is given by
\[
Z^\varepsilon_0(i) = X^\varepsilon_{\tau_{i-1}}.
\]
Then by condition (A1) it is not hard to see that
\[
Z^\varepsilon_t(i) = \begin{cases} X^\varepsilon_{\tau_{i-1}+t}, & 0 \leq t < \tau_i - \tau_{i-1}, \\ X^\varepsilon_{\tau_{i-1}}, & t = \tau_i - \tau_{i-1}, \end{cases} \tag{4.10}\]
and \(X^\varepsilon_{\tau_i} = X^\varepsilon_{\tau_{i-1}} + \varepsilon \sigma(X^\varepsilon_{\tau_{i-1}})\xi_i\).

**Step 1:** we consider the case of \(i = 1\). Here \(Z^\varepsilon_0(1) = x \in \mathbb{R}^d\). For any fixed \(\varepsilon > 0\), let \(\bar{\tau}_M^\varepsilon = \inf\{t : |Z^\varepsilon_t(1)| > M\} \text{ or } |Z^\varepsilon_0(1)| > M\} \text{ and let } \tilde{f} \in C^2_b(\mathbb{R}^d) \text{ be chosen such that } f(x) = |x|^2 \text{ if } |x| \leq M + K(M + 1). By Itô’s Formula, we have
\[
\tilde{f}(Z^\varepsilon_t(1)) - f(x) - \int_0^{t \wedge \bar{\tau}_M^\varepsilon} Af(Z^\varepsilon_s(1))ds
\]
is a martingale, where
\[
Af(x) = \sum_{k=1}^d b^k(x, \theta) \frac{\partial f}{\partial x^k}(x) + \frac{1}{2} \varepsilon^2 \sum_{k,j=1}^d \sum_{l=1}^r \sigma^k_l(x)\sigma^j_l(x) \frac{\partial^2}{\partial x^k \partial x^j} f(x)
\]
\[+ \int_{|u| \leq 1} \left[ f(x + \varepsilon \sigma(x)u) - f(x) - \varepsilon \sum_{k=1}^d \sum_{l=1}^r \sigma^k_l(x)u_l \frac{\partial}{\partial x^k} f(x) \right] \mu(du).\]
Consider
\[
Z^\varepsilon_t(1) - Z^\varepsilon_0(1) = \int_0^t (b(Z^\varepsilon_s(1), \theta) - b(Z^\varepsilon_0(1), \theta))ds + \varepsilon \int_0^t \sigma(Z^\varepsilon_s(1))dB_s(1)
\]
\[+ \varepsilon \int_0^t \int_{|u| \leq 1} \sigma(Z^\varepsilon_s(1))u\tilde{N}_1(ds, du). \tag{4.11}\]
Using some routine techniques such as Gronwall and Doob’s inequalities, we can show that
\[
E \left[ \sup_{0 \leq t \leq T} \left| Z^\varepsilon_t(1) - Z^\varepsilon_0(1) \right| \right] \leq c\varepsilon \left( T + \int_0^T (1 + |x|^2)e^{2cs}ds \right)^{1/2} e^{KT} \tag{4.12}\]
and consequently
\[ \lim_{\varepsilon \to 0} E \left[ \sup_{0 \leq t \leq T} \left| Z_{t}^{\varepsilon} (1) - Z_{t}^{0} (1) \right| \right] = 0. \] (4.13)

For any small \( \delta > 0 \), there exists a sufficiently large \( T \) such that \( P(\tau_{1} > T) < \delta \). Note that \( (X_{t}^{\varepsilon}) \) is càdlàg and \( (X_{t}^{0}) \) is continuous. By (4.13),
\[
P \left( \sup_{0 \leq t \leq \tau_{1}} \left| X_{t}^{\varepsilon} - X_{t}^{0} \right| > \gamma \right) \leq P \left( \sup_{0 \leq t \leq \tau_{1}} \left| X_{t}^{\varepsilon} - X_{t}^{0} \right| > \gamma; \tau_{1} \leq T \right) + P(\tau_{1} > T)
\[
\leq P \left( \sup_{0 \leq t \leq T} \left| Z_{t}^{\varepsilon} (1) - Z_{t}^{0} (1) \right| > \gamma \right) + \delta.
\]

Then
\[
\lim_{\varepsilon \to 0} P \left( \sup_{0 \leq t \leq \tau_{1}} \left| X_{t}^{\varepsilon} - X_{t}^{0} \right| > \gamma \right) = 0.
\] (4.14)

On the other hand, \( X_{\tau_{1}}^{\varepsilon} = X_{\tau_{1}}^{\varepsilon} + \varepsilon \sigma(X_{\tau_{1}}^{\varepsilon})\xi_{1} \). Since \( X_{\tau_{1}}^{\varepsilon} \xrightarrow{P} X_{\tau_{1}}^{0} \) as \( \varepsilon \to 0 \), it follows from condition (A2) and (4.14) that
\[
\left| X_{\tau_{1}}^{\varepsilon} - X_{\tau_{1}}^{0} \right| \leq \left| X_{\tau_{1}}^{\varepsilon} - X_{\tau_{1}}^{0} \right| + \varepsilon K(1 + |X_{\tau_{1}}^{\varepsilon}|)|\xi_{1}| \xrightarrow{P} 0,
\]
as \( \varepsilon \to 0 \). Then
\[
\lim_{\varepsilon \to 0} P \left( \sup_{0 \leq t \leq \tau_{1}} \left| X_{t}^{\varepsilon} - X_{t}^{0} \right| > \gamma \right) = 0.
\] (4.15)

**Step 2:** Consider \( \{Z_{t}^{\varepsilon} (2) : t \geq 0\} \). Here \( Z_{0}^{\varepsilon} (2) = X_{\tau_{1}}^{\varepsilon} \). Note that
\[
Z_{t}^{\varepsilon} (2) - Z_{t}^{0} (2) = X_{\tau_{1}}^{\varepsilon} - X_{\tau_{1}}^{0} + \int_{0}^{t} (b(Z_{s}^{\varepsilon} (2), \theta) - b(Z_{s}^{0} (2), \theta))ds
\[
+ \varepsilon \int_{0}^{t} \sigma(Z_{s}^{\varepsilon} (2))dB_{s} (2)
\[
+ \varepsilon \int_{0}^{t} \int_{|u| \leq 1} \sigma(Z_{s}^{\varepsilon} (2))u\tilde{N}_{2} (ds, du).
\]
Using some similar arguments as in step 1, we can show that
\[
\lim_{\varepsilon \to 0} E \left[ \sup_{0 \leq t \leq T} \left| Z_{t}^{\varepsilon} (2) - Z_{t}^{0} (2) \right| |X_{\tau_{1}}^{0} + |X_{\tau_{1}}^{0}| \leq M \right] \right] = 0.
\]

For any small \( \delta > 0 \), there exists a \( M > 0 \) such that \( P(|X_{\tau_{1}}^{0}| > M/4) < \delta \). Thus, we have
\[
\lim_{\varepsilon \to 0} \sup_{0 \leq t \leq T} P \left( \sup_{0 \leq t \leq T} \left| Z_{t}^{\varepsilon} (2) - Z_{t}^{0} (2) \right| > \gamma \right)
\[
\leq \lim_{\varepsilon \to 0} \sup_{0 \leq t \leq T} P(\sup_{0 \leq t \leq T} \left| Z_{t}^{\varepsilon} (2) - Z_{t}^{0} (2) \right| > \gamma; |X_{\tau_{1}}^{\varepsilon}| + |X_{\tau_{1}}^{0}| \leq M)
\[
+ \lim_{\varepsilon \to 0} \sup_{0 \leq t \leq T} P(|X_{\tau_{1}}^{\varepsilon} + X_{\tau_{1}}^{0}| > M/2) + P(|X_{\tau_{1}}^{0}| > M/4)
\[
< \delta.
\]
Then as \( \varepsilon \to 0 \), \( P(\sup_{0 \leq t \leq T}|Z^n_{\varepsilon}(t) - Z^0_{\varepsilon}(t)| > \gamma) \to 0 \). Note that \( X^\varepsilon_{\tau_2} = X^\varepsilon_{\tau_2-} + \varepsilon \sigma(X^\varepsilon_{\tau_2-})\xi_2 \). Thus as proved in (4.14) and (4.15), we have

\[
\lim_{\varepsilon \to 0} P\left(\sup_{\tau_1 \leq t \leq \tau_2} \left|X^\varepsilon_t - X^0_t\right| > \gamma\right) = 0.
\]

By induction we get for any integer \( i \geq 1 \),

\[
\lim_{\varepsilon \to 0} P\left(\sup_{\tau_{i-1} \leq t \leq \tau_i} \left|X^\varepsilon_t - X^0_t\right| > \gamma\right) = 0. \tag{4.16}
\]

Note that \( \tau_i \to \infty \) almost surely as \( i \to \infty \). For any fixed \( T > 0 \), for any \( \delta > 0 \), there exists \( i_0 \in \mathbb{N} \) such that \( P(\tau_i < T) < \delta/2 \) for \( i \geq i_0 \). Then, by (4.16), we have for each positive integer \( i \), for any \( \delta > 0 \), there exists \( \varepsilon_0 > 0 \) such that when \( \varepsilon < \varepsilon_0 \)

\[
P\left(\sup_{\tau_{i-1} \leq t \leq \tau_i} \left|X^\varepsilon_t - X^0_t\right| > \gamma\right) < \frac{\delta}{2i_0 \cdot 2^i}. \tag{4.17}
\]

**Step 3:** It follows from (4.17) that for any \( T > 0 \),

\[
P(\sup_{0 \leq t \leq T} |X^\varepsilon_t - X^0_t| > \gamma; \tau_{i-1} \leq T \leq \tau_i) \leq \sum_{k=1}^i P\left(\sup_{\tau_{k-1} \leq t \leq \tau_k} \left|X^\varepsilon_t - X^0_t\right| > \gamma\right) < \frac{\delta}{2i_0}.
\]

Then

\[
P(\sup_{0 \leq t \leq T} |X^\varepsilon_t - X^0_t| > \gamma) = P(\sup_{0 \leq t \leq T} |X^\varepsilon_t - X^0_t| > \gamma; \tau_{i_0} < T)
+ P(\sup_{0 \leq t \leq T} |X^\varepsilon_t - X^0_t| > \gamma; \tau_{i_0} \geq T)
\leq P(\tau_{i_0} < T) + \sum_{k=1}^{i_0} P(\sup_{0 \leq t \leq T} |X^\varepsilon_t - X^0_t| > \gamma; \tau_{k-1} \leq T < \tau_k)
< \delta, \tag{4.18}
\]

as \( \varepsilon < \varepsilon_0 \). By using Theorem II.11 of Protter [27], we have \( \sup_{0 \leq t \leq T}|Y^{n,\varepsilon}_t - X^\varepsilon_t| \to 0 \) in probability as \( n \to \infty \) for any fixed \( \varepsilon > 0 \) since \( X^\varepsilon_t \) is a semi-martingale. Therefore, we conclude that as \( \varepsilon \to 0 \) and \( n \to \infty \),

\[
P(\sup_{0 \leq t \leq T} |Y^{n,\varepsilon}_t - X^0_t| > \gamma) \to 0. \quad \Box
\]

We shall use \( \nabla_x f(x, \theta) = (\partial_{x_1} f(x, \theta), \ldots, \partial_{x_d} f(x, \theta))^T \) to denote the gradient operator of \( f(x, \theta) \) with respect to \( x \).

**Lemma 4.2.** Let \( f \in C^{1,1}_\uparrow(\mathbb{R}^d \times \Theta; \mathbb{R}) \). Assume (A1) and (A2). Then, we have

\[
\frac{1}{n} \sum_{k=1}^n f(X^\varepsilon_{h_{k-1}}, \theta) \overset{P_{n_0}}{\longrightarrow} \int_0^1 f(X^0_t, \theta) \, ds
\]

as \( \varepsilon \to 0 \) and \( n \to \infty \), uniformly in \( \theta \in \Theta \).

**Proof.** The proof is essentially the same as that of Lemma 3.3 of Long et al. [21].
Let
\[ \tau^e_m = \inf\{ t \geq 0 : |X^e_t| \geq m \text{ or } |X^e_{t-}| \geq m \}, \]
\[ \tau^0_m = \inf\{ t \geq 0 : |X^0_t| \geq m \}. \]

**Lemma 4.3.** Assume (A1) and (A2). We have that for any \( m > 0 \) \( \tau^e_m \xrightarrow{p} \tau^0_m \) as \( \varepsilon \to 0 \).

**Proof.** Let \( D([0, \infty), \mathbb{R}^d) \) be the space of all càdlàg functions: \([0, \infty) \to \mathbb{R}^d\) equipped with Skorokhod topology. For \( \omega \in D([0, \infty), \mathbb{R}^d) \), define \( \tau^e_m(\omega) = \inf\{ t \geq 0 : |\omega_t| \geq m \text{ or } |\omega_{t-}| \geq m \} \). By Proposition VI.2.11 of Jacod and Shiryaev [13], the function \( \omega \to \tau^e_m(\omega) \) is continuous at each point \( \omega \) such that \( m \not\in \{m > 0 : \tau^e_m(\omega) < \tau^e_m(\omega)\} \). Note that the process \( \{X^0_t\} \) is continuous and by **Lemma 4.1** we also have \( X^e_t \xrightarrow{P} X^0_t \) in the topology of Skorokhod space \( D([0, \infty), \mathbb{R}^d) \); see (4.18). It follows from the continuous mapping theorem that \( \tau^e_m \xrightarrow{P} \tau^0_m \) as \( \varepsilon \to 0 \). \( \square \)

**Lemma 4.4.** Let \( f \in C^{1,1}_{\ell}(\mathbb{R}^d \times \Theta; \mathbb{R}) \). Assume (A1) and (A2). Then, we have that for \( 1 \leq i \leq d \),
\[ \sum_{k=1}^{n} f(X^e_{t_{k-1}}, \theta)(X^e_{t_k} - X^e_{t_{k-1}} - b^i(X^e_{t_{k-1}}, \theta) \Delta t_{k-1}) \xrightarrow{P} 0 \]
as \( \varepsilon \to 0 \) and \( n \to \infty \), uniformly in \( \theta \in \Theta \), where \( X^e_{t_k} \) and \( b^i \) are the \( i \)th components of \( X^e_t \) and \( b \), respectively.

**Proof.** The proof is very similar to that of Lemma 3.5 in Long et al. [21]. Here we just point out the main differences. It is easy to see that
\[ \sum_{k=1}^{n} f(X^e_{t_{k-1}}, \theta)(X^e_{t_k} - X^e_{t_{k-1}} - b^i(X^e_{t_{k-1}}, \theta) \Delta t_{k-1}) \]
\[ = \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} f(X^e_{t_{k-1}}, \theta)(b^i(X^e_s, \theta_0) - b^i(X^e_{t_{k-1}}, \theta_0)) ds \]
\[ + \varepsilon \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} f(X^e_{t_{k-1}}, \theta) \sum_{l=1}^{r} \sigma_l^i(X^e_s) dL^l_s \]
\[ = \int_0^t f(Y^{n,e}_{s}, \theta)(b^i(X^e_s, \theta_0) - b^i(Y^{n,e}_{s}, \theta_0)) ds + \varepsilon \int_0^t f(Y^{n,e}_{s}, \theta) \sum_{l=1}^{r} \sigma_l^i(X^e_s) dL^l_s. \]

By the given condition on \( f \) and the Lipschitz condition on \( b \), the first term on the right hand side converges to zero in probability as \( \varepsilon \to 0 \) and \( n \to \infty \), uniformly in \( \theta \in \Theta \) by **Lemma 4.1**. By (1.1), let \( M_t = \int_{0}^{t} \int_{|z| > 1} z N(ds, dz) \) with
\[ [M, M]_t = t + \int_{0}^{t} \int_{|z| \leq 1} |z|^2 N(ds, dz). \quad (4.19) \]

Then we have
\[ \sup_{\theta \in \Theta} \left| \varepsilon \int_0^t f(Y^{n,e}_{s}, \theta) \sum_{l=1}^{r} \sigma_l^i(X^e_s) dL^l_s \right| \leq \varepsilon \sup_{\theta \in \Theta} \left| \varepsilon \int_0^t f(Y^{n,e}_{s}, \theta) \sum_{l=1}^{r} \sigma_l^i(X^e_s) dM^l_s \right| \]
\[ + \varepsilon \sup_{\theta \in \Theta} \left| \int_0^t \int_{|z| > 1} f(Y^{n,e}_{s}, \theta) \sum_{l=1}^{r} \sigma_l^i(X^e_s) z_l N(ds, dz) \right|. \]
Next by using Lemma 4.3 and some similar arguments as in the proof of Lemma 3.5 of Long et al. [21], we can show that

$$\varepsilon \sup_{\theta \in \Theta} \left| \int_0^1 f(Y^{n,\varepsilon}_s, \theta) \sum_{l=1}^r \sigma_i^l(X^{\varepsilon}_s) dL^l_s \right| \xrightarrow{P} 0$$

as $\varepsilon \to 0$ and $n \to \infty$. Thus, the proof is complete. □

Since $X^0_t$ is continuous in $t \in [0, 1]$, by Condition (A4) we can find a compact set $F \subset \mathcal{U}$ such that $X^0_t \in F$ for any $t \in [0, 1]$. Then there exists some real number $\zeta > 0$ and some compact set $\tilde{F}$ defined by

$$\tilde{F} = \{ x \in \mathbb{R}^d : d(x, F) \leq \zeta \}$$

such that $F \subset \tilde{F} \subset \mathcal{U}$. For the above $\tilde{F}$, it follows from Lemma 6 of [11] that there exists some smooth functions $\tilde{\sigma}$ such that $\tilde{\sigma}(x) = \sigma(x)$ for $x \in \tilde{F}$; $\inf_{x \in \mathbb{R}^d} \det \tilde{\sigma} \tilde{\sigma}^*(x) > 0$ and $\tilde{\sigma}$ is constant except on some compact set. Then the functions $\tilde{\sigma}$ and $[\tilde{\sigma} \tilde{\sigma}^*]^{-1}$ are bounded and smooth with bounded derivatives of any order on $\mathbb{R}^d$. Let

$$\tilde{\Psi}_{n,\varepsilon}(\theta) = \sum_{k=1}^n \varepsilon^{-2} n P_k^*(\theta) \tilde{\Lambda}_{k-1}^{-1} P_k(\theta),$$

where $P_k(\theta)$ is defined as in (2.4) and $\Lambda_k = [\tilde{\sigma} \tilde{\sigma}^*](X_{\theta_k})$. Let $\tilde{\Phi}_{n,\varepsilon}(\theta) = \varepsilon^2 (\tilde{\Psi}_{n,\varepsilon}(\theta) - \tilde{\Psi}_{n,\varepsilon}(\theta_0))$.

Define

$$\tilde{\theta}_{n,\varepsilon} = \arg\min_{\theta \in \Theta} \tilde{\Psi}_{n,\varepsilon}(\theta),$$

or equivalently,

$$\tilde{\theta}_{n,\varepsilon} = \arg\min_{\theta \in \Theta} \tilde{\Phi}_{n,\varepsilon}(\theta).$$

**Lemma 4.5.** Assume Conditions (A1)–(A5). Then $P(\hat{\theta}_{n,\varepsilon} \neq \tilde{\theta}_{n,\varepsilon}) \to 0$ as $\varepsilon \to 0$ and $n \to \infty$.

**Proof.** Since $X^0_t$ is continuous in $t \in [0, 1]$, it follows from Lemma 4.1 that

$$P(\exists t \in [0, 1], Y^{n,\varepsilon}_t \notin \tilde{F}) \leq P \left( \sup_{0 \leq t \leq 1} |Y^{n,\varepsilon}_t - X^0_t| > \zeta \right) \to 0,$$

as $\varepsilon \to 0$ and $n \to \infty$. On the event that $Y^{n,\varepsilon}_t \in \tilde{F}$ for all $t \in [0, 1]$, $\tilde{\Psi}_{n,\varepsilon}(\theta) = \Psi_{n,\varepsilon}(\theta)$ and $\tilde{\Phi}_{n,\varepsilon}(\theta) = \Phi_{n,\varepsilon}(\theta)$, which also implies that $\hat{\theta}_{n,\varepsilon} = \tilde{\theta}_{n,\varepsilon}$. Thus

$$P(\hat{\theta}_{n,\varepsilon} \neq \tilde{\theta}_{n,\varepsilon}) \leq P(\exists t \in [0, 1], Y^{n,\varepsilon}_t \notin \tilde{F}),$$

which goes to 0 as $\varepsilon \to 0$ and $n \to \infty$. □

Now we are in a position to prove Theorem 2.1.
Proof of Theorem 2.1. The proof is similar to that of Theorem 2.1 in Long et al. [21]. We just sketch the ideas. Note that
\[
\tilde{\phi}_{n, \varepsilon}(\theta) = -2 \sum_{k=1}^{n} (b(X^E_{t_k-1}, \theta) - b(X^E_{t_k-1}, \theta_0))^* \tilde{A}_{k-1}^{-1}(X^E_{t_k} - X^E_{t_{k-1}} - n^{-1}b(X^E_{t_k-1}, \theta_0))
\]
\[
+ \frac{1}{n} \sum_{k=1}^{n} (b(X^E_{t_k-1}, \theta) - b(X^E_{t_k-1}, \theta_0))^* \tilde{A}_{k-1}^{-1}(b(X^E_{t_k-1}, \theta) - b(X^E_{t_k-1}, \theta_0))
\]
\[
:= \tilde{\phi}^{(1)}_{n, \varepsilon}(\theta) + \tilde{\phi}^{(2)}_{n, \varepsilon}(\theta).
\]
By using Lemmas 4.2 and 4.4, we can show that
\[
\sup_{\theta \in \Theta} |\tilde{\phi}_{n, \varepsilon}(\theta) - F(\theta)| \stackrel{P_{\theta_0}}{\longrightarrow} 0
\]
as \(\varepsilon \to 0\) and \(n \to \infty\), where
\[
F(\theta) = \int_{0}^{1} (b(X^0_t, \theta) - b(X^0_t, \theta_0))^*[\tilde{\sigma}\tilde{\sigma}^*]^{-1}(X^0_t)(b(X^0_t, \theta) - b(X^0_t, \theta_0))dt.
\]
In addition, (A5) and the continuity of \(X^0\) yield that
\[
\inf_{|\theta - \theta_0| > \delta} F(\theta) > F(\theta_0) = 0,
\]
for each \(\delta > 0\). Therefore, by Theorem 5.9 of van der Vaart [45], we have \(\tilde{\theta}_{n, \varepsilon} \stackrel{P_{\theta_0}}{\longrightarrow} \theta_0\) as \(\varepsilon \to 0\) and \(n \to \infty\). It follows from Lemma 4.5 that for any \(\xi > 0\),
\[
P(|\tilde{\theta}_{n, \varepsilon} - \theta_0| > \xi) \leq P(|\tilde{\theta}_{n, \varepsilon} - \theta_0| > \xi) + P(\tilde{\theta}_{n, \varepsilon} \neq \tilde{\theta}_{n, \varepsilon}),
\]
which goes to 0 as \(\varepsilon \to 0\) and \(n \to \infty\). \(\Box\)

Note that
\[
\nabla_\theta \tilde{\phi}_{n, \varepsilon}(\theta) = -2 \sum_{k=1}^{n} (\nabla_\theta b)^*(X^E_{t_k-1}, \theta) \tilde{A}_{k-1}^{-1}(X^E_{t_k} - X^E_{t_{k-1}} - b(X^E_{t_k-1}, \theta) \Delta t_{k-1}).
\]
Let \(G_{n, \varepsilon}(\theta) = (G^1_{n, \varepsilon}, \ldots, G^p_{n, \varepsilon})^*\) with
\[
G^i_{n, \varepsilon}(\theta) = \sum_{k=1}^{n} (\partial_{\theta_i} b)^*(X^E_{t_k-1}, \theta) \tilde{A}_{k-1}^{-1}(X^E_{t_k} - X^E_{t_{k-1}} - b(X^E_{t_k-1}, \theta) \Delta t_{k-1}), \quad i = 1, \ldots, p.
\]
Let
\[
K_{n, \varepsilon}(\theta) = \nabla_\theta G_{n, \varepsilon}(\theta),
\]
which is a \(p \times p\) matrix consisting of elements \(K^i_{n, \varepsilon}(\theta) = \partial_{\theta_j} G^i_{n, \varepsilon}(\theta), 1 \leq i, j \leq p\). Moreover, we introduce the following function
\[
K^{ij}(\theta) = \int_{0}^{1} (\partial_{\theta_i} \partial_{\theta_j} b)^*(X^0_s, \theta)[\sigma\sigma^*]^{-1}(X^0_s)(b(X^0_s, \theta_0) - b(X^0_s, \theta))ds - I^{ij}(\theta),
\]
\(1 \leq i, j \leq p\).
Then we define the matrix function

\[ K(\theta) = (K_{ij}(\theta))_{1 \leq i, j \leq p}. \]  

(4.21)

Since \( X^0_t \in F \subset \tilde{F} \) for all \( t \in [0, 1] \), we also have that

\[ F(\theta) = \int_0^1 (b(X^0_t, \theta) - b(X^0_t, \theta_0)) * (\sigma_\sigma)^{-1}(X^0_t)(b(X^0_t, \theta) - b(X^0_t, \theta_0)) dt. \]

**Lemma 4.6.** Let \( f \in C^{1,1}_d(\mathbb{R}^d \times \Theta; \mathbb{R}) \). Assume (A1), (A2) and (A5). Then we have that for \( 1 \leq i \leq d \) and each \( \theta \in \Theta \),

\[ \sum_{k=1}^n f(X_{t_k-1}^0, \theta) \int_{t_{k-1}}^{t_k} \sum_{l=1}^{r} \sigma_l^j(X_{s-}^0) dL_s^l \to \int_0^1 f(X_s^0, \theta) \sum_{l=1}^{r} \sigma_l^j(X_{s-}^0) dL_s^l, \]

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

**Proof.** Note that

\[ \sum_{k=1}^n f(X_{t_k-1}^0, \theta) \int_{t_{k-1}}^{t_k} \sum_{l=1}^{r} \sigma_l^j(X_{s-}^0) dL_s^l = \int_0^1 f(Y_{s-}^{n,0}, \theta) \sum_{l=1}^{r} \sigma_l^j(X_{s-}^0) dL_s^l. \]

Recall (1.1) and \( M_t = L_t - \int_0^1 \int_{|z|>1} zN(ds, dz) \). Then

\[ \int_0^1 f(Y_{s-}^{n,0}, \theta) \sum_{l=1}^{r} \sigma_l^j(X_{s-}^0) dL_s^l = \int_0^1 \int_{|z|>1} f(Y_{s-}^{n,0}, \theta) \sum_{l=1}^{r} \sigma_l^j(X_{s-}^0) z_l N(ds, dz) \]

\[ + \int_0^1 f(Y_{s-}^{n,0}, \theta) \sum_{l=1}^{r} \sigma_l^j(X_{s-}^0) dM_s^l. \]

Since \( f(x, \theta) \) and \( \sigma_l^j(x) \) are continuous in \( x \) for any fixed \( \theta \), it follows from Ethier and Kurtz [5, Problem 13, p. 151], Lemma 4.1 and the continuous mapping theorem that for any \( T \geq 0 \),

\[ \sup_{0 \leq t \leq T} \left| f(Y_{t}^{n,0}, \theta) \sigma_l^j(X_{t}^0) - f(X_{t}^0, \theta) \sigma_l^j(X_{t}^0) \right| \to 0. \]  

(4.22)

We can use similar arguments in the proof of Lemma 3.4 in Long et al. [21] to show that

\[ \int_0^1 \int_{|z|>1} f(Y_{s-}^{n,0}, \theta) \sum_{l=1}^{r} \sigma_l^j(X_{s-}^0) z_l N(ds, dz) \]

\[ \to \int_0^1 \int_{|z|>1} f(X_{s-}^0, \theta) \sum_{l=1}^{r} \sigma_l^j(X_{s-}^0) z_l N(ds, dz) \]

and

\[ \int_0^1 f(Y_{s-}^{n,0}, \theta) \sum_{l=1}^{r} \sigma_l^j(X_{s-}^0) dM_s^l \to \int_0^1 f(X_{s-}^0, \theta) \sum_{l=1}^{r} \sigma_l^j(X_{s-}^0) dM_s^l \]

as \( \varepsilon \to 0 \) and \( n \to \infty \). \( \square \)
Lemma 4.7. Assume (A1)–(A5) and (B). Then we have that for each $i = 1, \ldots, p,$

$$\varepsilon^{-1} G^i_{n,\varepsilon}(\theta_0) \xrightarrow{P_{\theta_0}} \int_0^1 (\partial_{\theta_i} b)^*(X^0_t, \theta_0)[\sigma \sigma^*]^{-1}(X^0_t)\sigma(X^0_t) dL_t,$$

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

Proof. Note that for $1 \leq i \leq p,$

$$\varepsilon^{-1} G^i_{n,\varepsilon}(\theta_0) = \varepsilon^{-1} \sum_{k=1}^n (\partial_{\theta_i} b)^*(X^\varepsilon_{tk-1}, \theta_0) \tilde{A}^{-1}_{k-1}(X^\varepsilon_t - X^\varepsilon_{tk-1} - b(X^\varepsilon_{tk-1}, \theta_0) \Delta t_{k-1})$$

$$= \varepsilon^{-1} \sum_{k=1}^n (\partial_{\theta_i} b)^*(X^\varepsilon_{tk-1}, \theta_0) \tilde{A}^{-1}_{k-1} \int_{tk-1}^{tk} (b(X^\varepsilon_s, \theta_0) - b(X^\varepsilon_{tk-1}, \theta_0)) ds$$

$$+ \sum_{k=1}^n (\partial_{\theta_i} b)^*(X^\varepsilon_{tk-1}, \theta_0) \tilde{A}^{-1}_{k-1} \int_{tk-1}^{tk} \sigma(X^\varepsilon_s) dL_s$$

$$:= H^{(1)}_{n,\varepsilon}(\theta_0) + H^{(2)}_{n,\varepsilon}(\theta_0).$$

By using Lemma 4.6 and letting $f_j(x, \theta)$ be the $j$th component of $((\partial_{\theta_i} b(x, \theta))^*[\tilde{\sigma} \tilde{\sigma}^*]^{-1}(x))$ for $1 \leq j \leq d$ with $\theta = \theta_0,$ we have

$$H^{(2)}_{n,\varepsilon}(\theta_0) = \sum_{k=1}^n (\partial_{\theta_i} b)^*(X^\varepsilon_{tk-1}, \theta_0) \tilde{A}^{-1}_{k-1} \int_{tk-1}^{tk} \sigma(X^\varepsilon_s) dL_t$$

$$P_{\theta_0} \xrightarrow{\varepsilon \to 0, n \to \infty} \int_0^1 (\partial_{\theta_i} b)^*(X^0_t, \theta_0)[\tilde{\sigma} \tilde{\sigma}^*]^{-1}(X^0_t)\sigma(X^0_t) dL_t$$

as $\varepsilon \to 0$ and $n \to \infty.$ Note that $[\tilde{\sigma} \tilde{\sigma}^*]^{-1}(X^0_t) = [\sigma \sigma^*]^{-1}(X^0_t)$ for all $t \in [0, 1].$ By using some delicate estimate on the process $X^\varepsilon_t$, the fact that $\sup_{x \in \mathbb{R}^d} ||[\tilde{\sigma} \tilde{\sigma}^*]^{-1}(x)|| \leq K_1$ for some constant $K_1 > 0,$ and similar technical arguments in the proof of Lemma 3.6 in Long et al. [21], we can show that $H^{(1)}_{n,\varepsilon}(\theta_0)$ converges to zero in probability. The details are omitted. \qed

Lemma 4.8. Assume (A1)–(A5) and that $\theta_0 \in \Theta$ with the matrix $I(\theta_0)$ given in (2.6) being positive definite. Then, for $K_{n,\varepsilon}(\theta)$ defined in (4.20) and $K(\theta)$ defined in (4.21), we have

$$\sup_{\theta \in \Theta} |K_{n,\varepsilon}(\theta) - K(\theta)| P_{\theta_0} \xrightarrow{\varepsilon \to 0, n \to \infty} 0$$

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

Proof. The proof is similar to that of Lemma 3.7 in Long et al. [21]. It suffices to prove that for $1 \leq i, j \leq p$

$$\sup_{\theta \in \Theta} |K^{ij}_{n,\varepsilon}(\theta) - K^{ij}(\theta)| P_{\theta_0} \xrightarrow{\varepsilon \to 0, n \to \infty} 0$$

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

$$K^{ij}_{n,\varepsilon}(\theta) = \partial_{\theta_j} G^i_{n,\varepsilon}(\theta)$$

$$= \sum_{k=1}^n (\partial_{\theta_j} \partial_{\theta_i} b)^*(X^\varepsilon_{tk-1}, \theta) \tilde{A}^{-1}_{k-1}(X^\varepsilon_t - X^\varepsilon_{tk-1} - b(X^\varepsilon_{tk-1}, \theta_0) \Delta t_{k-1})$$
By using Lemmas 4.4 and 4.2, we have sup_{\theta \in \Theta} |K_{n,\varepsilon}^{ij,(1)}(\theta)| \overset{P_0}{\longrightarrow} 0 and sup_{\theta \in \Theta} |K_{n,\varepsilon}^{ij,(2)}(\theta) - K^{ij}(\theta)| \overset{P_0}{\longrightarrow} 0 as \varepsilon \to 0 and n \to \infty. \square

**Proof of Theorem 2.2.** First we consider \( S_{n,\varepsilon} = \varepsilon^{-1}(\hat{\theta}_{n,\varepsilon} - \theta_0) \). Using Lemmas 4.7 and 4.8, and following the same proof of Theorem 2.2 in Long et al. [21], we can show that

\[
S_{n,\varepsilon} \overset{P_0}{\longrightarrow} (I(\theta_0))^{-1}\left(\int_0^1 (\partial_\theta b)^*(X_t^\varepsilon, \theta_0)[\sigma\sigma^*]^{-1}(X_t^0)\sigma(X_t^0) dL_t\right)_{1 \leq i \leq p}^*.
\]

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

\[
W(\theta_0) = (I(\theta_0))^{-1}\left(\int_0^1 (\partial_\theta b)^*(X_t^0, \theta_0)[\sigma\sigma^*]^{-1}(X_t^0)\sigma(X_t^0) dL_t\right)_{1 \leq i \leq p}^*.
\]

By Lemma 4.5, for any \( \zeta > 0 \), we have

\[
P(|\varepsilon^{-1}(\hat{\theta}_{n,\varepsilon} - \theta_0) - W(\theta_0)| > \zeta) \leq P(|\varepsilon^{-1}(\hat{\theta}_{n,\varepsilon} - \theta_0) - W(\theta_0)| > \zeta) + P(\hat{\theta}_{n,\varepsilon} \neq \hat{\theta}_{n,\varepsilon}),
\]

which goes to 0 as \( \varepsilon \to 0, n \to \infty \), and \( n\varepsilon \to \infty \). \square

**Acknowledgments**

The authors are grateful to the anonymous referees, the associate editor and the editor for their insightful and valuable comments which have greatly improved the presentation of the paper. The second author was supported by NSFC (Grant Number 11271204). The third author was supported by JSPS KAKENHI (Grant Number 15K05009, 25245033) and JST/CREST.

**References**