Least squares estimators for discretely observed stochastic processes driven by small Lévy noises

Hongwei Long\textsuperscript{a,\ast}, Yasutaka Shimizu\textsuperscript{b}, Wei Sun\textsuperscript{c}

\textsuperscript{a} Department of Mathematical Sciences, Florida Atlantic University, Boca Raton, FL 33431-0991, USA
\textsuperscript{b} Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka 560-8531, Japan
\textsuperscript{c} Department of Mathematics and Statistics, Concordia University, Montreal, Quebec H3G 1M8, Canada

Abstract

We study the problem of parameter estimation for discretely observed stochastic processes driven by additive small Lévy noises. We do not impose any moment condition on the driving Lévy process. Under certain regularity conditions on the drift function, we obtain consistency and rate of convergence of the least squares estimator (LSE) of the drift parameter when a small dispersion coefficient $\varepsilon \to 0$ and $n \to \infty$ simultaneously. The asymptotic distribution of the LSE in our general setting is shown to be the convolution of a normal distribution and a distribution related to the jump part of the Lévy process. Moreover, we briefly remark that our methodology can be easily extended to the more general case of semi-martingale noises.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Let $(\mathcal{S}, \mathcal{F}, \mathbb{P})$ be a basic probability space equipped with a right continuous and increasing family of $\sigma$-algebras $(\mathcal{F}_t, t \geq 0)$. Let $(L_t, t \geq 0)$ be an $\mathbb{R}^d$-valued Lévy process, which is given by

$$L_t = at + \sigma B_t + \int_0^t \int_{|z| \leq 1} z \tilde{N}(ds, dz) + \int_0^t \int_{|z| > 1} zN(ds, dz), \quad (1.1)$$

where $a = (a_1, \ldots, a_d) \in \mathbb{R}^d$, $\sigma = (\sigma_i)_i$ is a $d \times r$ real-valued matrix, $B_t = (B^1_t, \ldots, B^r_t)$ is an $r$-dimensional standard Brownian motion, $N(ds, dz)$ is an independent Poisson random measure on $\mathbb{R}_+ \times (\mathbb{R}^d \setminus \{0\})$ with characteristic measure $dt \nu(dz)$, and $N(ds, dz) = N(ds, dz) - \nu(dz)ds$ is a martingale measure. Here we assume that $\nu(dz)$ is a Lévy measure on $\mathbb{R}^d \setminus \{0\}$ satisfying $\int_{|z| \leq 1} \nu(dz) < \infty$ with $|z| = \sqrt{\sum_{i=1}^d z^2}$. The stochastic process $X = (X_t, t \geq 0)$, starting from $x_0 \in \mathbb{R}^d$, is defined as the unique strong solution to the following stochastic differential equation (SDE)

$$dx_t = b(X_t, \theta) dt + \varepsilon dL_t, \quad t \in [0, 1]; \quad X_0 = x_0. \quad (1.2)$$

where $\theta \in \Theta$ (the closure of $\Theta_0$) with $\Theta_0$ being an open bounded convex subset of $\mathbb{R}^p$, and $b = (b_1, \ldots, b_d) : \mathbb{R}^d \times \Theta \to \mathbb{R}^d$ is a known function. Without loss of generality, we assume that $\varepsilon \in (0, 1]$. The regularity conditions on $b$ will

\textsuperscript{\ast} Corresponding author.
E-mail address: hlong@fau.edu (H. Long).

0047-259X/$-$ see front matter © 2013 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2013.01.012
be provided in Section 2. Assume that this process is observed at regularly spaced time points \( \{t_k = k/n, \ k = 1, 2, \ldots, n\} \). The only unknown quantity in SDE (1.2) is the parameter \( \theta \). Let \( \theta_0 = \theta_0 \) be the true value of the parameter \( \theta \). The purpose of this paper is to study the least squares estimator for the true value \( \theta_0 \) based on the sampling data \( (X_t)_{t=1}^n \) with small dispersion \( \varepsilon \) and large sample size \( n \).

In the case of diffusion processes driven by Brownian motion, a popular method is the maximum likelihood estimator (MLE) based on the Girsanov density when the processes can be observed continuously (see Prakasa Rao [31], Lipster and Shiryaev [19], Kutoyants [16], and Bishwal [2]). When a diffusion process is observed only at discrete times, in most cases the transition density and hence the likelihood function of the observations is not explicitly computable. In order to overcome this difficulty, some approximate likelihood methods have been proposed by Lo [20], Pedersen [27,28], Poulsen [29], and Ait-Sahalia [1]. For a comprehensive review on MLE and other related methods, we refer to Sørensen [37]. The least squares estimator (LSE) is asymptotically equivalent to the MLE. For the LSE, the convergence in probability was proved in Dorogovcev [5] and Le Breton [18], the strong consistency was studied in Kasonga [12], and the asymptotic distribution was studied in Prakasa Rao [30]. For a more recent comprehensive discussion, we refer to Prakasa Rao [31], Kutoyants [16], Bishwal [2] and the references therein.

The parametric estimation problems for diffusion processes with jumps based on discrete observations have been studied by Shimizu and Yoshida [35] and Shimizu [33] via the quasi-maximum likelihood. They established consistency and asymptotic normality for the proposed estimators. Moreover, Ogihara and Yoshida [26] showed some stronger results than the ones by Shimizu and Yoshida [35], and also investigated an adaptive Bayes-type estimator with its asymptotic properties. The driving jump processes considered in Shimizu and Yoshida [35], Shimizu [33] and Ogihara and Yoshida [26] include a large class of Lévy processes such as compound Poisson processes, gamma, inverse Gaussian, variance gamma, normal inverse Gaussian or some generalized tempered stable processes. Masuda [24] dealt with the consistency and asymptotic normality of the TFE (trajectory-fitting estimator) and LSE when the driving process is a zero-mean adapted process (including Lévy process) with finite moments. The parametric estimation for Lévy-driven Ornstein–Uhlenbeck processes was also studied by Brockwell et al. [3], Spiliopoulos [39], and Valdivieso et al. [46]. However, the aforementioned papers were unable to cover an important class of driving Lévy processes, namely \( \alpha \)-stable Lévy motions with \( \alpha \in (0, 2) \). Recently, Hu and Long [9,10] have started the study on parameter estimation for Ornstein–Uhlenbeck processes driven by \( \alpha \)-stable Lévy motions. They obtained some new asymptotic results on the proposed TFE and LSE under continuous or discrete observations, which are different from the classical cases where asymptotic distributions are normal. Fasen [6] extended the results of Hu and Long [10] to multivariate Ornstein–Uhlenbeck processes driven by \( \alpha \)-stable Lévy motions. Masuda [25] proposed a self-weighted least absolute deviation estimator for discretely observed ergodic Ornstein–Uhlenbeck processes driven by symmetric Lévy processes.

The asymptotic theory of parametric estimation for diffusion processes with small white noise based on continuous-time observations has been well developed (see, e.g., Kutoyants [14,15], Yoshida [48,50], Uchida and Yoshida [44]). There have been many applications of small noise asymptotics to mathematical finance, see for example Yoshida [49], Takahashi [40], Kunitomo and Takahashi [13], Takahashi and Yoshida [41], Uchida and Yoshida [45]. 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 [7] and Laredo [17] studied the efficient estimation of drift parameters of small diffusions from discrete observations when \( \varepsilon \to 0 \) and \( n \to \infty \). Sørensen [36] used martingale estimating functions to establish consistency and asymptotic normality of the estimators of drift and diffusion coefficient parameters when \( \varepsilon \to 0 \) and \( n \) is fixed. Sørensen and Uchida [38] and Gloter and Sørensen [8] used a contrast function to study the efficient estimation for unknown parameters in both drift and diffusion coefficient functions. Uchida [42,43] 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 [21] 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 \( \theta \) (\( b(x, \theta) = - \theta x \)), the driving Lévy process is \( L_t = \alpha B_t + b Z_t, \) 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. In a similar framework, Long [22] discussed the statistical estimation of the drift parameter for a class of SDEs with special drift function \( b(x, \theta) = \theta b(x) \). Ma [23] extended the results of Long [21] to the case when the driving noise is a general Lévy process. However, all the drift functions discussed in Long [21,22] and Ma [23] are linear in \( \theta \), which restricts the applicability of their models and results. In this paper, we allow the drift function \( b(x, \theta) \) to be nonlinear in both \( x \) and \( \theta \), and the driving noise to be a general Lévy process. We are interested in estimating the drift parameter in SDE (1.2) 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.

Consider the following contrast function

\[
\Psi_{n, \varepsilon}(\theta) = \sum_{k=1}^{n} \frac{|X_{t_k} - X_{t_{k-1}} - b(X_{t_{k-1}}, \theta) \cdot \Delta t_{k-1}|^2}{\varepsilon^2 \Delta t_{k-1}},
\]
where $\Delta t_{k-1} = t_k - t_{k-1} = 1/n$. Then the LSE $\hat{\theta}_{n,\varepsilon}$ is defined as
\[
\hat{\theta}_{n,\varepsilon} := \arg\min_{\theta \in \Theta} \Psi_{n,\varepsilon}(\theta).
\]
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
\[
\hat{\theta}_{n,\varepsilon} = \arg\min_{\theta \in \Theta} \Phi_{n,\varepsilon}(\theta).
\]
We shall use this fact later for convenience of the proofs.

In the nonlinear case, 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 [47] and Shimizu [34] for a more general criterion) to establish the consistency of the LSE as well as its asymptotic behaviors (asymptotic distribution and rate of convergence). In this paper, we consider the asymptotics of the LSE $\hat{\theta}_{n,\varepsilon}$ with high frequency ($n \to \infty$) and small dispersion ($\varepsilon \to 0$). Our goal is to prove that $\hat{\theta}_{n,\varepsilon} \to \theta_0$ in probability and to establish its rate of convergence and asymptotic distributions. We obtain some new asymptotic distributions for the LSE in our general setting, which are the convolutions of normal distribution and a distribution related to the jump part of the driving Lévy process. Some similar but more general results are also established when the driving Lévy process is replaced by a general semi-martingale.

The paper is organized as follows. In Section 2, we state our main result with some remarks and examples. We establish the consistency of the LSE $\hat{\theta}_{n,\varepsilon}$ and give its asymptotic distribution, which is a natural extension of the classical small-diffusion cases. All the proofs are given in Section 3. In Section 4, we discuss the extension of main results in Section 2 to the general case when the driving noise is a semi-martingale. Some simulation studies are provided in Section 5.

2. Main results

2.1. Notation and assumptions

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 $\partial^m := \partial_{z_1}^{m_1} \cdots \partial_{z_k}^{m_k}$, where $\partial_{z_i}^{m_i} := \partial^{m_i} / \partial z_i^{m_i}$. Let $C^k(\mathbb{R}^d \times \Theta; \mathbb{R}^q)$ 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 $C^{k,l}(\mathbb{R}^d \times \Theta; \mathbb{R}^q)$ is a class of $f \in C^{k,l}(\mathbb{R}^d \times \Theta; \mathbb{R}^q)$ satisfying that $\sup_{\theta \in \Theta} |\partial^\beta f(x, \theta)| \leq C(1 + |x|)^\lambda$ 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 $0 \leq \sum_{i=1}^p \alpha_i \leq l$ and $0 \leq \sum_{i=1}^d \beta_i \leq k$, respectively.

We introduce the following set of assumptions.

(A1) There exists a constant $K > 0$ such that
\[
|b(x, \theta) - b(y, \theta)| \leq K|x - y|; \quad |b(x, \theta)| \leq K(1 + |x|)
\]
for each $x, y \in \mathbb{R}^d$ and $\theta \in \Theta$.

(A2) $b(\cdot, \cdot) \in C^{1,3}(\mathbb{R}^d \times \Theta; \mathbb{R}^d)$.

(A3) $\theta \neq \theta_0 \Leftrightarrow b(X^0_t, \theta) \neq b(X^0_t, \theta_0)$ for at least one value of $t \in [0, 1]$.

(A4) $I(\theta) = (I^i(\theta_0))_{1 \leq i, j \leq p}$ is positive definite, where
\[
I^i(\theta) = \int_0^1 (\partial_0 b)^T(X^0_s, \theta) \partial_0 b(X^0_s, \theta)ds.
\]

It is well-known that SDE (1.2) has a unique strong solution under (A1). For convenience, we shall use $C$ to denote a generic constant whose value may vary from place to place. For a matrix $A$, we define $|A|^2 = \text{tr}(AA^T)$, where $A^T$ is the transpose of $A$. In particular, $|\sigma|^2 = \sum_{i=1}^d \sum_{j=1}^d \sigma_{ij}^2$.

2.2. Asymptotic behavior of LSE

The consistency of our estimator $\hat{\theta}_{n,\varepsilon}$ is given as follows.

**Theorem 2.1.** Under conditions (A1)–(A3), we have
\[
\hat{\theta}_{n,\varepsilon} \xrightarrow{p} \theta_0
\]
as $\varepsilon \to 0$ and $n \to \infty$. 

Author's personal copy
The next theorem gives the asymptotic distribution of $\hat{\theta}_{n,\varepsilon}$. As is easily seen, our result includes the case of Sørensen and Uchida [38] as a special case.

**Theorem 2.2.** Under conditions (A1)–(A4), we have

\[ \varepsilon^{-1}(\hat{\theta}_{n,\varepsilon} - \theta_0) \overset{p}{\to} I^{-1}(\theta_0)S(\theta_0), \quad (2.1) \]

as $\varepsilon \to 0$, $n \to \infty$ and $n\varepsilon \to \infty$, where

\[ S(\theta_0) := \left( \int_0^1 (\partial_{\theta} b)^T(X_{\theta}^0, \theta_0) dL_s, \ldots, \int_0^1 (\partial_{\theta} b)^T(X_{\theta}^0, \theta_0) dL_s \right)^T. \]

**Remark 2.3.** One of our main contributions is that we no longer require any high-order moments condition on $X$ as in, e.g., Sørensen and Uchida [38] and others, which makes our results applicable in many practical situations.

**Remark 2.4.** In general, the limiting distribution on the right-hand side of (2.1) is a convolution of a normal distribution and a distribution related to the jump part of the Lévy process. In particular, if the driving Lévy process $L$ is the linear combination of standard Brownian motion and $\alpha$-stable motion, Theorem 2.2 was discussed in Long [21] and Ma [23].

**Example 2.6.** We consider a one-dimensional stochastic process in (1.2) with drift function $b(x, \theta) = \theta_1 + \theta_2 x$. We assume that the true value $\theta_0 = (\theta_0^0, \theta_0^2)$ of $\theta = (\theta_1, \theta_2)$ belongs to $\Theta_0 = (c_1, c_2) \times (c_3, c_4) \subset \mathbb{R}^2$ with $c_1 < c_2$ and $c_3 < c_4$. Then, $X_0^0$ satisfies the following ODE

\[ dX_t^0 = (\theta_1^0 + \theta_2^0 x^0) dt, \quad X_0^0 = x_0. \quad (2.2) \]

The explicit solution is given by $X_t^0 = e^{\theta_1^0 t} x_0 + \theta_1^0 (e^{\theta_2^0 t} - 1)$ when $\theta_2^0 \neq 0$; $X_t^0 = x_0 + \theta_1^0 t$ when $\theta_2^0 = 0$. The LSE $\hat{\theta}_{n,\varepsilon} = (\hat{\theta}_{n,\varepsilon,1}, \hat{\theta}_{n,\varepsilon,2})^T$ of $\theta_0$ is given by

\[ \hat{\theta}_{n,\varepsilon,1} = (X_1 - X_0) - \frac{1}{n} \sum_{k=1}^n X_{k-1} \]

\[ \hat{\theta}_{n,\varepsilon,2} = \frac{\sum_{k=1}^n (X_k - X_{k-1}) X_{k-1} - (X_1 - X_0) \left( \frac{1}{n} \sum_{k=1}^n X_{k-1} \right)}{\frac{1}{n} \sum_{k=1}^n X_{k-1}^2 - \left( \frac{1}{n} \sum_{k=1}^n X_{k-1} \right)^2}. \]

Note that $\partial_{\theta_1} b(x, \theta) = 1$ and $\partial_{\theta_2} b(x, \theta) = x$. In this case, the limiting random vector in Theorem 2.2 is $I^{-1}(\theta_0)(\int_0^1 dL_s, \int_0^1 X_0^0 dL_s)^T$, where

\[ I(\theta_0) = \left( \begin{array}{cc} \int_0^1 ds & \int_0^1 X_0^0 ds \\ \int_0^1 X_0^0 ds & \int_0^1 (X_0^0)^2 ds \end{array} \right). \]

**Example 2.7.** We consider a one-dimensional stochastic process in (1.2) with drift function $b(x, \theta) = \sqrt{\theta + x^2}$. We assume that the true value $\theta_0$ of $\theta = (c_1, c_2) \subset \mathbb{R}$ with $0 < c_1 < c_2 < \infty$. Then, $X_0^0$ satisfies the following ODE

\[ dX_t^0 = \sqrt{\theta_0 + (X_0^0)^2} dt, \quad X_0^0 = x_0. \]

The explicit solution is given by $X_t^0 = \frac{(x_0 + \sqrt{x_0^2 + \theta_0 t^2} - \sqrt{\theta_0})}{2(x_0 + \sqrt{x_0^2 + \theta_0 t^2})}$. It is easy to verify that the LSE $\hat{\theta}_{n,\varepsilon}$ of $\theta$ is a solution to the following nonlinear equation

\[ \sum_{k=1}^n \frac{X_k - X_{k-1}}{\sqrt{\theta + X_{k-1}^2}} = 1. \]
Since it is impossible to get the explicit expression for \( \hat{\theta}_{n,v} \), we solve the above equation numerically (e.g. by using Newton’s method). Note that \( \partial_\theta b(x, \theta) = \frac{1}{2\sqrt{\theta + x^2}} \). It is clear that the limiting random variable in Theorem 2.2 is
\[
I^{-1}(\theta_0) \int_0^1 \frac{1}{2\sqrt{\theta_0 + x^2}} dl, \text{ where } I(\theta_0) = \int_0^1 \frac{1}{4(\theta_0 + x^2)} ds.
\]
In particular, we assume that \( L_t = aB_t + \sigma Z_t \), where \( B_t \) is the standard Brownian motion and \( Z_t \) is a standard \( \alpha \)-stable Lévy motion independent of \( B_t \). Let us denote by \( N \) a random variable with the standard normal distribution and \( U \) a random variable with the standard \( \alpha \)-stable distribution \( S_\alpha(1, \beta, 0) \), where \( \alpha \in (0, 2) \) is the index of stability and \( \beta \in [-1, 1] \) is the skewness parameter. By using the self-similarity and time change, we can easily show that the limiting random variable in Theorem 2.2 has distribution given by
\[
al^{-\frac{1}{2}}(\theta_0)N + \sigma I^{-1}(\theta_0) \left[ \int_0^1 \left( \frac{1}{2\sqrt{\theta_0 + x^2}} \right)^d \right]^{1/7} U.
\]

**Example 2.8.** We consider a two-dimensional stochastic process in (1.2) with drift function \( b(x, \theta) = C + Ax \), where \( C = (c_1, c_2)^T \), \( A = (A_{ij})_{1 \leq i, j \leq 2} \) and \( x = (x_1, x_2)^T \). We assume that the eigenvalues of \( A \) have positive real parts. We want to estimate \( \theta = (\theta_1, \ldots, \theta_6)^T = (\theta_1, A_{11}, A_{12}, c_2, A_{21}, A_{22})^T \in \Theta \subset \mathbb{R}^6 \), whose true value is \( \theta_0 = (\theta_1, A_{11}^0, A_{12}^0, c_2^0, A_{21}^0, A_{22}^0)^T \). Then \( X_0^2 \) satisfies the following ODE
\[
dX_t^0 = (C_0 + A_0X_t^0)dt, \quad X_0^0 = x_0.
\]
The explicit solution is given by \( X_t^0 = e^{t\theta}X_0 + \int_0^t e^{\lambda(t-s)}C_0ds \). After some basic calculation, we find that the LSE \( \hat{\theta}_{n,v} = (\hat{\theta}_{n,v,1}, \ldots, \hat{\theta}_{n,v,6}) \) is given by
\[
\left( \hat{\theta}_{n,v,1}, \ldots, \hat{\theta}_{n,v,6} \right) = A_n^{-1} \left( \begin{array}{c} n \sum_{k=1}^n Y_k^{(1)}X_k^{(1)} \\ n \sum_{k=1}^n Y_k^{(1)}X_k^{(2)} \\ n \sum_{k=1}^n Y_k^{(2)}X_k^{(1)} \\ n \sum_{k=1}^n Y_k^{(2)}X_k^{(2)} \end{array} \right),
\]
where \( Y_k^{(i)} \) and \( X_k^{(i)} \) are the components of \( X_k \), \( Y_k^{(i)} \) and \( X_k^{(i)} \) are the components of \( Y_k = X_k - X_{k-1} \), and
\[
A_n = \left( \begin{array}{ccc} n \sum_{k=1}^n X_k^{(1)} & n \sum_{k=1}^n X_k^{(2)} & n \sum_{k=1}^n X_k^{(1)}X_k^{(2)} \\ n \sum_{k=1}^n X_k^{(1)} & (X_k^{(1)})^2 & n \sum_{k=1}^n X_k^{(2)}X_k^{(1)} \\ n \sum_{k=1}^n X_k^{(2)} & n \sum_{k=1}^n X_k^{(1)}X_k^{(2)} & (X_k^{(2)})^2 \end{array} \right).
\]
Since it is easy and straightforward to compute the partial derivatives \( \partial_\theta b(x, \theta) \), \( 1 \leq i \leq 6 \), and the limiting random vector in Theorem 2.2, we omit the details here.

3. Proofs

3.1. Proof of Theorem 2.1

We first establish some preliminary lemmas. In the sequel, we shall use the notation
\[
Y_t^{n,v} := X_{[nt]/n}
\]
for the stochastic process \( X \) defined by (1.2), where \( [nt] \) denotes the integer part of \( nt \).

**Lemma 3.1.** The sequence \( \{Y_t^{n,v}\} \) converges to the deterministic process \( \{X_t^0\} \) uniformly on compacts in probability as \( v \rightarrow 0 \) and \( n \rightarrow \infty \).

**Proof.** Note that
\[
X_t - X_t^0 = \int_0^t (b(X_s, \theta_0) - b(X_s^0, \theta_0)) ds + vL_t.
\]
By the Lipschitz condition on $b(\cdot)$ in (A1) and the Cauchy–Schwarz inequality, we find that

$$|X_t - X^0_t|^2 \leq 2 \left[ \int_0^t (b(X_s, \theta_0) - b(X^0_s, \theta_0)) ds \right]^2 + 2e^2 |L_t|^2$$

$$\leq 2t \int_0^t |b(X_s, \theta_0) - b(X^0_s, \theta_0)|^2 ds + 2e^2 \sup_{0 \leq s \leq t} |L_s|^2$$

$$\leq 2K^2 t \int_0^t |X_s - X^0_s|^2 ds + 2e^2 \sup_{0 \leq s \leq t} |L_s|^2.$$  

By Gronwall’s inequality, it follows that

$$|X_t - X^0_t|^2 \leq 2e^2 e^{2K^2 t} \sup_{0 \leq s \leq t} |L_s|^2$$

and consequently

$$\sup_{0 \leq t \leq T} |X_t - X^0_t| \leq \sqrt{2e} e^{2K^2 t} \sup_{0 \leq s \leq T} |L_s|,$$  \hspace{1cm} (3.1)

which goes to zero in probability as $\varepsilon \to 0$ for each $T > 0$. Since $[nt]/n \to t$ as $n \to \infty$, we conclude that the statement holds. \hfill $\Box$

**Lemma 3.2.** Let $\tau^{n, \varepsilon} = \inf\{t \geq 0 : |X^0_t| \geq m \text{ or } |Y^t_{n, \varepsilon}| \geq m\}$. Then, $\tau^{n, \varepsilon} \to \infty$ a.s. uniformly in $n$ and $\varepsilon$ as $m \to \infty$.

**Proof.** Note that

$$X_t = X_0 + \int_0^t b(X_s, \theta_0) ds + \varepsilon L_t.$$  

By the linear growth condition on $b$ and the Cauchy–Schwarz inequality, we get

$$|X_t|^2 \leq 2(|X_0| + \varepsilon |L_t|)^2 + 2 \left( \int_0^t b(X_s, \theta_0) ds \right)^2$$

$$\leq 2 \left( |X_0| + \varepsilon \sup_{0 \leq s \leq t} |L_s| \right)^2 + 2t \int_0^t |b(X_s, \theta_0)|^2 ds$$

$$\leq 2 \left( |X_0| + \varepsilon \sup_{0 \leq s \leq t} |L_s| \right)^2 + 2K^2 t \int_0^t (1 + |X_s|)^2 ds$$

$$\leq \left[ 2 \left( |X_0| + \varepsilon \sup_{0 \leq s \leq t} |L_s| \right)^2 + 4K^2 t^2 \right] + 4K^2 t \int_0^t |X_s|^2 ds.$$  

Gronwall’s inequality yields that

$$|X_t|^2 \leq \left[ 2 \left( |X_0| + \varepsilon \sup_{0 \leq s \leq t} |L_s| \right)^2 + 4K^2 t^2 \right] e^{4K^2 t^2}$$

and

$$|X_t| \leq \sqrt{2} \left( |X_0| + \varepsilon \sup_{0 \leq s \leq t} |L_s| \right) + 2Kt \right] e^{2K^2 t^2}.$$  

Thus, it follows that

$$|Y^t_{n, \varepsilon}| = |X_{[nt]/n}| \leq \left[ \sqrt{2} \left( |X_0| + \sup_{0 \leq s \leq t} |L_s| \right) + 2Kt \right] e^{2K^2 t^2},$$

which is almost surely finite. Therefore the proof is complete. \hfill $\Box$

We shall use $\nabla f(x, \theta) = (\partial_x f(x, \theta), \ldots, \partial_\theta f(x, \theta))^T$ to denote the gradient operator of $f(x, \theta)$ with respect to $x$.

**Lemma 3.3.** Let $f \in C^{1,1}_x(\mathbb{R}^d \times \Theta; \mathbb{R}).$ Assume (A1)–(A2). Then, we have

$$\frac{1}{n} \sum_{k=1}^n f(X_{n-1}, \theta) \rightharpoonup \int_0^1 f(X_0, \theta) ds$$

as $\varepsilon \to 0$ and $n \to \infty$, uniformly in $\theta \in \Theta.$
Proof. By the differentiability of the function $f(x, \theta)$ and Lemma 3.1, we find that

$$\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{k=1}^{n} f(X_{i_k-1}, \theta) - \int_0^1 f(X^0_i, \theta) \, ds \right| \leq \sup_{\theta \in \Theta} \left| \int_0^1 f(Y^{n,e}_s, \theta) \, ds - \int_0^1 f(X^0_i, \theta) \, ds \right|$$

$$\leq \sup_{\theta \in \Theta} \int_0^1 \left| f(Y^{n,e}_s, \theta) - f(X^0_i, \theta) \right| \, ds$$

$$\leq \sup_{\theta \in \Theta} \int_0^1 \left| \int_0^1 (\nabla f)(X^0_i + u(Y^{n,e}_s - X^0_i), \theta) \cdot (Y^{n,e}_s - X^0_i) \, du \right| \, ds$$

$$\leq \int_0^1 \left( \int_0^1 \sup_{\theta \in \Theta} |\nabla f(X^0_i + u(Y^{n,e}_s - X^0_i), \theta)| \, du \right) |Y^{n,e}_s - X^0_i| \, ds$$

$$\leq C \left( 1 + \sup_{0 \leq s \leq 1} |X^0_i| + \sup_{0 \leq s \leq 1} |X_i| \right)^\lambda \sup_{0 \leq s \leq 1} |Y^{n,e}_s - X^0_i|$$

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

Lemma 3.4. Let $f \in C^1_1(R^d \times \Theta; R)$. Assume (A1)–(A2). Then, we have that for each $1 \leq i \leq d$ and each $\theta \in \Theta$,

$$\sum_{k=1}^{n} f(X_{i_k-1}, \theta)(l^i_{k} - l^i_{i_k-1}) \overset{p_{\Theta}}{\to} \int_0^1 f(X^0_i, \theta) \, dt^i_l$$

as $\varepsilon \to 0$ and $n \to \infty$, where

$$l^i_{i} = a_t + \sum_{j=1}^{r} \alpha_j B^j_t + \int_0^t \int_{|z| \leq 1} z_i \tilde{N}(ds, dz) + \int_0^t \int_{|z| > 1} z_i \bar{N}(ds, dz)$$

is the $i$-th component of $L_t$.

Proof. Note that

$$\sum_{k=1}^{n} f(X_{i_k-1}, \theta)(l^i_{k} - l^i_{i_k-1}) = \int_0^1 f(Y^{n,e}_s, \theta) \, dt^i_l.$$ 

Let $\tilde{L}^i = L^i_t - \int_0^t \int_{|z| > 1} z_i \bar{N}(ds, dz)$. Then, we have the following decomposition

$$\int_0^1 f(Y^{n,e}_s, \theta) \, dt^i_l - \int_0^1 f(X^0_i, \theta) \, dt^i_l = \int_0^1 \int_{|z| > 1} (f(Y^{n,e}_s, \theta) - f(X^0_i, \theta))z_i \bar{N}(ds, dz) + \int_0^1 (f(Y^{n,e}_s, \theta) - f(X^0_i, \theta)) \, d\tilde{L}^i.$$ 

Similar to the proof of Lemma 3.3, we have

$$\left| \int_0^1 \int_{|z| > 1} (f(Y^{n,e}_s, \theta) - f(X^0_i, \theta))z_i \bar{N}(ds, dz) \right| \leq C \int_0^1 \int_{|z| > 1} |Y^{n,e}_s - X^0_i| |z_i| \bar{N}(ds, dz)$$

$$\leq C \left( 1 + \sup_{0 \leq s \leq 1} |X^0_i| + \sup_{0 \leq s \leq 1} |X_i| \right)^\lambda \sup_{0 \leq s \leq 1} |Y^{n,e}_s - X^0_i|$$

$$\times \int_0^1 \int_{|z| > 1} |z_i| \bar{N}(ds, dz),$$
which converges to zero in probability as $\varepsilon \to 0$ and $n \to \infty$ by Lemma 3.1. By using the stopping time $\tau_{m}^{n,\varepsilon}$, Lemma 3.1, Markov inequality and dominated convergence, we find that for any given $\eta > 0$ and some fixed $m$

\[
P\left(\left|\int_{0}^{1} (f(Y_{s}^{n,\varepsilon}, \theta) - f(X_{s}^{0}, \theta))1_{|s| \leq m} \, d\tilde{L}_{s}\right| > \eta\right) \leq \frac{|a_{i}|}{\eta} \int_{0}^{1} \mathbb{E}\left[|f(Y_{s}^{n,\varepsilon}, \theta) - f(X_{s}^{0}, \theta)|1_{|s| \leq m} \right] \, ds
\]

\[
+ \frac{\sqrt{\sum_{j=1}^{2} \sigma_{j}^{2}}}{\eta} \left(\int_{0}^{1} \mathbb{E}\left[|f(Y_{s}^{n,\varepsilon}, \theta) - f(X_{s}^{0}, \theta)|21_{|s| \leq m} \right] \, ds\right)^{1/2}
\]

\[
+ \frac{1}{\eta} \left(\int_{0}^{1} \mathbb{E}\left[|f(Y_{s}^{n,\varepsilon}, \theta) - f(X_{s}^{0}, \theta)|21_{|s| \leq m} \right] \, ds \int_{|z| \leq 1} |z|^{2} \nu(dz)\right)^{1/2},
\]

which goes to zero as $\varepsilon \to 0$ and $n \to \infty$. Then, we have

\[
P\left(\left|\int_{0}^{1} (f(Y_{s}^{n,\varepsilon}, \theta) - f(X_{s}^{0}, \theta)) \, d\tilde{L}_{s}\right| > \eta\right) \leq P(\tau_{m}^{n,\varepsilon} < 1) + P\left(\left|\int_{0}^{1} (f(Y_{s}^{n,\varepsilon}, \theta) - f(X_{s}^{0}, \theta))1_{|s| \leq m} \, d\tilde{L}_{s}\right| > \eta\right),
\]

which converges to zero as $\varepsilon \to 0$ and $n \to \infty$ by Lemma 3.2 and (3.2). This completes the proof.}

**Lemma 3.5.** Let $f \in C_{c}^{1,1}(\mathbb{R}^{d} \times \Theta; \mathbb{R})$. Assume (A1)–(A2). Then, we have that for $1 \leq i \leq d$,

\[
\sum_{k=1}^{n} f(X_{k-1}, \theta)(X_{k}^{i} - X_{k-1}^{i} - b_{i}(X_{k-1}, \theta_{0}) \Delta t_{k-1}) = \frac{p_{0}}{n}
\]

as $\varepsilon \to 0$ and $n \to \infty$, uniformly in $\theta \in \Theta$, where $X_{i}^{i}$ and $b_{i}$ are the $i$-th components of $X_{i}$ and $b$, respectively.

**Proof.** Note that

\[
X_{k}^{i} = X_{k-1}^{i} + \int_{t_{k-1}}^{t_{k}} b_{i}(X_{k-1}, \theta_{0}) \, ds + \varepsilon(L_{k}^{i} - L_{k-1}^{i}).
\]

It is easy to see that

\[
\sum_{k=1}^{n} f(X_{k-1}, \theta)(X_{k}^{i} - X_{k-1}^{i} - b_{i}(X_{k-1}, \theta_{0}) \Delta t_{k-1}) = \sum_{k=1}^{n} \int_{t_{k-1}}^{t_{k}} f(X_{k-1}, \theta)(b_{i}(X_{k-1}, \theta_{0}) - b_{i}(X_{k-1}, \theta_{0})) \, ds
\]

\[
+ \varepsilon \sum_{k=1}^{n} f(X_{k-1}, \theta)(L_{k}^{i} - L_{k-1}^{i})
\]

\[
= \int_{0}^{1} f(Y_{s}^{n,\varepsilon}, \theta)(b_{i}(X_{s}, \theta_{0}) - b_{i}(Y_{s}^{n,\varepsilon}, \theta_{0})) \, ds + \varepsilon \int_{0}^{1} f(Y_{s}^{n,\varepsilon}, \theta) \, dL_{s}^{i}.
\]

By the given condition on $f$ and the Lipschitz condition on $b$, we have

\[
\sup_{\theta \in \Theta} \left|\int_{0}^{1} f(Y_{s}^{n,\varepsilon}, \theta)(b_{i}(X_{s}, \theta_{0}) - b_{i}(Y_{s}^{n,\varepsilon}, \theta_{0})) \, ds\right| \leq \int_{0}^{1} \sup_{\theta \in \Theta} |f(Y_{s}^{n,\varepsilon}, \theta)| \cdot K|X_{s} - Y_{s}^{n,\varepsilon}| \, ds
\]

\[
\leq KC \int_{0}^{1} (1 + |Y_{s}^{n,\varepsilon}|)^{\lambda}(|X_{s} - X_{0}^{0}| + |Y_{s}^{n,\varepsilon} - X_{0}^{0}|) \, ds
\]

\[
\leq KC \left(1 + \sup_{0 \leq s \leq 1} |X_{s}| \right)^{\lambda} \left(\sup_{0 \leq s \leq 1} |X_{s} - X_{0}^{0}| + \sup_{0 \leq s \leq 1} |Y_{s}^{n,\varepsilon} - X_{0}^{0}| \right),
\]

which converges to zero in probability as $\varepsilon \to 0$ and $n \to \infty$ by Lemma 3.1. Next using the decomposition of $L_{s}$, we have

\[
\sup_{\theta \in \Theta} \varepsilon \int_{0}^{1} f(Y_{s}^{n,\varepsilon}, \theta) \, dL_{s}^{i} \leq \sup_{\theta \in \Theta} a_{i} \int_{0}^{1} f(Y_{s}^{n,\varepsilon}, \theta) \, ds + \varepsilon \sup_{\theta \in \Theta} \int_{0}^{1} f(Y_{s}^{n,\varepsilon}, \theta) \sum_{j=1}^{2} \sigma_{j} \, dB_{j} \]

\[
+ \varepsilon \sup_{\theta \in \Theta} \int_{0}^{1} \int_{|z| \leq 1} f(Y_{s}^{n,\varepsilon}, \theta)z_{i} \tilde{N}(ds, \, dz) \leq \sup_{\theta \in \Theta} \int_{0}^{1} f(Y_{s}^{n,\varepsilon}, \theta)z_{i} \tilde{N}(ds, \, dz) \leq \sup_{\theta \in \Theta} \int_{0}^{1} f(Y_{s}^{n,\varepsilon}, \theta)z_{i} \tilde{N}(ds, \, dz).
\]
It is clear that
\[ \varepsilon \sup_{\theta \in \Theta} \left| a_i \int_0^1 f(Y_{s+}^{n, \varepsilon}, \theta)ds \right| \leq \varepsilon |a_i| C \int_0^1 (1 + |Y_{s+}^{n, \varepsilon}|)^\lambda ds \]
\[ \leq \varepsilon |a_i| C \left( 1 + \sup_{0 \leq i \leq 1} |X_i| \right)^\lambda, \]
which converges to zero in probability as \( \varepsilon \to 0 \) and \( n \to \infty \), and
\[ \varepsilon \sup_{\theta \in \Theta} \left| \int_0^1 \int_{|z| > 1} f(Y_{s+}^{n, \varepsilon}, \theta)zN(ds, dz) \right| \leq \varepsilon \int_0^1 \int_{|z| > 1} \sup_{\theta \in \Theta} |f(Y_{s+}^{n, \varepsilon}, \theta)| \cdot |z|N(ds, dz) \]
\[ \leq \varepsilon \int_0^1 \int_{|z| > 1} C(1 + |Y_{s+}^{n, \varepsilon}|)^\lambda \cdot |z|N(ds, dz) \]
\[ \leq \varepsilon C \left( 1 + \sup_{0 \leq i \leq 1} |X_i| \right) \int_0^1 \int_{|z| > 1} |z|N(ds, dz), \]
which converges to zero in probability. Note that
\[ P \left( \varepsilon \sup_{\theta \in \Theta} \left| \int_0^1 f(Y_{s+}^{n, \varepsilon}, \theta) \sum_{j=1}^{r} \sigma_{ij} dB_s^i \right| > \eta \right) \leq P\left( \varepsilon \sup_{\theta \in \Theta} \left| \int_0^1 f(Y_{s+}^{n, \varepsilon}, \theta) 1_{|s| \leq \tau_n, \varepsilon} \sum_{j=1}^{r} \sigma_{ij} dB_s^i \right| > \eta \right) \]
(3.3)

Let
\[ u_{n, \varepsilon}^i(\theta) = \varepsilon \int_0^1 f(Y_{s+}^{n, \varepsilon}, \theta) 1_{|s| \leq \tau_n, \varepsilon} \sum_{j=1}^{r} \sigma_{ij} dB_s^i, \quad 1 \leq i \leq d. \]

We want to prove that \( u_{n, \varepsilon}^i(\theta) \to 0 \) in probability as \( \varepsilon \to 0 \) and \( n \to \infty \), uniformly in \( \theta \in \Theta \). It suffices to show the pointwise convergence and the tightness of the sequence \( \{u_{n, \varepsilon}^i(\cdot)\} \). For the pointwise convergence, by the Chebyshev inequality and Itô's isometry, we have
\[ P(|u_{n, \varepsilon}^i(\theta)| > \eta) \leq \varepsilon^2 \eta^{-2} \mathbb{E} \left[ \int_0^1 |f(Y_{s+}^{n, \varepsilon}, \theta) 1_{|s| \leq \tau_n, \varepsilon} \sum_{j=1}^{r} \sigma_{ij} dB_s^i|^2 \right] \]
\[ \leq \left( \sum_{j=1}^{r} \sigma_{ij}^2 \right) \varepsilon^2 \eta^{-2} \int_0^1 \mathbb{E} \left[ |f(Y_{s+}^{n, \varepsilon}, \theta) 1_{|s| \leq \tau_n, \varepsilon}|^2 \right] ds \]
\[ \leq \left( \sum_{j=1}^{r} \sigma_{ij}^2 \right) \varepsilon^2 \eta^{-2} C^2 (1 + |Y_{s+}^{n, \varepsilon}|)^{2\lambda} 1_{|s| \leq \tau_n, \varepsilon} ds \]
\[ \leq \left( \sum_{j=1}^{r} \sigma_{ij}^2 \right) \varepsilon^2 \eta^{-2} C^2 (1 + m)^{2\lambda}, \quad (3.4) \]
which converges to zero as \( \varepsilon \to 0 \) and \( n \to \infty \) with fixed \( m \). For the tightness of \( \{u_{n, \varepsilon}^i(\cdot)\} \), by using Theorem 20 in Appendix I of Ibragimov and Has'minskii [11], it is enough to prove the following two inequalities
\[ \mathbb{E}[|u_{n, \varepsilon}^i(\theta)|^{2q}] \leq C, \quad (3.5) \]
\[ \mathbb{E}[|u_{n, \varepsilon}^i(\theta_2) - u_{n, \varepsilon}^i(\theta_1)|^{2q}] \leq C|\theta_2 - \theta_1|^{2q} \quad (3.6) \]
for \( \theta, \theta_1, \theta_2 \in \Theta \), where \( 2q > p \). The proof of (3.5) is very similar to moment estimates in (3.4) by replacing Itô’s isometry with the Burkholder–Davis–Gundy inequality. So we omit the details here. For (3.6), by using Taylor’s formula and the Burkholder–Davis–Gundy inequality, we have
\[ \mathbb{E}[|u_{n, \varepsilon}^i(\theta_2) - u_{n, \varepsilon}^i(\theta_1)|^{2q}] \leq \varepsilon^{2q} C_q \left( \sum_{j=1}^{r} \sigma_{ij}^2 \right)^q \mathbb{E} \left[ \left( \int_0^1 |f(Y_{s+}^{n, \varepsilon}, \theta_2) - f(Y_{s+}^{n, \varepsilon}, \theta_1)|^2 1_{|s| \leq \tau_n, \varepsilon} ds \right)^q \right] \]
\[ \leq \varepsilon^{2q} C_q \left( \sum_{j=1}^{r} \sigma_{ij}^2 \right)^q \mathbb{E} \left[ \left( \int_0^1 |\theta_2 - \theta_1|^2 |\nabla f(Y_{s+}^{n, \varepsilon}, \theta_1 + v(\theta_2 - \theta_1))|^2 1_{|s| \leq \tau_n, \varepsilon} dvds \right)^q \right] \]
Combining (3.3) and the above arguments, we have that \( \varepsilon \sup_{\theta \in \Theta} \left| \int_0^1 f(Y_{n\varepsilon}^n, \theta) \sum_{j=1}^r \sigma_j dB^j \right| \) converges to zero in probability as \( \varepsilon \to 0 \) and \( n \to \infty \). Similarly, we can prove that \( \varepsilon \sup_{\theta \in \Theta} \left| \int_0^1 \int_{|z| \leq 1} f(Y_{n\varepsilon}^n, \theta) z_0 N(ds, dz) \right| \) converges to zero in probability as \( \varepsilon \to 0 \) and \( n \to \infty \). Therefore, the proof is complete. \( \square \)

**Proof of Theorem 2.1.** Note that

\[
\Phi_{n,\varepsilon}(\theta) = \sum_{k=1}^n \left( b(X_{k-1}, \theta) - b(X_{k-1}, \theta_0) \right)^T (X_k - X_{k-1} - n^{-1} b(X_{k-1}, \theta_0)) + \frac{1}{n} \sum_{k=1}^n \left| b(X_{k-1}, \theta) - b(X_{k-1}, \theta_0) \right|^2
\]

By Lemma 3.5 and let \( f(x, \theta) = b(x, \theta) - b(x, \theta_0) \) \( (1 \leq i \leq d) \), we have \( \sup_{\theta \in \Theta} |\Phi_{n,\varepsilon}^{(1)}(\theta)| \xrightarrow{P_{\theta_0}} 0 \) as \( \varepsilon \to 0 \) and \( n \to \infty \). By using Lemma 3.3 with \( f(x, \theta) = |b(x, \theta) - b(x, \theta_0)|^2 \), we find \( \sup_{\theta \in \Theta} |\Phi_{n,\varepsilon}^{(2)}(\theta) - F(\theta)| \xrightarrow{P_{\theta_0}} 0 \) as \( \varepsilon \to 0 \) and \( n \to \infty \), where

\[
F(\theta) = \int_0^1 |b(X_0^0, \theta) - b(X_0^0, \theta_0)|^2 dt.
\]

Thus combining the previous arguments, we have

\[
\sup_{\theta \in \Theta} |\Phi_{n,\varepsilon}(\theta) - F(\theta)| \xrightarrow{P_{\theta_0}} 0
\]

as \( \varepsilon \to 0 \) and \( n \to \infty \), and that (A3) and the continuity of \( X_0^0 \) 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 [47], we have the desired consistency, i.e., \( \hat{\theta}_{n,\varepsilon} \xrightarrow{P_{\theta_0}} \theta_0 \) as \( \varepsilon \to 0 \) and \( n \to \infty \). This completes the proof. \( \square \)

3.2. **Proof of Theorem 2.2**

Note that

\[

\nabla_\theta \Phi_{n,\varepsilon}(\theta) = -2 \sum_{k=1}^n (\nabla_\theta b)^T (X_{k-1}, \theta) (X_k - X_{k-1} - b(X_{k-1}, \theta) \Delta t_{k-1}).
\]

Let \( G_{n,\varepsilon}(\theta) = (G_{n,\varepsilon}^1, \ldots, G_{n,\varepsilon}^p)^T \) with

\[

G_{n,\varepsilon}^i(\theta) = \sum_{k=1}^n (\partial_{\theta_i} b)^T (X_{k-1}, \theta) (X_k - X_{k-1} - b(X_{k-1}, \theta) \Delta t_{k-1}), \quad i = 1, \ldots, p,
\]

and let \( K_{n,\varepsilon}(\theta) = \nabla_\theta G_{n,\varepsilon}(\theta) \), which is a \( p \times p \) matrix consisting of elements \( k_{n,\varepsilon}^{ij}(\theta) = \partial_{\theta_j} G_{n,\varepsilon}^i(\theta), \quad 1 \leq i, j \leq p \). Moreover, we introduce the following function

\[
K(\theta) = \int_0^1 (\partial_{\theta_j} b)^T (X_0^0, \theta) (b(X_0^0, \theta) - b(X_0^0, \theta)) ds - l^j(\theta), \quad 1 \leq i, j \leq p.
\]

Then we define the matrix function \( K(\theta) = (K(\theta))_{i,j} \), \( 1 \leq i, j \leq p \).

**Lemma 3.6.** Assume (A1)–(A2). Then, we have that for each \( i = 1, \ldots, p \)

\[
\varepsilon^{-1} G_{n,\varepsilon}^i(\theta_0) \xrightarrow{P_{\theta_0}} \int_0^1 (\partial_{\theta_j} b)^T (X_0^0, \theta_0) dL_s
\]

as \( \varepsilon \to 0, n \to \infty \) and \( n\varepsilon \to \infty \).
Thus, by the Lipschitz condition on \( b \) and the Cauchy–Schwarz inequality, we find that

\[
|X_{t_{k-1}} - X_{t_{k-1}}|^2 \leq 2 \int_{t_{k-1}}^{s} \left( b(X_{u}, \theta_0) - b(X_{t_{k-1}}, \theta_0) \right)^2 \, du + 2 \left( |b(X_{t_{k-1}}, \theta_0)| (s - t_{k-1}) + \varepsilon |L_s - L_{t_{k-1}}| \right)^2
\]

By Gronwall’s inequality, we get

\[
|X_s - X_{t_{k-1}}|^2 \leq 2 \left( n^{-1} |b(X_{t_{k-1}}, \theta_0)| + \varepsilon \sup_{t_{k-1} \leq s \leq t_k} |L_s - L_{t_{k-1}}| \right)^2. 
\]

It further follows that

\[
\sup_{t_{k-1} \leq s \leq t_k} |X_s - X_{t_{k-1}}| \leq \sqrt{2} \left( n^{-1} |b(X_{t_{k-1}}, \theta_0)| + \varepsilon \sup_{t_{k-1} \leq s \leq t_k} |L_s - L_{t_{k-1}}| \right) e^{2K^2 n^{-1} (s - t_{k-1})}. 
\]

Thus, by the Lipschitz condition on \( b \) and (3.7), we get

\[
|H_{n,e}^{(1)}(\theta_0)| \leq e^{-1} \sum_{k=1}^{n} |\partial_{\theta} b(X_{t_{k-1}}, \theta_0) : \int_{t_{k-1}}^{s} \left( b(X_{u}, \theta_0) - b(X_{t_{k-1}}, \theta_0) \right) \, du|
\]

By Lemma 3.4 and letting \( f(x, \theta) = \partial_{\theta} b_j(x, \theta) \) \( (1 \leq i \leq p, 1 \leq j \leq d) \) with \( \theta = \theta_0 \), we have

\[
H_{n,e}^{(2)}(\theta_0) = \int_{0}^{1} (\partial_{\theta} b) (Y_{n,e}, \theta_0) \, dL_\varepsilon \int_{0}^{1} (\partial_{\theta} b) (X_0, \theta_0) \, dL_\varepsilon
\]

\[
\xrightarrow[\varepsilon \to 0]{} 0 \quad \text{and} \quad \frac{1}{n} \sum_{k=1}^{n} |\partial_{\theta} b(X_{t_{k-1}}, \theta_0) \cdot b(X_{t_{k-1}}, \theta_0)| \leq CK \left( 1 + \sup_{0 \leq s \leq 1} |X_s| \right)^{\lambda + 1} \quad \text{a.s.}
\]
(cf. (3.1)). By using the basic fact that

$$\frac{1}{n} \sum_{k=1}^{n} \sup_{0 \leq i \leq 1} |L_k - L_{k-1}| = o_p(1),$$

we find that

$$H_{n,\epsilon}^{(1,2)}(\theta_0) \leq \sqrt{2Ke^{K^2/n^2} C \left(1 + \sup_{0 \leq i \leq 1} |U_i|\right)} \frac{1}{n} \sum_{k=1}^{n} \sup_{0 \leq i \leq 1} |L_k - L_{k-1}|,$$

which converges to zero in probability as $\epsilon \to 0$ and $n \to \infty$. Therefore the proof is complete. □

**Lemma 3.7.** Assume (A1)–(A4). Then, we have

$$\sup_{\theta \in \Theta} |K_{n,\epsilon}(\theta) - K(\theta)| \xrightarrow{p_{\theta_0}} 0$$

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

**Proof.** It suffices to prove that for $1 \leq i, j \leq p$

$$\sup_{\theta \in \Theta} |K_{n,\epsilon}^{ij}(\theta) - K^{ij}(\theta)| \xrightarrow{p_{\theta_0}} 0$$

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

$$K_{n,\epsilon}^{ij}(\theta) = \frac{\partial_{ij} C_{n,\epsilon}(\theta)}{\sup_{D(\theta_0)} |\hat{\theta}_{n,\epsilon} - \theta_0|} \sup_{D(\theta_0)} |\hat{\theta}_{n,\epsilon} - \theta_0|$$

By using Lemma 3.5 and letting $f(x, \theta) = \partial_{ij} b(x, \theta)$ ($1 \leq i, j \leq p, 1 \leq l \leq d$), we have that $\sup_{\theta \in \Theta} |K_{n,\epsilon}^{ij}(\theta)|$ converges to zero in probability as $\epsilon \to 0$ and $n \to \infty$. By using Lemma 3.3 and letting $f(x, \theta) = (\partial_{ij} b)^T(x, \theta)(b(x, \theta_0) - b(x, \theta)) - (\partial_{ij} b)^T(x, \theta)\partial_{ij} b(x, \theta)$, it follows that $\sup_{\theta \in \Theta} |K_{n,\epsilon}^{ij}(\theta)|$ converges to zero in probability as $\epsilon \to 0$ and $n \to \infty$. Thus, the proof is complete. □

Finally we are ready to prove Theorem 2.2.

**Proof of Theorem 2.2.** The proof ideas mainly follow Uchida [42]. Let $B(\theta_0; \rho) = \{\theta : |\theta - \theta_0| \leq \rho\}$ for $\rho > 0$. Then, by the consistency of $\hat{\theta}_{n,\epsilon}$, there exists a sequence $\eta_{n,\epsilon} \to 0$ as $\epsilon \to 0$ and $n \to \infty$ such that $B(\theta_0; \eta_{n,\epsilon}) \subset \Theta_0$, and that $P_{\theta_0}[\hat{\theta}_{n,\epsilon} \in B(\theta_0; \eta_{n,\epsilon})] \to 1$. When $\hat{\theta}_{n,\epsilon} \in B(\theta_0; \eta_{n,\epsilon})$, it follows by Taylor’s formula that

$$D_{n,\epsilon}S_{n,\epsilon} = \epsilon^{-1}G_{n,\epsilon}(\theta_0) - e^{-1}G_{n,\epsilon}(\theta_0),$$

where $D_{n,\epsilon} = \int_{0}^{1} K_{\epsilon,\epsilon}(\theta_0 + t(\hat{\theta}_{n,\epsilon} - \theta_0)) dt$ and $S_{n,\epsilon} = \epsilon^{-1}(\hat{\theta}_{n,\epsilon} - \theta_0)$ since $B(\theta_0; \eta_{n,\epsilon})$ is a convex subset of $\Theta_0$. We have

$$|D_{n,\epsilon} - K_{n,\epsilon}(\theta_0)| \leq \sup_{\theta \in \Theta_0} |K_{n,\epsilon}(\theta) - K_{n,\epsilon}(\theta_0)|$$

Consequently, it follows from Lemma 3.7 that

$$D_{n,\epsilon} \xrightarrow{p_{\theta_0}} K(\theta_0), \quad \epsilon \to 0, \quad n \to \infty.$$
Then, for any \( \varepsilon \in (0, \varepsilon(\delta)) \) and \( n > N(\delta) \), we have, on \( \Gamma_{n,\varepsilon} \),
\[
\sup_{|u|=1} |(D_{n,\varepsilon} - K(\theta_0)) w| \leq \sup_{|u|=1} \left| D_{n,\varepsilon} - \int_0^1 K(\theta_0 + u(\hat{\theta}_{n,\varepsilon} - \theta_0)) \, du \right| w
\]
\[
\quad + \sup_{|u|=1} \left| \int_0^1 K(\theta_0 + u(\hat{\theta}_{n,\varepsilon} - \theta_0)) \, du - K(\theta_0) \right| w
\]
\[
\quad \leq \sup_{|\theta - \theta_0| \leq \varepsilon n} |K_n(\theta) - K(\theta)| + \frac{\delta}{2} < \delta.
\]

Thus, on \( \Gamma_{n,\varepsilon} \),
\[
\inf_{|u|=1} |D_{n,\varepsilon} w| \geq \inf_{|u|=1} |K(\theta_0)| - \sup_{|u|=1} |(D_{n,\varepsilon} - K(\theta_0)) w| > 2\delta - \delta = \delta > 0.
\]

Hence, letting
\[
\delta_{n,\varepsilon} = \{ D_{n,\varepsilon} \text{ is invertible, } \hat{\theta}_{n,\varepsilon} \in B(\theta_0; \eta_{n,\varepsilon}) \},
\]
we see that \( P_{\theta_0}[\delta_{n,\varepsilon}] \geq P_{\theta_0}[\Gamma_{n,\varepsilon}] \to 1 \) as \( \varepsilon \to 0 \) and \( n \to \infty \) by Lemma 3.7. Now set
\[
U_{n,\varepsilon} = D_{n,\varepsilon}^{-1} 1_{\delta_{n,\varepsilon}} + I_{p \times p} 1_{\delta_{n,\varepsilon}}^c,
\]
where \( I_{p \times p} \) is the identity matrix. Then it is easy to see that
\[
|U_{n,\varepsilon} - K(\theta_0)| \leq |D_{n,\varepsilon} - K(\theta_0)| 1_{\delta_{n,\varepsilon}} + |I_{p \times p} - K(\theta_0)| 1_{\delta_{n,\varepsilon}}^c \to 0,
\]
since \( P_{\theta_0}[\delta_{n,\varepsilon}] \to 1 \). Thus, by Lemma 3.6, we obtain that
\[
S_{n,\varepsilon} = U_{n,\varepsilon}^{-1} D_{n,\varepsilon} S_{\theta_0} 1_{\delta_{n,\varepsilon}} + S_{n,\varepsilon} 1_{\delta_{n,\varepsilon}}^c
\]
\[
= U_{n,\varepsilon}^{-1} (-\varepsilon^{-1} G_{n,\varepsilon}(\theta_0)) 1_{\delta_{n,\varepsilon}} + S_{n,\varepsilon} 1_{\delta_{n,\varepsilon}}^c
\]
\[
\to (I(\theta_0))^{-1} \left( \int_0^1 (\partial_{\theta_0} b)^T (X_0^0, \theta_0) \, dL, \ldots, \int_0^1 (\partial_{\theta_0} b)^T (X_0^0, \theta_0) \, dL \right)^T
\]
as \( \varepsilon \to 0 \), \( n \to \infty \) and \( n\varepsilon \to \infty \). This completes the proof. \( \square \)

4. Generalization to semi-martingale noises

In this section, we discuss the extension of our main results in Section 2 to the general case when the driving noise is a semi-martingale. Let \( Q_t = Q_{t0} + M_t + A_t \) be a semi-martingale, where \( M_t \) is a local martingale and \( A_t \) is a finite variation process. Then, we can replace the driving Lévy process \( L_t \) in the SDE (1.2) by the semi-martingale \( Q_t \) to get
\[
dX_t = b(X_t, \theta) \, dt + \varepsilon \, dQ_t, \quad t \in [0, 1], \quad X_0 = x_0.
\]
(4.1)
All the related information about the LSE of \( \theta \) discussed in the Introduction and Section 2 shall be the same. We are interested in the consistency and asymptotic behavior of the LSE \( \hat{\theta}_{n,\varepsilon} \) under the general model (4.1).

We state the new results as follows.

**Theorem 4.1.** Under conditions (A1)–(A3), we have
\[
\hat{\theta}_{n,\varepsilon} \overset{P_{\theta_0}}{\to} \theta_0
\]
as \( \varepsilon \to 0 \) and \( n \to \infty \).

**Theorem 4.2.** Under conditions (A1)–(A4), we have
\[
\varepsilon^{-1}(\hat{\theta}_{n,\varepsilon} - \theta_0) \overset{P_{\theta_0}}{\to} I^{-1}(\theta_0)\hat{S}(\theta_0),
\]
as \( \varepsilon \to 0 \), \( n \to \infty \) and \( n\varepsilon \to \infty \), where
\[
\hat{S}(\theta_0) := \left( \int_0^1 (\partial_{\theta_0} b)^T (X_0^0, \theta_0) \, dQ, \ldots, \int_0^1 (\partial_{\theta_0} b)^T (X_0^0, \theta_0) \, dQ \right)^T.
\]
**Remark 4.3.** Since the proofs of Theorems 4.1 and 4.2 are similar to those of Theorems 2.1 and 2.2 given in Section 3, we mention only the necessary modifications. Note that $M_n$ is a local martingale, we can use the standard localization procedure to make $M$ and $[M, M]$ being bounded up to the stopping times $\{T_m\}$, $m = 1, 2, \ldots$, with $\lim_{m \to \infty} T_m = \infty$ almost surely. For example, we can define $T_m = \inf\{t > 0 : [M, M] > m\}$. Then, we can modify the definition of $r_m^{n, \tau}$ by

$$r_m^{n, \tau} = \inf\{t \geq 0 : |\hat{X}^0_t| \geq m \text{ or } |\hat{Y}^{1, \tau}_t| \geq m\} \wedge T_m.$$ 

Lemma 3.2 still holds, i.e. $r_m^{n, \tau} \to \infty$ a.s. uniformly in $n$ and $\varepsilon$ as $m \to \infty$. When the proofs are based on pathwise arguments, they can be carried over to the semi-martingale noise case easily. When the proofs are based on the Markov inequality (or the Chebyshov inequality), Itô’s isometry and the Burkholder–Davis–Gundy inequality (cf. Theorem 54 and remark on page 175 in Chapter IV of Protter [32]), we can apply the modified stopping times $r_m^{n, \tau}$ to the stochastic integrals with respect to the local martingale $M_n$. Thus all the proofs will still be valid in terms of the modifications described as above. We omit the details here.

5. Simulations

Consider a 2-dimensional model for (1.2) with

$$b(x, \theta) = \left(\sqrt{\theta_1 + x_1^2 + x_2^2}, -\frac{\theta_2 x_2}{\sqrt{1 + x_1^2 + x_2^2}}\right)^T, \quad L_t = \left(\begin{array}{c} V_t^{\beta, \gamma} + B_t \\
S_t \end{array}\right),$$

(5.1)

where $B$ is the standard Brownian motion, $S^\alpha$ is a standard symmetric $\alpha$-stable process $S_\alpha(1, 0, 0)$, and $V^{\beta, \gamma}$ is a variance gamma process with Lévy density

$$p_V(z) = \frac{\delta}{|z|^\gamma} e^{-\gamma |z|}, \quad z \in \mathbb{R}, \quad \delta, \gamma > 0.$$ 

In this example, we find that our LSE of $\theta$, say $\hat{\theta} = (\hat{\theta}_{n,r,1}, \hat{\theta}_{n,r,2})$, satisfies

$$\sum_{k=1}^n \frac{X_{tk}^k - X_{tk-1}^k}{\sqrt{\hat{\theta}_{n,r,1} + (X_{tk-1}^k)^2 + (X_{tk-1}^k)^2}} = 1; \quad \hat{\theta}_{n,r,2} = -\frac{\sum_{k=1}^n \frac{(X_{tk}^2 - X_{tk-1}^2)}{1 + (X_{tk-1}^1)^2 + (X_{tk-1}^2)^2}}{n - 1}. \quad (5.2)$$

In the sequel, we set values of parameters as

$$(X_{t0}^0, X_{t0}^2) = (1, 1), \quad (\theta_1, \theta_2) = (2, 1), \quad (\delta, \gamma, \alpha) = (5, 3, 3/2).$$

Then both $X^1$ and $X^2$ are infinite activity jump-processes, but the jump activity of $X^1$ is bounded variation, and the one of $X^2$ is unbounded variation. A sample path of $X = (X^1_t, X^2_t)_{t \in [0, 1]}$ with $\varepsilon = 0.3$ is given in Fig. 1.

In each experiment, we generate a discrete sample $(X_k)_{k=0, 1, \ldots, n}$ and compute $\hat{\theta}$ from the sample. This procedure is iterated 10,000 times, and the mean and the standard deviation of 10,000 sampled estimators are computed in each case of $\varepsilon, n$. To optimize the nonlinear equation (5.2), we used nlm function in R. On generating discrete samples of Lévy processes, see e.g., Cont and Tankov [4], Section II.6 and references therein, or one can find some random number generator in yuima package of R, which is a package for simulating SDEs with jumps; see https://r-forge.r-project.org/projects/yuima/. For example, we use rstable to generate random samples from $\alpha$-stable distributions. The results are shown in Tables 1–4. From those tables, we can observe the consistency result holds true when $\varepsilon \to 0$. We also note that the size of $n$ is often less important in practice for estimating the drift parameter than the size of $\varepsilon$ although $n \to \infty$ is necessary in theory. It is intuitively clear because the accuracy of estimating drift highly depends on the terminal time $T$ of observations in general, that is, the larger $T$ becomes, the more accurately $\hat{\theta}$ is estimated. However, the terminal $T = 1$ is now fixed. Note that, in the small noise model, letting $\varepsilon \to 0$ corresponds to observing a process from a macro point of view, which corresponds to the case $T \to \infty$ in some sense. Therefore, increasing $n$ under fixed $\varepsilon$ does not improve a bias of estimators, which is improved only if $\varepsilon \to 0$. In general, a large $n$ can decrease the standard error (or standard deviation) of estimation, but the effect seems small in this example.

Comparing standard deviations between $\hat{\theta}_{n,r,1}$ and $\hat{\theta}_{n,r,2}$, the former seems to be estimated more ‘stably’ than the latter. This is because “big” jumps of $X^1$ are less frequent than those of $X^2$. If $\varepsilon$ is small enough, the path of $X^1$ is almost similar to the deterministic curve of $X^0$ since “big” jumps do not occur so frequently. However, $X^2$ can have more “big” jumps that are not ignorable even if $\varepsilon$ is “small”, which makes the estimator fluctuating.

To observe the asymptotic distribution of $\hat{\theta}$, we shall compare the above example, say Model A (non-Gaussian noise), with a 2-dimensional process with the same drift $b$ as in (5.1), but the driving noise $L$ is 2-dimensional Brownian motion, say Model
Fig. 1. A sample path of Model (5.1) with \((\theta_1, \theta_2, \delta, \gamma, \alpha) = (2, 1, 5, 3/2)\) and \(\epsilon = 0.3\).

### Table 1
Mean (upper) and standard deviation (parentheses) of estimates through 10,000 experiments in the case \(\epsilon = 0.3\) and \((\delta, \gamma, \alpha) = (5, 3, 3/2)\).

<table>
<thead>
<tr>
<th>(\epsilon = 0.3)</th>
<th>(n = 500)</th>
<th>(n = 1000)</th>
<th>(n = 3000)</th>
<th>True</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\theta}_{n,1})</td>
<td>2.30885</td>
<td>2.31618</td>
<td>2.29381</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>(1.8770)</td>
<td>(1.8248)</td>
<td>(1.7926)</td>
<td></td>
</tr>
<tr>
<td>(\hat{\theta}_{n,2})</td>
<td>1.54087</td>
<td>1.50664</td>
<td>1.52753</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>(2.8493)</td>
<td>(2.8685)</td>
<td>(2.7667)</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2
Mean (upper) and standard deviation (parentheses) of estimates through 10,000 experiments in the case \(\epsilon = 0.1\) and \((\delta, \gamma, \alpha) = (5, 3, 3/2)\).

<table>
<thead>
<tr>
<th>(\epsilon = 0.1)</th>
<th>(n = 500)</th>
<th>(n = 1000)</th>
<th>(n = 3000)</th>
<th>True</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\theta}_{n,1})</td>
<td>2.03134</td>
<td>2.02699</td>
<td>2.02389</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>(0.5829)</td>
<td>(0.5836)</td>
<td>(0.5833)</td>
<td></td>
</tr>
<tr>
<td>(\hat{\theta}_{n,2})</td>
<td>1.10165</td>
<td>1.09839</td>
<td>1.09709</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>(1.2024)</td>
<td>(1.1212)</td>
<td>(1.0971)</td>
<td></td>
</tr>
</tbody>
</table>

### Table 3
Mean (upper) and standard deviation (parentheses) of estimates through 10,000 experiments in the case \(\epsilon = 0.05\) and \((\delta, \gamma, \alpha) = (5, 3, 3/2)\).

<table>
<thead>
<tr>
<th>(\epsilon = 0.05)</th>
<th>(n = 500)</th>
<th>(n = 1000)</th>
<th>(n = 3000)</th>
<th>True</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\theta}_{n,1})</td>
<td>2.00583</td>
<td>2.00599</td>
<td>2.01071</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>(0.2961)</td>
<td>(0.2951)</td>
<td>(0.2913)</td>
<td></td>
</tr>
<tr>
<td>(\hat{\theta}_{n,2})</td>
<td>1.04883</td>
<td>1.05963</td>
<td>1.04438</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>(1.4364)</td>
<td>(1.3026)</td>
<td>(0.6773)</td>
<td></td>
</tr>
</tbody>
</table>

B (Gaussian noise). Figs. 2 and 3 respectively show (normal) QQ-plots for 10,000 iterated samples of \(\epsilon^{-1}(\hat{\theta}_{n,t,i} - \theta_i) (i = 1, 2)\) in Model A with \((\epsilon, n) = (0.01, 3000)\), and Figs. 4 and 5 are those for Model B with \((\epsilon, n) = (0.01, 3000)\). In Model B, (marginal) asymptotic distributions of \(\epsilon^{-1}(\hat{\theta}_{n,t,i} - \theta_i) (i = 1, 2)\) must theoretically be normal, which are supported by Figs. 4 and 5. On the other hand, tails of the corresponding distributions in Model A should be heavier than normal distributions due
Table 4
Mean (upper) and standard deviation (parentheses) of estimates through 10,000 experiments in the case \( \varepsilon = 0.01 \) and \((\delta, \gamma, \alpha) = (5, 3, 3/2)\).

<table>
<thead>
<tr>
<th>( \varepsilon = 0.01 )</th>
<th>( n = 500 )</th>
<th>( n = 1000 )</th>
<th>( n = 3000 )</th>
<th>True</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\theta}_{n,1} )</td>
<td>2.00051</td>
<td>2.00061</td>
<td>2.00108</td>
<td>2.0</td>
</tr>
<tr>
<td>( \varepsilon )</td>
<td>(0.0583)</td>
<td>(0.0583)</td>
<td>(0.0578)</td>
<td></td>
</tr>
<tr>
<td>( \hat{\theta}_{n,2} )</td>
<td>1.00308</td>
<td>1.00572</td>
<td>0.99958</td>
<td>1.0</td>
</tr>
<tr>
<td>( \varepsilon )</td>
<td>(0.2626)</td>
<td>(0.1454)</td>
<td>(0.1371)</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 2. Normal QQ-plot for 10,000 iterated samples of \( \varepsilon^{-1}(\hat{\theta}_{n,1} - \theta_1) \) in Model A (non-Gaussian); \((\varepsilon, n) = (0.01, 3000)\).

Fig. 3. Normal QQ-plot for 10,000 iterated samples of \( \varepsilon^{-1}(\hat{\theta}_{n,2} - \theta_2) \) in Model A (non-Gaussian); \((\varepsilon, n) = (0.01, 3000)\).

to jump activities, and we can observe those facts from Figs. 2 and 3. We can also observe that the asymptotic distribution of \( \varepsilon^{-1}(\hat{\theta}_{n,2} - \theta_2) \) is much heavier than the one of \( \varepsilon^{-1}(\hat{\theta}_{n,1} - \theta_1) \) because of the high frequency of jumps in \( X^{(2)} \). These facts are consistent with the theory.
Fig. 4. Normal QQ-plot for 10,000 iterated samples of $e^{-1}(\hat{\theta}_{n,1} - \theta_1)$ in Model B (Gaussian); $(\epsilon, n) = (0.01, 3000)$.

Fig. 5. Normal QQ-plot for 10,000 iterated samples of $e^{-1}(\hat{\theta}_{n,2} - \theta_2)$ in Model B (Gaussian); $(\epsilon, n) = (0.01, 3000)$.

Acknowledgments

The authors are grateful to anonymous referees for suggesting to add sections for semi-martingale noise and simulations (Sections 4 and 5). This research was supported by JSPS KAKENHI Grant Number 24740061, Japan Science and Technology Agency, CREST (the 2nd author), and NSERC Grant Number 311945-2008 (the 3rd author).

References

[34] Y. Shimizu, Quadratic type contrast functions for discretely observed non-ergodic diffusion processes, Research Report Series 09-04, Division of Mathematical Science, Osaka University, 2010.