

Nadaraya-Watson estimator for stochastic processes driven by stable Lévy motions

Hongwei Long* and Lianfen Qian†

*Department of Mathematical Sciences
Florida Atlantic University
777 Glades Road, Boca Raton, FL 33431
USA*

e-mail: hlong@fau.edu; lqian@fau.edu

Abstract: We discuss the nonparametric Nadaraya-Watson (N-W) estimator of the drift function for ergodic stochastic processes driven by α -stable noises and observed at discrete instants. Under geometrical mixing condition, we derive consistency and rate of convergence of the N-W estimator of the drift function. Furthermore, we obtain a central limit theorem for stable stochastic integrals. The central limit theorem has its own interest and is the crucial tool for the proofs. A simulation study illustrates the finite sample properties of the N-W estimator.

AMS 2000 subject classifications: Primary 62G20, 62M05; secondary 60G52, 65C30.

Keywords and phrases: Central limit theorem, consistency, geometrically strong mixing, kernel estimator, Lévy motion, Nadaraya-Watson estimator, rate of convergence, stable stochastic integrals.

Received June 2012.

Contents

1	Introduction	1388
2	Asymptotic properties of the Nadaraya-Watson estimator	1390
3	A simulation study	1394
4	Proofs	1398
4.1	Preliminary Lemmas	1398
4.2	Proof of Theorem 2.3	1402
4.3	Proof of Theorem 2.4	1405
4.4	Proof of Theorem 2.5	1409
	Acknowledgements	1415
	References	1415

*Corresponding author, partially supported by FAU Start-up funding at the C. E. Schmidt College of Science.

†Specially-appointed Professor, Wenzhou University, Zhejiang, China, 325035.

1. Introduction

We consider the following nonlinear stochastic differential equation (SDE) driven by an α -stable Lévy motion ($0 < \alpha < 2$):

$$dX_t = b(X_{t-})dt + \sigma(X_{t-})dZ_t, X_0 = \eta, \quad (1.1)$$

where $b : \mathbb{R} \rightarrow \mathbb{R}$ is an unknown measurable function, $\sigma : \mathbb{R} \rightarrow \mathbb{R}_+$ is an unknown positive function (which is considered as a nuisance parameter), and $\{Z_t, t \geq 0\}$ is a standard α -stable Lévy motion defined on a probability space (Ω, \mathcal{F}, P) equipped with a right continuous and increasing family of σ -algebras $\{\mathcal{F}_t, t \geq 0\}$, and η is a random variable independent of $\{Z_t\}$. In this case, Z_1 has a α -stable distribution $S_\alpha(1, \beta, 0)$, where $\beta \in [-1, 1]$ is the skewness parameter of the distribution. When $\beta = 0$, the underlying stable distribution is symmetric. For more detailed discussion on stable distributions, we refer to Samorodnitsky and Taqqu [46], Janicki and Weron [25], and Sato [47]. We assume that the stochastic process $\{X_t\}$ is observed at discrete time points $\{t_i = i\Delta, i = 0, 1, \dots, n\}$, where Δ is the time frequency for observation and n is the sample size. The purpose of this paper is to study the nonparametric estimation of the unknown drift function b based on the sampling data $\{X_{t_i}\}_{i=0}^n$. The nonparametric estimation of the dispersion function σ is much harder, which will be addressed separately.

The SDEs driven by Lévy noises have attracted a lot of attention recently, especially in view of applications to finance (see Schoutens [49], Kyprianou, Schoutens and Wilmott [31]), network traffic (see Mikosch et al. [37]), physics (see, e.g. Schertzer et al. [48]), and climate dynamics (see Ditlevsen [10, 11]). The existence and uniqueness of solutions to (1.1) under Lipschitz conditions are standard results in stochastic calculus (see e.g., Applebaum [1]). For simplicity, we assume that the solution X_t is stationary and geometrically strong mixing (in this case, the initial distribution is taken from the invariant measure). Recently, Masuda [35] provided sets of ergodic conditions for a multidimensional diffusion process with jumps for any initial distribution to be exponential β -mixing. These conditions build up the bridge between mixing sequences and diffusion processes with jumps.

When the drift function in (1.1) is known to be linear, i.e. $b(x) = -\theta x$ with unknown parameter θ , the estimation of θ based on discrete or continuous observations of X_t was studied in the parametric framework by Hu and Long [21, 22]. If the dispersion coefficient is small and the driving noise is a Lévy process, the least squares estimation of θ was also discussed by Long [32] and Ma [34]. Masuda [36] proposed a self-weighted least absolute deviation estimator for discretely observed ergodic Ornstein-Uhlenbeck processes driven by symmetric Lévy processes. But in reality, the drift function is seldom known. Hence in this paper, we focus on estimation of the drift function $b(\cdot)$ in model (1.1) using nonparametric smoothing approach. Nonparametric smooth approach provides a versatile method of exploring the relationship between variables with no prior specified models. The classical Nadaraya-Watson (N-W) estimator of the regression function was proposed independently by Nadaraya [38] and Waston [52].

The major statistical properties (e.g. consistency and rate of convergence) of nonparametric methods for the N-W estimators under independent and identical distribution observations are developed between 1980 and 1990 (see Hardle [19]). These properties have been extended to dependent situations in the 1990s (see Bosq [6]), typically for $(\alpha, \beta$ and $\phi)$ -mixing. These results have been further generalized to stationary processes with so-called uniform predictive dependent structure, which can be regarded as a natural alternative to strong mixing conditions, by Wu [53] and [54].

Many authors have investigated nonparametric estimation for the drift function b in the setting of diffusions driven by Brownian motions. Pham [40] and Prakasa Rao [41] gave a non-parametric estimator for b by mimicking the construction of the well-known Nadaraya-Watson estimator and the asymptotic behavior of the N-W estimator was established. Arfi [2] discussed the uniformly strong consistency of the N-W estimator for the drift function b under ergodic conditions. Recently, Bandi and Phillips [3] extended the N-W estimators to non-stationary recurrent processes.

Other related methods of estimating the drift function have also been proposed. Banon [4] constructed a drift estimator purely based on the kernel estimator and a relation between the drift and the density function along with its derivative, which was further extended by van Zanten [51], Dalalyan and Kutoyants [8, 9], Dalalyan [7] (see also the monograph by Kutoyants [30] and references therein). Locally linear (or polynomial) estimators have been proposed by Fan [12] and further discussed in Fan and Gijbels [14], Spokoiny [50] and Fan and Zhang [15]. For a complete review of non-parametric methods for diffusion processes with applications in financial econometrics, see the excellent survey paper by Fan [13]. Hoffmann [20] and Gobet, Hoffmann and Reiss [17] applied wavelet approach.

In the stable setting, Hall, Peng and Yao [18] and Peng and Yao [39] discussed the non-parametric regression estimation for time series with heavy tails. In this paper, we shall consider the regression type of estimation for stochastic processes driven by Lévy motions, which is a natural extension of the discrete time series with heavy tails. We shall discuss the Nadaraya-Watson estimator of the drift function b in this paper. The basic idea of N-W estimator is to minimize an object function given below with certain weights:

$$\sum_{i=0}^{n-1} W_{n,i}(x)(Y_i - a\Delta)^2$$

over the parameter space of a and any given $x \in R$, where $Y_i := X_{t_{i+1}} - X_{t_i}$, $\Delta = t_{i+1} - t_i, i = 0, 1, \dots, n - 1$. The weight function is given by

$$W_{n,i}(x) = \frac{K_h(X_{t_i} - x)}{\sum_{i=0}^{n-1} K_h(X_{t_i} - x)}, i = 0, 1, \dots, n - 1$$

where $K_h(\cdot) = K(\cdot/h)/h$, K is a kernel density function with mean zero and finite variance, and h is the bandwidth for the kernel. Then, the N-W estimator

is given by the following expression

$$\hat{b}_n(x) = \frac{\sum_{i=0}^{n-1} Y_i K_h(X_{t_i} - x)}{\Delta \sum_{i=0}^{n-1} K_h(X_i - x)}. \quad (1.2)$$

It turns out that N-W estimator is a simple class of a large family called “local polynomial estimator” (see Fan and Gijbels [14]). Hence N-W estimator is also called *local constant estimator*. As pointed out in Gobet, Hoffmann and Reiss [17], the N-W estimator of the drift or diffusion coefficient in the classical diffusion cases is not consistent for low frequency data (i.e. Δ is fixed). So, we shall focus on the consistency and asymptotic distribution of the N-W estimator of the drift function for high frequency data (i.e. $\Delta \rightarrow 0$) in this paper.

The paper is organized as follows. In Section 2, we state our main results on the asymptotic properties of the N-W estimator $\hat{b}_n(x)$. In Section 3, we conduct a simulation study to confirm the finite sample property. Finally, all the proofs are given in Section 4. We conjecture that the results of this paper can be extended to local polynomial estimator of order p . Throughout the paper, we shall use notation “ \rightarrow_P ” to denote “convergence in probability” and notation “ \Rightarrow ” to denote “convergence in distribution”.

2. Asymptotic properties of the Nadaraya-Watson estimator

In this section, we consider asymptotic behavior of the Nadaraya-Watson estimator of the drift coefficient in our stable setting.

For convenience, we provide the definitions of α -stable distributions and α -stable Lévy motions.

Definition 2.1. A random variable ξ is said to follow a stable distribution, denoted by $\xi \sim S_\alpha(\sigma, \beta, \mu)$, if it has the characteristic function of the following form:

$$\varphi_\xi(u) = E \exp\{iu\xi\} = \begin{cases} \exp\{-\sigma^\alpha |u|^\alpha (1 - i\beta \operatorname{sgn}(u) \tan \frac{\alpha\pi}{2}) + i\mu u\}, & \text{if } \alpha \neq 1, \\ \exp\{-\sigma |u| (1 + i\beta \frac{2}{\pi} \operatorname{sgn}(u) \log |u|) + i\mu u\}, & \text{if } \alpha = 1, \end{cases}$$

where $\alpha \in (0, 2]$, $\sigma \in (0, \infty)$, $\beta \in [-1, 1]$, and $\mu \in (-\infty, \infty)$ are the index of stability, the scale, skewness, and location parameters, respectively.

If $\xi \sim S_\alpha(\sigma, \beta, \mu)$ and $\alpha \in (0, 2)$, then $\mathbb{E}|\xi|^p < \infty$ when $p \in (0, \alpha)$ and $\mathbb{E}|\xi|^p = \infty$ when $p \in [\alpha, \infty)$. When $\mu = 0$, we say ξ is *strictly α -stable*. If in addition $\beta = 0$, we call ξ *symmetric α -stable*. Note that ξ is strictly 1-stable ($\alpha = 1$) if and only if $\beta = 0$ (symmetric case). We refer to [25, 46] and [47] for more details on stable distributions.

Definition 2.2. An \mathcal{F}_t -adapted stochastic process $\{Z_t\}_{t \geq 0}$ is called a standard α -stable Lévy motion if

- (i) $Z_0 = 0$, a.s.
- (ii) $Z_t - Z_s \sim S_\alpha((t - s)^{1/\alpha}, \beta, 0)$, $t > s \geq 0$;

(iii) For any finite time points $0 \leq s_0 < s_1 < \dots < s_m < \infty$, the random variables $Z_{s_0}, Z_{s_1} - Z_{s_0}, \dots, Z_{s_m} - Z_{s_{m-1}}$ are independent.

Note that α -stable Lévy motion has the scaling (or self-similarity) property $Z_{at} \stackrel{d}{=} a^{1/\alpha} Z_t$ with $a > 0$. In terms of Theorem 30 in Chapter I of Protter [43], we know that every α -stable Lévy process has a càdlàg (right continuous with left limits) modification. We will henceforth assume that we are using the càdlàg version of any given α -stable Lévy process. When $0 < \alpha < 1$, Z_t is a pure jump process with locally finite variation; when $1 < \alpha < 2$, it is a purely discontinuous martingale; when $\alpha = 1$, it is a semi-martingale. For more path properties of α -stable Lévy processes, we refer to Sato [47].

Since the strong mixing property of the solution X_t to SDE (1.1) is essential in our framework, we give its definition here. The strong mixing coefficient of X is defined by

$$\alpha_X(t) = \sup_{s \in \mathbb{R}_+} \sup |P(A \cap B) - P(A)P(B)|,$$

where the second supremum is taken over measurable sets A and B in the σ -algebras generated by $\{X_u, u \leq s\}$ and $\{X_u, u \geq s + t\}$, respectively. We say that $\{X_t, t \geq 0\}$ is strongly mixing if $\alpha_X(t) \rightarrow 0$ as $t \rightarrow \infty$.

We will make use of the following assumptions:

(A.1). The drift function $b(\cdot)$ and dispersion function $\sigma(\cdot)$ satisfy a global Lipschitz condition, i.e., there exists a positive constant $L > 0$ such that

$$|b(x_1) - b(x_2)| + |\sigma(x_1) - \sigma(x_2)| \leq L|x_1 - x_2|, \quad x_1, x_2 \in \mathbb{R}.$$

(A.2). The dispersion function $\sigma(\cdot)$ satisfies the following bounded condition: there exist positive constants σ_0 and σ_1 such that $0 < \sigma_0 \leq \sigma(x) \leq \sigma_1$ for each $x \in \mathbb{R}$.

(A.3). The solution X_t admits a unique invariant distribution μ and is geometrically strong mixing (GSM), i.e. there exists $c_0 > 0$ and $\rho \in (0, 1)$ such that $\alpha_X(t) \leq c_0 \rho^t$, $t \geq 0$. Consequently, X_t is ergodic. We also assume that X_t is stationary.

(A.4). The density function $f(x)$ of the stationary distribution μ is continuous.

(A.5). The kernel function $K(\cdot)$ is a symmetric and nonnegative probability density function satisfying

$$\sup_u (1 \vee |u|)K(u) < M_0 < \infty, \quad \int_{-\infty}^{\infty} u^2 K(u) du < \infty, \quad \int_{-\infty}^{\infty} K^2(u) du < \infty.$$

(A.6). As $n \rightarrow \infty$, $h \rightarrow 0$, $\Delta \rightarrow 0$, $n\Delta \rightarrow \infty$ and $\frac{n\Delta h}{(\log(n\Delta))^2} \rightarrow \infty$.

The Lipschitz condition (A.1) is a typical condition which ensures that SDE (1.1) admits a unique non-explosive càdlàg adapted solution. For some sufficient conditions which guarantee (A.3), we refer to Theorem 2.2 and Lemma 2.4 in Masuda [35].

We first state some results about the consistency and inconsistency of the N-W estimator separately in terms of the range of α , i.e. $\alpha \in (1, 2)$ and $\alpha \in (0, 1]$.

Theorem 2.3. Assume that (A.1)-(A.6) hold. If $f(x) > 0$ and $\alpha \in (1, 2)$, then $\hat{b}_n(x) \rightarrow_P b(x)$ as $n \rightarrow \infty$.

Theorem 2.4. Assume that (A.1)-(A.6) hold. Let $f(x) > 0$ and $\alpha \in (0, 1]$. If there exists some $r \in (\frac{\alpha}{1+\alpha}, \alpha)$ such that $nh\Delta^{\frac{r}{\alpha(1-r)}} \rightarrow 0$, then $\hat{b}_n(x)$ is not consistent.

To study the asymptotic distribution and rate of convergence of the N-W estimator for the drift function, we impose some new conditions as follows:

(A.7). The drift function $b(\cdot)$ is twice continuously differentiable with bounded first and second order derivatives.

(A.8). The density function $f(x)$ of the stationary distribution μ is continuously differentiable ($f'(x)$ is continuous).

Note that (A.7) is stronger than the Lipschitz condition on $b(\cdot)$ in (A.1), and (A.8) is stronger than (A.4). We consider the rate of convergence of the N-W estimator when $1 < \alpha < 2$.

Let

$$\Lambda(x) = \frac{f(x)^{1-\frac{1}{\alpha}}}{\sigma(x) \left(\int_{-\infty}^{\infty} K^\alpha(u) du \right)^{1/\alpha}}$$

and

$$\Gamma_b(x) = \left[b'(x) \frac{f'(x)}{f(x)} + \frac{1}{2} b''(x) \right] \int_{-\infty}^{\infty} u^2 K(u) du.$$

Theorem 2.5. (a) Let $\alpha \in (1, 2)$ and assume that (A.1)-(A.3) and (A.5)-(A.8) are satisfied. We also assume that $f(x) > 0$.

(i) If $(n\Delta h)^{1-\frac{1}{\alpha}} h^2 = o(1)$ and $(n\Delta h)^{1-\frac{1}{\alpha}} \Delta^{1/\alpha} = o(1)$ then

$$(n\Delta h)^{1-\frac{1}{\alpha}} \Lambda(x) (\hat{b}_n(x) - b(x)) \Rightarrow S_\alpha(1, \beta, 0). \tag{2.1}$$

(ii) If the condition $(n\Delta h)^{1-\frac{1}{\alpha}} h^2 = o(1)$ in (i) is replaced by $(n\Delta h)^{1-\frac{1}{\alpha}} h^2 = O(1)$, then

$$(n\Delta h)^{1-\frac{1}{\alpha}} \Lambda(x) (\hat{b}_n(x) - b(x) - h^2 \Gamma_b(x)) \Rightarrow S_\alpha(1, \beta, 0). \tag{2.2}$$

(b) For $\alpha = 1$, if $nh\Delta^{\frac{q}{1-q}} = o(1)$ for some $\frac{1}{2} < q < 1$, then

$$\frac{1}{\sigma(x)} (\hat{b}_n(x) - b(x)) \Rightarrow S_\alpha(1, \beta, 0). \tag{2.3}$$

Remark 2.6. (1) Since the set of conditions (A.1)-(A.3) and (A.5)-(A.8) in Theorem 2.5 is stronger than the set of conditions (A.1)-(A.6), Theorem 2.3 and Theorem 2.4 are still valid under (A.1)-(A.3) and (A.5)-(A.8).

(2) It is easy to see that $\hat{b}_n(x)$ is not consistent when $\alpha = 1$, which is compatible with Theorem 2.4.

(3) Theorem 2.5 (a)-(ii) specifies that the bias of the N-W estimator $\hat{b}_n(x)$ is in the order of $(n\Delta h)^{1-\frac{1}{\alpha}} h^2$. If $(n\Delta h)^{1-\frac{1}{\alpha}} h^2 = o(1)$, then the estimator $\hat{b}_n(x)$ has asymptotically zero bias as stated in Theorem 2.5 (a)-(i).

Remark 2.7. We would like to give some comments on the estimation of α which is involved in the rate of convergence of the N-W estimator $\hat{b}_n(x)$ of the drift function $b(x)$ in Theorem 2.5. Note that we are dealing with non-parametric models with unknown functions $b(x)$ and $\sigma(x)$, which makes the estimation of α much harder. In order to estimate α , we need also consider the non-parametric estimation of the dispersion function $\sigma(x)$, which is our ongoing research project and will be addressed in a separate article. After getting estimation for both the drift function $b(x)$ and dispersion function $\sigma(x)$, we can use the Euler scheme to approximate the SDE (1.1):

$$\Delta X_{t_k} \approx b(X_{t_k})\Delta + \sigma(X_{t_k})\Delta Z_{t_k},$$

where $\Delta X_{t_k} = X_{t_{k+1}} - X_{t_k}$ and $\Delta Z_{t_k} = Z_{t_{k+1}} - Z_{t_k}$, $k = 0, 1, \dots, n - 1$. By replacing $b(X_{t_k})$ and $\sigma(X_{t_k})$ with $\hat{b}(X_{t_k})$ and $\hat{\sigma}(X_{t_k})$ respectively, we can regard $\frac{\Delta X_{t_k} - \hat{b}(X_{t_k})\Delta}{\hat{\sigma}(X_{t_k})}$ as an estimate of ΔZ_{t_k} for $k = 0, 1, \dots, n - 1$, which follows the stable distribution $S_\alpha(\Delta^{1/\alpha}, \beta, 0)$. Motivated by the parameter estimation for stable distributions discussed in Press [42], we can define the approximate sample characteristic function by

$$\hat{\varphi}_n(u) = \frac{1}{n} \sum_{k=0}^{n-1} \exp\{iu[\Delta X_{t_k} - \hat{b}(X_{t_k})\Delta](\hat{\sigma}(X_{t_k}))^{-1}\},$$

which shall be asymptotically equal to the characteristic function $\varphi_{Z_\Delta}(u)$ of $Z_\Delta \sim S_\alpha(\Delta^{1/\alpha}, \beta, 0)$. Note that $|\varphi_{Z_\Delta}(u)| = \exp\{-\Delta|u|^\alpha\}$. It follows that

$$\alpha = \frac{\log(-\log|\varphi_{Z_\Delta}(u)|) - \log \Delta}{\log|u|}, |u| \neq 1.$$

Then, by replacing $\varphi_{Z_\Delta}(u)$ with $\hat{\varphi}_n(u)$, we can get a point estimator of α as

$$\hat{\alpha} = \frac{\log(-\log|\hat{\varphi}_n(u)|) - \log \Delta}{\log|u|}, |u| \neq 1.$$

Of course, this estimator depends on the choice of the value of u . The consistency and asymptotic behavior of $\hat{\alpha}$ will be addressed in our ongoing research project. One can also use the regression method to estimate α as well as other parameters as discussed in Koutrouvelis [28] and [29].

Example 2.8. Here we provide one example of the SDE (1.1) which satisfies all the required conditions in our main results. We consider the following SDE

$$dX_t = \left(c - \frac{dX_{t-}^3}{\sqrt{1 + X_{t-}^4}} \right) dt + \rho(1 + \sin^2 X_{t-})dZ_t, X_0 = \eta, \quad (2.4)$$

where $c \in \mathbb{R}$, $d > 0$ and $\rho > 0$ are constants. Namely, the drift function $b(x)$ and dispersion function $\sigma(x)$ are given by

$$b(x) = c - \frac{dx^3}{\sqrt{1 + x^4}}, \quad \sigma(x) = \rho(1 + \sin^2 x).$$

It is clear that $b(x)$ and $\sigma(x)$ satisfy the Lipschitz condition (A.1), and $\sigma(x)$ satisfies (A.2). It is easy to verify that $b(x)$ is twice continuously differentiable with bounded first and second derivatives, namely $b(x)$ satisfies (A.7). In order to ensure that the SDE (2.4) satisfies (A.3), we shall employ Theorem 2.2 and Lemma 2.4 of Masuda [35] and verify that Assumption 2 and Assumption 3* in Masuda [35] are fulfilled. Assumption 2(a) is satisfied since the coefficient functions $b(x)$ and $\sigma(x)$ are smooth (see the discussion in Section 4.2 of Masuda [35]). Basic calculation shows that there exists a constant $0 < c_1 < d$ such that $xb(x) \leq -c_1|x|^2$ for every $|x|$ large enough, which implies that Assumption 3* is satisfied. Hence we conclude that (A.3) is satisfied when $\eta = \mu$. For the regularity conditions (A.4) and (A.8) of the density function of the stationary distribution μ , we can use results from Schertzer et al [48]. According to Proposition 1 of [48], we know that the transition density function $p(t, x)$ of the solution X_t to SDE (2.4) satisfies a fractional Fokker-Planck equation. In terms of Section IX of [48], when the coefficient functions $b(x)$ and $\sigma(x)$ are C^2 and Lipschitz, there exists a unique classical solution to the corresponding fractional Fokker-Planck equation and consequently the density function of the stationary distribution μ should be continuously differentiable. Thus, (A.4) and (A.8) are satisfied in our example. Finally, for condition (A.5), we can choose the normal kernel function $K(u) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{u^2}{2})$, $u \in \mathbb{R}$. Therefore the SDE (2.4) satisfies all the required conditions.

3. A simulation study

In this section, we conduct a simulation study to confirm the finite sample properties of the asymptotic results developed in Section 2. Let T be the length of observation time interval, n be the sample size, and $\Delta = T/n$ be the observation time frequency. For simplicity, let the dispersion function $\sigma(\cdot)$ be constant. The stable index α considered is 1.5 and the skewness parameter β is zero (symmetric case). We simulate and approximate X_t by using the Euler scheme (see e.g., Jacod [23], Jacod et al. [24]):

$$X_{t_{i+1}} = X_{t_i} + b(X_{t_i})\Delta + \sigma \cdot \Delta Z_{t_i}, \quad (3.1)$$

where $t_i = i\Delta$ and $\Delta Z_{t_i} = Z_{t_{i+1}} - Z_{t_i}$, $i = 0, 1, \dots, n-1$. We consider various length of observation time interval T and sample size n . The length of observation interval of the process considered is 10, 50, 100, while the sample size n considered is 1000, 5000, 10000, respectively. The drift function considered in the simulation is one of the following:

- (i). $b_1(x) = -cx + d\sqrt{1+x^2}$
- (ii). $b_2(x) = -cx + d\sin(2\pi x)$,

for various c and d . Two different values of c and d are tested. Two possible values of σ considered are 0.05 and 0.1. We consider six cases of parameter settings:

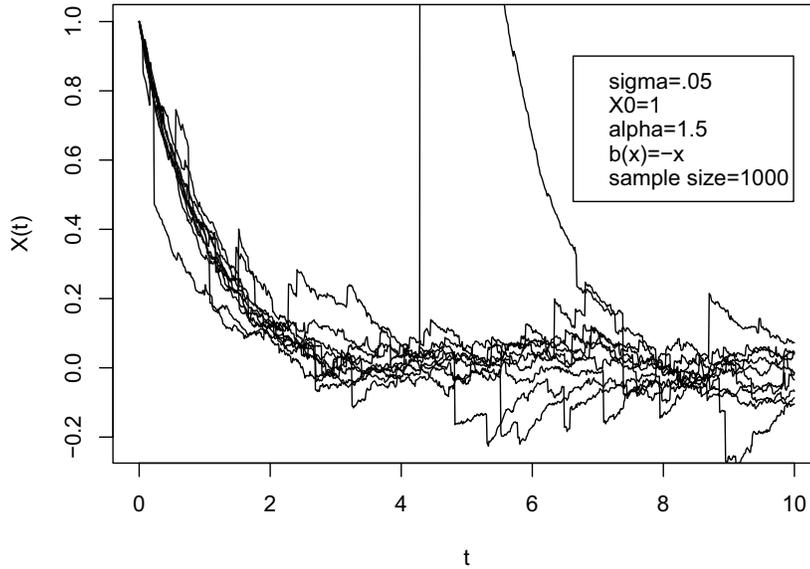


FIG 1. The ten sample paths of the process X_t .

- Case 1: $c = 1, d = 0, \sigma = 0.05$ and $b(x) = b_1(x)$;
- Case 2: $c = 1, d = 0.5, \sigma = 0.05$ and $b(x) = b_1(x)$;
- Case 3: $c = 3, d = 0.5, \sigma = 0.05$ and $b(x) = b_2(x)$.
- Case 4: $c = 1, d = 0, \sigma = 0.1$ and $b(x) = b_1(x)$;
- Case 5: $c = 1, d = 0.5, \sigma = 0.1$ and $b(x) = b_1(x)$;
- Case 6: $c = 3, d = 0.5, \sigma = 0.1$ and $b(x) = b_2(x)$.

The purpose of the six cases are two folds. For cases 1-3, we are testing the sensitivity of the N-W drift function estimator away from linearity. By changing the value of σ from 0.05 to 0.1 in cases 1, 2 and 3, we obtain cases 4, 5 and 6, respectively. The purpose of these changes is to test the sensitivity of the N-W estimators with respect to different sizes of the scale parameter σ in the Lévy driven error term.

We adapt the direct method from Janicki and Weron [25] (1994, pp. 48) to generate the α -stable process. That is: we first generate a random sample V_{t_i} uniformly distributed on $(-\pi/2, \pi/2)$ and an exponential random sample W_{t_i} with mean 1. Then we compute the symmetric α -stable random sample

$$S_{t_i} = \frac{\sin(\alpha V_{t_i})}{\{\cos(V_{t_i})\}^{1/\alpha}} \times \left\{ \frac{\cos(V_{t_i} - \alpha V_{t_i})}{W_{t_i}} \right\}^{\frac{(\alpha-1)}{\alpha}}.$$

Last, we generate the Lévy sample path by putting $\Delta Z_{t_i} = \Delta^{1/\alpha} S_{t_i}$. The initial value of the process X_t is set to be one.

Figure 1 shows the ten simulated sample paths of the process X_t . Notice that the jump error term could affect the process a lot at some typical time points.

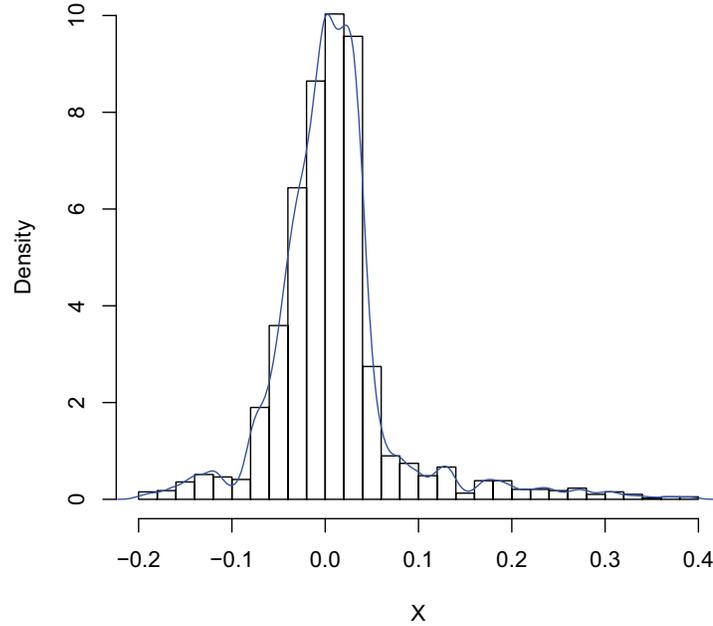


FIG 2. The estimated density function of the process X_t .

Figure 2 shows the kernel density estimate of a realization of X_t overlay with the histogram.

In computing the N-W estimate, the normal kernel is used and the bandwidth is selected according to the sample size n . In the simulation, we use $h = n^{-1/5}$. Figure 3 represents the estimated $\hat{b}(\cdot)$ from a random sample with $n = 1000$ and $h = n^{-1/5}$ and other information given in the figure. It shows that the N-W estimator performs reasonably well.

The estimator $\hat{b}(x)$ is assessed via the square-Root of Average Square Errors (RASE)

$$RASE = \left[\frac{1}{n} \sum_{k=0}^n \left\{ \hat{b}(x_k) - b(x_k) \right\}^2 \right]^{1/2},$$

where $\{x_k\}_1^n$ are chosen uniformly to cover the range of sample path of X_t . Table 1 below reports the results on $RASE$ of the N-W estimator of the drift function $b(\cdot)$ based on 1000 replicates. In each cell, the first, second and third numbers represent the sample mean, sample standard deviation and sample median of $RASE$ for three sample sizes $n = 1000$, $n = 5000$ and $n = 10000$, respectively.

Notice that as the time interval expands longer, the estimation is better as expected for whatever sample size even though the time frequency becomes larger. This means that T tending to ∞ is more important than Δ tending to 0 in the asymptotic behavior of the N-W estimator. The estimates are not so

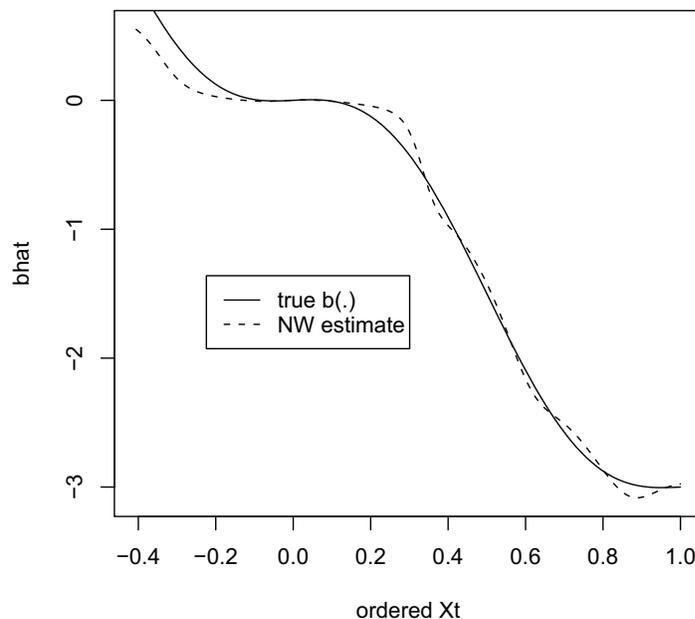


FIG 3. The N-W estimate of $b(\cdot)$ for $\sigma = 0.05, c = 3, d = 0.5$ and $b(x) = -cx + d \sin(2\pi x)$ for $T = 100, n = 1000$.

TABLE 1
Simulation results on RASE for three lengths of time interval and three sample sizes for six parameter settings

RASE	mean	sd	median	mean	sd	median	mean	sd	median
T	n=1000			n=5000			n=10000		
	Case=1								
10	0.0781	0.1234	0.0409	0.0901	0.1141	0.0558	0.1013	0.1170	0.0646
50	0.0460	0.0556	0.0306	0.0570	0.0720	0.0368	0.0656	0.0820	0.0425
100	0.0352	0.0516	0.0250	0.0434	0.0458	0.0302	0.0487	0.0686	0.0339
	Case=2								
10	0.0407	0.0682	0.0283	0.0464	0.1417	0.0264	0.0423	0.0627	0.0267
50	0.0304	0.0537	0.0215	0.0388	0.1754	0.0184	0.0346	0.0763	0.0191
100	0.0287	0.0415	0.0197	0.0318	0.0859	0.0183	0.0424	0.1316	0.0195
	Case=3								
10	0.1637	0.5541	0.1341	0.1055	0.2617	0.0684	0.0983	0.2344	0.0614
50	0.1083	0.0840	0.0995	0.0926	0.3041	0.0535	0.0875	0.2103	0.0492
100	0.0940	0.1016	0.0782	0.0672	0.1029	0.0464	0.0684	0.1655	0.0397
	Case=4								
10	0.1279	0.5408	0.0631	0.1254	0.3521	0.0611	0.1457	0.4305	0.0662
50	0.0865	0.2032	0.0472	0.0895	0.1659	0.0493	0.1140	0.3014	0.0509
100	0.0691	0.1190	0.0401	0.0912	0.1770	0.0536	0.1103	0.3473	0.0546
	Case=5								
10	0.0789	0.2310	0.0441	0.0837	0.1919	0.0474	0.0853	0.1629	0.0495
50	0.0630	0.1299	0.0371	0.0782	0.1652	0.0395	0.0927	0.2317	0.0411
100	0.0576	0.0920	0.0366	0.0803	0.2088	0.0428	0.0864	0.2096	0.0437
	Case=6								
10	0.1686	0.3104	0.1202	0.1514	0.3264	0.0929	0.2029	1.2164	0.0896
50	0.1113	0.1590	0.0771	0.1262	0.2626	0.0710	0.1615	0.5427	0.0744
100	0.0869	0.1216	0.0601	0.1128	0.2156	0.0659	0.1181	0.2026	0.0663

sensitive to the linearity assumption on the drift function. As the σ increases, the summary statistics of RASE confirm the results in Theorem 2.5. Namely, the increase of σ slows down the convergence of the N-W estimator.

As both T and n get larger, the summary statistics of RASE are getting better seen through the diagonal numerals in Table 1. The bias is getting smaller, the standard deviation is getting smaller most of the time. So, in general, when Th becomes larger, the N-W estimator becomes better with smaller RASE.

One notices that for a fixed length of observation interval T , as the sample size gets larger, the N-W estimator does not behave better, which is consistent with the asymptotic theory of the N-W estimators for stochastic processes driven by Lévy motions. This confirms that the drift function can not be identified in a fixed time interval in the framework of N-W estimator, no matter how frequently the observations are sampled. This phenomenon is also consistent with the sample paths shown in Figure 1. That is, the more often observed diffusion process, it is more affected by the Lévy driven error term which may produce many huge jumps.

4. Proofs

In this section, we provide proofs for all the results stated in Section 2. The N-W estimator is closely related to the kernel estimator $\hat{f}_n(x)$ of the density function $f(x)$ of the stationary distribution, which is defined by

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=0}^{n-1} K_h(X_{t_i} - x). \quad (4.1)$$

Denote

$$\hat{g}_n(x) = \frac{1}{n\Delta} \sum_{i=0}^{n-1} Y_i K_h(X_{t_i} - x). \quad (4.2)$$

Then, the N-W estimator given in (1.2) can be represented as

$$\hat{b}_n(x) = \hat{g}_n(x) / \hat{f}_n(x). \quad (4.3)$$

4.1. Preliminary Lemmas

Before proving our main theorems in Section 2, we prepare some preliminary lemmas. The following result (consistency of the kernel estimator) is an analogue to Lemma 2.1 in Bosq [6].

Lemma 4.1. *Under the conditions (A.1)-(A.6), we have*

$$\hat{f}_n(x) \rightarrow_P f(x) \quad \text{as } n \rightarrow \infty. \quad (4.4)$$

Proof. Note that

$$\hat{f}_n(x) - f(x) = \hat{f}_n(x) - \mathbb{E}[\hat{f}_n(x)] + \mathbb{E}[\hat{f}_n(x)] - f(x).$$

By the stationarity of the process X_t , we have

$$\begin{aligned} \mathbb{E}[\hat{f}_n(x)] &= \mathbb{E}[K_h(X_0 - x)] = \int_{-\infty}^{\infty} K_h(y - x)f(y)dy \\ &= \int_{-\infty}^{\infty} K(u)f(x + uh)du \end{aligned}$$

which converges to $f(x)$ for each x as $n \rightarrow \infty$ by Lebesgue dominated convergence theorem. Thus, it suffices to prove that $\hat{f}_n(x) - \mathbb{E}[\hat{f}_n(x)] \rightarrow_P 0$. We have

$$\begin{aligned} \hat{f}_n(x) - \mathbb{E}[\hat{f}_n(x)] &= \frac{1}{n} \sum_{i=0}^{n-1} K_h(X_{t_i} - x) - \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}[K_h(X_{t_i} - x)] \\ &= \frac{1}{n} \sum_{i=0}^{n-1} [K_h(X_{t_i} - x) - \mathbb{E}K_h(X_{t_i} - x)]. \end{aligned}$$

Let $\eta_{n,i}(x) = K_h(X_{t_{i-1}} - x) - \mathbb{E}K_h(X_{t_{i-1}} - x), i = 1, 2, \dots, n$. Note that $\sup_{1 \leq i \leq n} |\eta_{n,i}(x)| \leq C_0 h^{-1}$ a.s. for some positive constant $C_0 < \infty$. Applying Theorem 1.3 of Bosq [6], we have for each integer $q \in [1, \frac{n}{2}]$ and each $\varepsilon > 0$

$$\begin{aligned} \mathbb{P}\left(\frac{1}{n} \left| \sum_{i=1}^n \eta_{n,i}(x) \right| > \varepsilon\right) &\leq 4 \exp\left(-\frac{\varepsilon^2 q}{8v^2(q)}\right) \\ &\quad + 22 \left(1 + \frac{4C_0 h^{-1}}{\varepsilon}\right)^{1/2} q\alpha_X([p]\Delta), \end{aligned} \tag{4.5}$$

where

$$v^2(q) = \frac{2}{p^2} s(q) + \frac{C_0 h^{-1} \varepsilon}{2}$$

with $p = \frac{n}{2q}$ and

$$\begin{aligned} s(q) &= \max_{0 \leq j \leq 2q-1} \mathbb{E}[(\lfloor jp \rfloor + 1 - jp)\eta_{n,\lfloor jp \rfloor + 1}(x) + \eta_{n,\lfloor jp \rfloor + 2}(x) \\ &\quad + \dots + \eta_{n,\lfloor (j+1)p \rfloor}(x) + ((j+1)p - \lfloor (j+1)p \rfloor)\eta_{n,\lfloor (j+1)p \rfloor + 1}(x)]^2. \end{aligned}$$

Here we set $\eta_{n,n+1}(x) = 0$ for the well-definedness of $s(q)$. By using Cauchy-Schwarz inequality and stationarity of $\eta_{n,i}(x)$, it is easy to find that $s(q) = O(p^2 h^{-1})$. By choosing $q = \lceil \sqrt{n\Delta}/\sqrt{h} \rceil$ and $p = \frac{n}{2q} = O(\sqrt{nh}/\sqrt{\Delta})$, we obtain

$$\frac{\varepsilon^2 q}{8v^2(q)} = \varepsilon^2 \cdot O(qh) = O(\varepsilon^2 \sqrt{n\Delta h}). \tag{4.6}$$

We claim that

$$22 \left(1 + \frac{4C_0 h^{-1}}{\varepsilon}\right)^{1/2} q\alpha_X([p]\Delta) \leq C(\varepsilon) \exp(-O(\sqrt{n\Delta h})). \tag{4.7}$$

By using the GSM property of X_t , we find that there exists some $a > 0$ such that

$$22 \left(1 + \frac{4C_0 h^{-1}}{\varepsilon}\right)^{1/2} q_{\alpha_X}([p]\Delta) = O\left(\frac{\sqrt{n\Delta}}{h} \exp(-a\sqrt{n\Delta h})\right).$$

Hence it suffices to show that

$$\frac{\sqrt{n\Delta}}{h} \exp(-a\sqrt{n\Delta h}) \leq \exp(-a'\sqrt{n\Delta h}) \quad (4.8)$$

for some $a' > 0$. Some basic calculation yields that (4.8) is equivalent to

$$a - a' \geq \frac{3}{2} \cdot \frac{\log(n\Delta)}{\sqrt{n\Delta h}} - \frac{\log(n\Delta h)}{\sqrt{n\Delta h}}.$$

The existence of such a a' (e.g. $a' = \frac{a}{2}$) is justified since the right hand side of the above inequality tends to zero under (A.6). Finally, combining (4.5), (4.6) and (4.7), we have

$$\mathbb{P}\left(\frac{1}{n} \left| \sum_{i=1}^n \eta_{n,i}(x) \right| > \varepsilon\right) \leq C(\varepsilon) \exp\{-O(\varepsilon^2 \sqrt{n\Delta h})\}. \quad (4.9)$$

Therefore, the desired convergence result (4.4) follows from given conditions. \square

Next we establish a central limit theorem (CLT) for stable stochastic integrals, which has some independent interest. The CLT will be crucial in establishing the consistency (or inconsistency) and asymptotic distribution of the N-W estimator. Let $\phi(t)$ be a predictable process satisfying $\int_0^T |\phi(t)|^\alpha dt < \infty$ almost surely for $T < \infty$. Then the stochastic integral $\int_0^t \phi(s) dZ_s$ is well-defined (see e.g., Rosinski and Woyczynski [45], Kallenberg [27]). We assume that either $\phi(t)$ is nonnegative or Z is symmetric. Then, we have the following version of Lengart's inequality in the stable setting.

Lemma 4.2. *For any given $\varepsilon > 0$ and $\delta > 0$, there is some constant $c > 0$ such that*

$$P\left(\sup_{0 \leq t \leq T} \left| \int_0^t \phi(s) dZ_s \right| > \varepsilon\right) \leq \frac{c\delta}{\varepsilon^\alpha} + P\left(\int_0^T |\phi(t)|^\alpha dt > \delta\right). \quad (4.10)$$

Proof. Let $S_t = \int_0^t |\phi(s)|^\alpha ds$. By Theorem 4.1 and Theorem 4.2 of Kallenberg [27] (see also Theorem 3.1 in Rosinski and Woyczynski [45] for the symmetric case), there exists a strictly α -stable process Z' with the same finite-dimensional distributions as Z such that $\int_0^t \phi(s) dZ_s = Z'(S_t)$ almost surely. By the classical maximal inequality (see e.g., Proposition 10.2 of Fristedt [16]), we find that for

some $c > 0$

$$\begin{aligned}
 & P\left(\sup_{0 \leq t \leq T} \left| \int_0^t \phi(s) dZ_s \right| > \varepsilon\right) \\
 & \leq P\left(\sup_{0 \leq t \leq T} \left| \int_0^t \phi(s) dZ_s \right| > \varepsilon, S_T \leq \delta\right) + P(S_T > \delta) \\
 & \leq P\left(\sup_{0 \leq t \leq T} |Z'(S_t)| > \varepsilon, S_T \leq \delta\right) + P(S_T > \delta) \\
 & \leq P\left(\sup_{0 \leq s \leq \delta} |Z'(s)| > \varepsilon\right) + P(S_T > \delta) \\
 & \leq \frac{c\delta}{\varepsilon^\alpha} + P\left(\int_0^T |\phi(t)|^\alpha dt > \delta\right).
 \end{aligned}$$

This completes the proof. □

The following result is a version of the CLT for stable stochastic integrals.

Lemma 4.3. *Suppose that there is a deterministic and nonnegative function Φ such that*

$$\Phi^\alpha(T) \int_0^T |\phi(t)|^\alpha dt \rightarrow_P 1 \text{ as } T \rightarrow \infty.$$

Then, we have

$$\Phi(T) \int_0^T \phi(t) dZ_t \Rightarrow S_\alpha(1, \beta, 0). \tag{4.11}$$

Proof. Let $R_t = \Phi^\alpha(T) \int_0^t |\phi(s)|^\alpha ds$. For a fixed T , we redefine the function ϕ on the interval $(T, T + 1]$ as $\phi(t) = \Phi^{-1}(T)$ and define the stopping time $\tau_T = \inf\{t \geq 0 : R_t > 1\}$. Then, $\tau_T \in [0, T + 1]$ almost surely. Note that there is a strictly α -stable process Z' with the same finite-dimensional distributions as Z such that $\Phi(T) \int_0^t \phi(t) dZ_t = Z'_{R_t}$. It is easy to see that

$$\Phi(T) \int_0^{\tau_T} \phi(t) dZ_t = Z'_1 \sim S_\alpha(1, \beta, 0).$$

By using Lemma 4.2 and following exactly the same arguments as in the proof of Theorem 1.19 in Kutoyants [30], we can show that the characteristic function of $\Phi(T) \int_0^T \phi(t) dZ_t$ converges to the characteristic function of $\Phi(T) \int_0^{\tau_T} \phi(t) dZ_t$ as $T \rightarrow \infty$. Therefore, by the continuity theorem (see Theorem 26.3 of Billingsley [5]), it immediately follows that (4.11) holds. □

We say that a continuous function $F : [0, \infty) \rightarrow [0, \infty)$ grows more slowly than u^α ($\alpha > 0$) if there exist positive constants c, λ_0 and $\alpha_0 < \alpha$ such that $F(\lambda u) \leq c\lambda^{\alpha_0} F(u)$ for all $u > 0$ and all $\lambda \geq \lambda_0$. The following lemma (Lemma 2.4 of Long [33]) provides moment inequalities for stable stochastic integrals, which will be a crucial tool in the proofs of our main results.

Lemma 4.4. Let $\phi(t)$ be a predictable process satisfying $\int_0^T |\phi(t)|^\alpha dt < \infty$ almost surely for $T < \infty$. We assume that either ϕ is nonnegative or Z is symmetric. If $F(u)$ grows more slowly than u^α , then there exist positive constants c_1 and c_2 depending only on $\alpha, \alpha_0, \beta, c$ and λ_0 such that for each $T > 0$

$$\begin{aligned} c_1 \mathbb{E} \left[F \left(\left(\int_0^T |\phi(t)|^\alpha dt \right)^{1/\alpha} \right) \right] &\leq \mathbb{E} \left[F \left(\sup_{t \leq T} \left| \int_0^t \phi(s) dZ_s \right| \right) \right] \\ &\leq c_2 \mathbb{E} \left[F \left(\left(\int_0^T |\phi(t)|^\alpha dt \right)^{1/\alpha} \right) \right]. \end{aligned} \quad (4.12)$$

4.2. Proof of Theorem 2.3

By (4.3) and Lemma 4.1, it suffices to prove that

$$\hat{g}_n(x) \rightarrow_P f(x)b(x) \text{ as } n \rightarrow \infty. \quad (4.13)$$

We first note that

$$\begin{aligned} Y_i &= \int_{t_i}^{t_{i+1}} b(X_{s-}) ds + \int_{t_i}^{t_{i+1}} \sigma(X_{s-}) dZ_s \\ &= b(X_{t_i})\Delta + \int_{t_i}^{t_{i+1}} (b(X_{s-}) - b(X_{t_i})) ds + \int_{t_i}^{t_{i+1}} \sigma(X_{s-}) dZ_s. \end{aligned}$$

Then, by (4.2), it follows that

$$\begin{aligned} \hat{g}_n(x) &= \frac{1}{n} \sum_{i=0}^{n-1} b(X_{t_i}) K_h(X_{t_i} - x) \\ &\quad + \frac{1}{n\Delta} \sum_{i=0}^{n-1} K_h(X_{t_i} - x) \int_{t_i}^{t_{i+1}} (b(X_{s-}) - b(X_{t_i})) ds \\ &\quad + \frac{1}{n\Delta} \sum_{i=0}^{n-1} K_h(X_{t_i} - x) \int_{t_i}^{t_{i+1}} \sigma(X_{s-}) dZ_s \\ &=: g_{n,1}(x) + g_{n,2}(x) + g_{n,3}(x). \end{aligned} \quad (4.14)$$

We have the following claims:

- (i) $g_{n,1}(x) \rightarrow_P f(x)b(x)$ as $n \rightarrow \infty$;
- (ii) $g_{n,2}(x) \rightarrow_P 0$ as $n \rightarrow \infty$;
- (iii) $g_{n,3}(x) \rightarrow_P 0$ as $n \rightarrow \infty$.

These three claims guarantee that (4.13) holds.

Proof of Claim (i). Note that

$$\begin{aligned} g_{n,1}(x) &= b(x) \frac{1}{n} \sum_{i=0}^{n-1} K_h(X_{t_i} - x) + \frac{1}{n} \sum_{i=0}^{n-1} (b(X_{t_i}) - b(x)) K_h(X_{t_i} - x) \\ &=: B_{n,1}(x) + B_{n,2}(x). \end{aligned} \quad (4.15)$$

By Lemma 4.1, it is clear that $B_{n,1}(x) \rightarrow_P b(x)f(x)$ when $n \rightarrow \infty$. For $B_{n,2}(x)$, by the Lipschitz property of $b(\cdot)$ and stationarity of X_t , we have

$$\begin{aligned} |B_{n,2}(x)| &\leq \frac{1}{n} \sum_{i=0}^{n-1} L|X_{t_i} - x|K_h(X_{t_i} - x) \\ &\leq L \frac{1}{n} \sum_{i=0}^{n-1} (|X_{t_i} - x|K_h(X_{t_i} - x) - \mathbb{E}[|X_{t_i} - x|K_h(X_{t_i} - x)]) \\ &\quad + L \cdot \mathbb{E}[|X_0 - x|K_h(X_0 - x)]. \end{aligned} \tag{4.16}$$

Note that $|X_{t_i} - x|K_h(X_{t_i} - x) - \mathbb{E}[|X_{t_i} - x|K_h(X_{t_i} - x)]$ is uniformly bounded for each i under condition (A.5). By slightly modifying the proof of Lemma 4.1, we can show that

$$L \frac{1}{n} \sum_{i=0}^{n-1} (|X_{t_i} - x|K_h(X_{t_i} - x) - \mathbb{E}[|X_{t_i} - x|K_h(X_{t_i} - x)]) \rightarrow_P 0 \tag{4.17}$$

when $n \rightarrow \infty$. By using the continuity of $f(x)$ and Lebesgue dominated convergence theorem, we find

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{\mathbb{E}[|X_0 - x|K_h(X_0 - x)]}{h} &= \lim_{h \rightarrow 0} \frac{1}{h} \int_{-\infty}^{\infty} |y - x|K_h(y - x)f(y)dy \\ &= \lim_{h \rightarrow 0} \int_{-\infty}^{\infty} |u|K(u)f(x + uh)du \\ &= f(x) \int_{-\infty}^{\infty} |u|K(u)du. \end{aligned} \tag{4.18}$$

Combining (4.16), (4.17) and (4.18), it follows that $B_{n,2}(x) \rightarrow_P 0$ when $n \rightarrow \infty$. Hence the claim (i) holds. \square

Proof of Claim (ii). By using the Lipschitz condition of $b(\cdot)$, we have

$$\begin{aligned} |g_{n,2}(x)| &\leq \frac{1}{n\Delta} \sum_{i=0}^{n-1} K_h(X_{t_i} - x) \int_{t_i}^{t_{i+1}} |b(X_{s-}) - b(X_{t_i})|ds \\ &\leq \frac{L}{n\Delta} \sum_{i=0}^{n-1} K_h(X_{t_i} - x) \int_{t_i}^{t_{i+1}} |X_{s-} - X_{t_i}|ds \\ &\leq \frac{L}{n} \sum_{i=0}^{n-1} K_h(X_{t_i} - x) \sup_{t_i \leq t \leq t_{i+1}} |X_t - X_{t_i}|. \end{aligned} \tag{4.19}$$

Let us consider the estimate of $\sup_{t_i \leq t \leq t_{i+1}} |X_t - X_{t_i}|$. Note that for $t_i \leq t \leq t_{i+1}$

$$X_t - X_{t_i} = \int_{t_i}^t b(X_{s-})ds + \int_{t_i}^t \sigma(X_{s-})dZ_s.$$

By using Lipschitz condition on $b(\cdot)$ again, we find

$$\begin{aligned} |X_t - X_{t_i}| &\leq \int_{t_i}^t |b(X_{s-})| ds + \left| \int_{t_i}^t \sigma(X_{s-}) dZ_s \right| \\ &\leq \int_{t_i}^t (|b(X_{s-}) - b(X_{t_i})| + |b(X_{t_i})|) ds + \left| \int_{t_i}^t \sigma(X_{s-}) dZ_s \right| \\ &\leq |b(X_{t_i})| \Delta + \sup_{t_i \leq t \leq t_{i+1}} \left| \int_{t_i}^t \sigma(X_{s-}) dZ_s \right| + L \int_{t_i}^t |X_s - X_{t_i}| ds. \end{aligned}$$

By Gronwall's inequality, we have

$$|X_t - X_{t_i}| \leq \left(|b(X_{t_i})| \Delta + \sup_{t_i \leq t \leq t_{i+1}} \left| \int_{t_i}^t \sigma(X_{s-}) dZ_s \right| \right) e^{L(t-t_i)}.$$

It follows that

$$\sup_{t_i \leq t \leq t_{i+1}} |X_t - X_{t_i}| \leq e^{L\Delta} \left(|b(X_{t_i})| \Delta + \sup_{t_i \leq t \leq t_{i+1}} \left| \int_{t_i}^t \sigma(X_{s-}) dZ_s \right| \right). \quad (4.20)$$

By (4.19) and (4.20), we find

$$\begin{aligned} |g_{n,2}(x)| &\leq \frac{L}{n} \sum_{i=0}^{n-1} K_h(X_{t_i} - x) e^{L\Delta} \left(|b(X_{t_i})| \Delta + \sup_{t_i \leq t \leq t_{i+1}} \left| \int_{t_i}^t \sigma(X_{s-}) dZ_s \right| \right) \\ &\leq L\Delta e^{L\Delta} \cdot \frac{1}{n} \sum_{i=0}^{n-1} K_h(X_{t_i} - x) |b(X_{t_i})| \\ &\quad + L e^{L\Delta} \cdot \frac{1}{n} \sum_{i=0}^{n-1} K_h(X_{t_i} - x) \sup_{t_i \leq t \leq t_{i+1}} \left| \int_{t_i}^t \sigma(X_{s-}) dZ_s \right| \\ &=: B_{n,3}(x) + B_{n,4}(x). \end{aligned} \quad (4.21)$$

By the claim (i) with b being replaced by $|b|$, we know that

$$\frac{1}{n} \sum_{i=0}^{n-1} K_h(X_{t_i} - x) |b(X_{t_i})| \rightarrow_P |b(x)| f(x)$$

when $n \rightarrow \infty$. This implies that $B_{n,3}(x) \rightarrow_P 0$ when $n \rightarrow \infty$. By Markov inequality and Lemma 4.4, we have

$$\begin{aligned} &P \left(\frac{1}{n} \sum_{i=0}^{n-1} K_h(X_{t_i} - x) \sup_{t_i \leq t \leq t_{i+1}} \left| \int_{t_i}^t \sigma(X_{s-}) dZ_s \right| > \varepsilon \right) \\ &\leq \frac{1}{n\varepsilon} \sum_{i=0}^{n-1} \mathbb{E} \left[\sup_{t_i \leq t \leq t_{i+1}} \left| \int_{t_i}^t K_h(X_{t_i} - x) \sigma(X_{s-}) dZ_s \right| \right] \end{aligned}$$

$$\begin{aligned}
 &\leq \frac{C_1}{n\varepsilon} \sum_{i=0}^{n-1} \mathbb{E} \left[\left(\int_{t_i}^{t_{i+1}} K_h^\alpha(X_{t_i} - x) \sigma^\alpha(X_{s-}) ds \right)^{1/\alpha} \right] \\
 &\leq \frac{C_1}{n\varepsilon} \sum_{i=0}^{n-1} \mathbb{E}[K_h(X_{t_i} - x) \sigma_1 \Delta^{1/\alpha}] \\
 &\leq O(\Delta^{1/\alpha}).
 \end{aligned} \tag{4.22}$$

Hence, $B_{n,4}(x) \rightarrow_P 0$ as $n \rightarrow \infty$. This completes the proof of claim (ii). \square

Proof of Claim (iii). We define

$$\phi_n(t, x) = \sum_{i=0}^{n-1} \frac{1}{h^{1/\alpha}} K\left(\frac{X_{t_i} - x}{h}\right) \sigma(X_{t-}) 1_{(t_i, t_{i+1}]}(t), \tag{4.23}$$

so that $\phi_n(t, x)$ is a predictable process. Then, we have

$$g_{n,3}(x) = \frac{1}{n\Delta h^{\frac{\alpha-1}{\alpha}}} \int_0^{t_n} \phi_n(t, x) dZ_t.$$

By using Markov inequality, Lemma 4.4, boundedness of $\sigma(X_{t-})$, and stationarity of X_t , we find for some constant $C_2 > 0$

$$\begin{aligned}
 &P(|g_{n,3}(x)| > \varepsilon) \\
 &\leq \frac{1}{n\Delta h^{\frac{\alpha-1}{\alpha}} \varepsilon} \mathbb{E} \left| \int_0^{t_n} \phi_n(t, x) dZ_t \right| \\
 &\leq \frac{C_2}{n\Delta h^{\frac{\alpha-1}{\alpha}} \varepsilon} \mathbb{E} \left[\left(\int_0^{t_n} |\phi_n(t, x)|^\alpha dt \right)^{1/\alpha} \right] \\
 &\leq \frac{C_2}{n\Delta h^{\frac{\alpha-1}{\alpha}} \varepsilon} \left(\mathbb{E} \left[\sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \frac{1}{h} K^\alpha\left(\frac{X_{t_i} - x}{h}\right) \sigma^\alpha(X_{t-}) dt \right] \right)^{1/\alpha} \\
 &\leq \frac{C_2}{n\Delta h^{\frac{\alpha-1}{\alpha}} \varepsilon} \left(n\Delta \sigma_1^\alpha \mathbb{E} \left[\frac{1}{h} K^\alpha\left(\frac{X_0 - x}{h}\right) \right] \right)^{1/\alpha} \\
 &\leq \frac{C_2 \sigma_1}{n\Delta h^{\frac{\alpha-1}{\alpha}} \varepsilon} (n\Delta)^{\frac{1}{\alpha}} O(1) \\
 &= O((n\Delta h)^{\frac{1-\alpha}{\alpha}}),
 \end{aligned} \tag{4.24}$$

which goes to zero under condition (A.6). This shows that claim (iii) holds. \square

4.3. Proof of Theorem 2.4

The claims (i) in the proof of Theorem 2.3 is still true when $0 < \alpha \leq 1$. For claim (ii), we need to make some minor modifications on its proof. It is clear that we still have $B_{n,3}(x) \rightarrow_P 0$ under (A.1)-(A.6). For $B_{n,4}(x)$, by Markov

inequality, Lemma 4.4, condition (A.2), and stationarity of X_t , we have for $\frac{\alpha}{1+\alpha} < r < \alpha \leq 1$

$$\begin{aligned}
& P\left(\frac{1}{n} \sum_{i=0}^{n-1} K_h(X_{t_i} - x) \sup_{t_i \leq t \leq t_{i+1}} \left| \int_{t_i}^t \sigma(X_{s-}) dZ_s \right| > \varepsilon\right) \\
& \leq \frac{1}{n^r \varepsilon^r} \sum_{i=0}^{n-1} \mathbb{E} \left[\sup_{t_i \leq t \leq t_{i+1}} \left| \int_{t_i}^t K_h(X_{t_i} - x) \sigma(X_{s-}) dZ_s \right|^r \right] \\
& \leq \frac{C_1}{n^r \varepsilon^r} \sum_{i=0}^{n-1} \mathbb{E} \left[\left(\int_{t_i}^{t_{i+1}} K_h^\alpha(X_{t_i} - x) \sigma^\alpha(X_{s-}) ds \right)^{r/\alpha} \right] \\
& \leq \frac{C_1}{n^r \varepsilon^r} \sum_{i=0}^{n-1} \mathbb{E}[K_h^r(X_{t_i} - x) \sigma_1^r \Delta^{r/\alpha}] \\
& \leq O((nh\Delta^{\frac{r}{\alpha(1-r)}})^{1-r}). \tag{4.25}
\end{aligned}$$

which tends to zero when $nh\Delta^{\frac{r}{\alpha(1-r)}} \rightarrow 0$. So, under this extra condition, claim (ii) holds when $0 < \alpha \leq 1$. However, we shall prove that claim (iii) is not true, i.e., $g_{n,3}(x)$ is no longer converging to zero in probability as $n \rightarrow \infty$ and consequently $\hat{b}_n(x)$ is not consistent. Recall that

$$\phi_n(t, x) = \sum_{i=0}^{n-1} \frac{1}{h^{1/\alpha}} K\left(\frac{X_{t_i} - x}{h}\right) \sigma(X_{t-}) 1_{(t_i, t_{i+1}]}(t).$$

Let

$$\Phi_{t_n} = \left(t_n \sigma^\alpha(x) f(x) \int_{-\infty}^{\infty} K^\alpha(u) du \right)^{-\frac{1}{\alpha}}.$$

Then, we have

$$\begin{aligned}
g_{n,3}(x) &= \frac{1}{n\Delta h^{1-\frac{1}{\alpha}}} \int_0^{t_n} \phi_n(t, x) dZ_t \\
&= \frac{\sigma(x) f^{\frac{1}{\alpha}}(x) \left(\int_{-\infty}^{\infty} K^\alpha(u) du \right)^{\frac{1}{\alpha}}}{(n\Delta h)^{1-\frac{1}{\alpha}}} \cdot \Phi_{t_n} \int_0^{t_n} \phi_n(t, x) dZ_t,
\end{aligned}$$

or equivalently

$$\Phi_{t_n} \int_0^{t_n} \phi_n(t, x) dZ_t = \frac{(n\Delta h)^{1-\frac{1}{\alpha}}}{\sigma(x) f(x)^{\frac{1}{\alpha}} \left(\int_{-\infty}^{\infty} K^\alpha(u) du \right)^{\frac{1}{\alpha}}} g_{n,3}(x). \tag{4.26}$$

Note that

$$\begin{aligned}
\phi_n^\alpha(t, x) &= \left[\sum_{i=0}^{n-1} \frac{1}{h^{1/\alpha}} K\left(\frac{X_{t_i} - x}{h}\right) \sigma(X_{t-}) 1_{(t_i, t_{i+1}]}(t) \right]^\alpha \\
&= \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha\left(\frac{X_{t_i} - x}{h}\right) \sigma^\alpha(X_{t-}) 1_{(t_i, t_{i+1}]}(t)
\end{aligned}$$

and consequently

$$\begin{aligned}
 & \Phi_{t_n}^\alpha \cdot \int_0^{t_n} \phi_n^\alpha(t, x) dt \\
 = & \Phi_{t_n}^\alpha \cdot \int_0^{t_n} \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha \left(\frac{X_{t_i} - x}{h} \right) \sigma^\alpha(X_{t-}) 1_{(t_i, t_{i+1}]}(t) dt \\
 = & \Phi_{t_n}^\alpha \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha \left(\frac{X_{t_i} - x}{h} \right) \int_{t_i}^{t_{i+1}} \sigma^\alpha(X_{t-}) dt \\
 = & \Phi_{t_n}^\alpha \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha \left(\frac{X_{t_i} - x}{h} \right) \sigma^\alpha(X_{t_i}) \Delta \\
 & + \Phi_{t_n}^\alpha \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha \left(\frac{X_{t_i} - x}{h} \right) \int_{t_i}^{t_{i+1}} (\sigma^\alpha(X_{t-}) - \sigma^\alpha(X_{t_i})) dt \\
 =: & I + J. \tag{4.27}
 \end{aligned}$$

Similar to Lemma 4.1, it is easy to prove that

$$\frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha \left(\frac{X_{t_i} - x}{h} \right) \sigma^\alpha(X_{t_i}) \rightarrow_P \sigma^\alpha(x) f(x) \int_{-\infty}^{\infty} K^\alpha(u) du. \tag{4.28}$$

Therefore, we have

$$I = \frac{1}{\sigma^\alpha(x) f(x) \int_{-\infty}^{\infty} K^\alpha(u) du} \cdot \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha \left(\frac{X_{t_i} - x}{h} \right) \sigma^\alpha(X_{t_i}) \rightarrow_P 1. \tag{4.29}$$

Next we deal with the second term J . By Lipschitz condition (A.1) on $\sigma(\cdot)$, (4.20) and basic inequality $\|u + v\|^q - \|v\|^q \leq \|u\|^q$ for $u, v \in \mathbb{R}$ and $q \in (0, 1]$, we have

$$\begin{aligned}
 |J| &= \Phi_{t_n}^\alpha \left| \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha \left(\frac{X_{t_i} - x}{h} \right) \int_{t_i}^{t_{i+1}} (\sigma^\alpha(X_{t-}) - \sigma^\alpha(X_{t_i})) dt \right| \\
 &\leq \Phi_{t_n}^\alpha \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha \left(\frac{X_{t_i} - x}{h} \right) \int_{t_i}^{t_{i+1}} |\sigma^\alpha(X_{t-}) - \sigma^\alpha(X_{t_i})| dt \\
 &\leq \Phi_{t_n}^\alpha \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha \left(\frac{X_{t_i} - x}{h} \right) \int_{t_i}^{t_{i+1}} |\sigma(X_{t-}) - \sigma(X_{t_i})|^\alpha dt \\
 &\leq \Phi_{t_n}^\alpha \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha \left(\frac{X_{t_i} - x}{h} \right) L^\alpha \Delta \sup_{t_i \leq t \leq t_{i+1}} |X_t - X_{t_i}|^\alpha \\
 &\leq \frac{L^\alpha e^{\alpha L \Delta} \Delta^\alpha}{\sigma^\alpha(x) f(x) \int_{-\infty}^{\infty} K^\alpha(u) du} \cdot \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha \left(\frac{X_{t_i} - x}{h} \right) |b(X_{t_i})|^\alpha
 \end{aligned}$$

$$\begin{aligned}
 & + \frac{L^\alpha e^{\alpha L \Delta}}{\sigma^\alpha(x) f(x) \int_{-\infty}^\infty K^\alpha(u) du} \cdot \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha\left(\frac{X_{t_i} - x}{h}\right) \\
 & \times \sup_{t_i \leq t \leq t_{i+1}} \left| \int_{t_i}^t \sigma(X_{s-}) dZ_s \right|^\alpha \\
 = & J_1 + J_2. \tag{4.30}
 \end{aligned}$$

It is clear that $J_1 \rightarrow_P 0$ under given conditions since (similar to Lemma 4.1)

$$\frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha\left(\frac{X_{t_i} - x}{h}\right) |b(X_{t_i})|^\alpha \rightarrow_P |b(x)|^\alpha f(x) \int_{-\infty}^\infty K^\alpha(u) du. \tag{4.31}$$

To prove that $J_2 \rightarrow_P 0$, it is sufficient to show that

$$\frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha\left(\frac{X_{t_i} - x}{h}\right) \sup_{t_i \leq t \leq t_{i+1}} \left| \int_{t_i}^t \sigma(X_{s-}) dZ_s \right|^\alpha \rightarrow_P 0.$$

By using Markov inequality, Lemma 4.4, and condition (A.2), we have for $q < 1$

$$\begin{aligned}
 & P\left(\frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha\left(\frac{X_{t_i} - x}{h}\right) \sup_{t_i \leq t \leq t_{i+1}} \left| \int_{t_i}^t \sigma(X_{s-}) dZ_s \right|^\alpha > \varepsilon\right) \\
 \leq & \frac{1}{(nh\varepsilon)^q} \mathbb{E} \left[\sum_{i=0}^{n-1} K^\alpha\left(\frac{X_{t_i} - x}{h}\right) \sup_{t_i \leq t \leq t_{i+1}} \left| \int_{t_i}^t \sigma(X_{s-}) dZ_s \right|^{\alpha q} \right] \\
 \leq & \frac{1}{(nh\varepsilon)^q} \sum_{i=0}^{n-1} \mathbb{E} \left[\left(\sup_{t_i \leq t \leq t_{i+1}} \left| \int_{t_i}^t K\left(\frac{X_{t_i} - x}{h}\right) \sigma(X_{s-}) dZ_s \right| \right)^{q\alpha} \right] \\
 \leq & \frac{C}{(nh\varepsilon)^q} \sum_{i=0}^{n-1} \mathbb{E} \left[\left(\int_{t_i}^{t_{i+1}} K^\alpha\left(\frac{X_{t_i} - x}{h}\right) \sigma^\alpha(X_{s-}) ds \right)^q \right] \\
 \leq & \frac{C}{(nh\varepsilon)^q} \sum_{i=0}^{n-1} \mathbb{E} \left[K^{q\alpha}\left(\frac{X_{t_i} - x}{h}\right) \sigma_1^{q\alpha} \Delta^q \right] \\
 = & \frac{nC\sigma_1^{q\alpha} \Delta^q}{(nh\varepsilon)^q} \int_{-\infty}^\infty K^{q\alpha}\left(\frac{y - x}{h}\right) f(y) dy \\
 = & O((nh\Delta^{\frac{q}{1-q}})^{1-q}), \tag{4.32}
 \end{aligned}$$

which goes to zero if we choose $\frac{r}{\alpha+(1-\alpha)r} \leq q < 1$ so that $\frac{q}{1-q} \geq \frac{r}{\alpha(1-r)}$. Thus, combining (4.27), (4.29)-(4.32), we find that

$$\Phi_{t_n}^\alpha \cdot \int_0^{t_n} \phi_n^\alpha(t, x) dt \rightarrow_P 1. \tag{4.33}$$

Hence, by Lemma 4.3, we have

$$\Phi_{t_n} \int_0^{t_n} \phi_n(t, x) dZ_t \Rightarrow S_\alpha(1, \beta, 0). \tag{4.34}$$

If $g_{n,3}(x) \rightarrow_P 0$ as $n \rightarrow \infty$, then the right hand side of (4.26) converges to zero in probability as $n \rightarrow \infty$ since $(n\Delta h)^{1-\frac{1}{\alpha}} \rightarrow 0$ when $0 < \alpha < 1$ and $(n\Delta h)^{1-\frac{1}{\alpha}} = 1$ when $\alpha = 1$ under condition (A.6). This contradicts (4.34). Therefore, we conclude that $g_{n,3}(x)$ does not converge to zero in probability. This completes the proof. \square

4.4. Proof of Theorem 2.5

Proof of Theorem 2.5 (a)-(i). Note that

$$\begin{aligned} (n\Delta h)^{1-\frac{1}{\alpha}}\Lambda(x)(\hat{b}_n(x) - b(x)) &= \frac{(n\Delta h)^{1-\frac{1}{\alpha}}\Lambda(x)[\hat{g}_n(x) - b(x)\hat{f}_n(x)]}{\hat{f}_n(x)} \\ &:= \frac{V_n(x)}{\hat{f}_n(x)}. \end{aligned} \tag{4.35}$$

Since $\hat{f}_n(x) \rightarrow f(x)$ in probability as $n \rightarrow \infty$, it is enough to study the asymptotic behavior of $V_n(x)$. By (4.14), we find

$$\begin{aligned} V_n(x) &= (n\Delta h)^{1-\frac{1}{\alpha}}\Lambda(x)[g_{n,1}(x) - b(x)\hat{f}_n(x)] \\ &\quad + (n\Delta h)^{1-\frac{1}{\alpha}}\Lambda(x)g_{n,2}(x) \\ &\quad + (n\Delta h)^{1-\frac{1}{\alpha}}\Lambda(x)g_{n,3}(x) \\ &:= V_{n,1}(x) + V_{n,2}(x) + V_{n,3}(x). \end{aligned} \tag{4.36}$$

Under the given conditions, we have the following claims:

Claim 1.

$$\begin{aligned} &V_{n,1}(x) \\ &= (n\Delta h)^{1-\frac{1}{\alpha}}h^2\Lambda(x) \left[(b'(x)f'(x) + \frac{1}{2}b''(x)f(x)) \int_{-\infty}^{\infty} u^2K(u)du + o(1) \right] \\ &\quad + o_P(1). \end{aligned} \tag{4.37}$$

Claim 2.

$$V_{n,2}(x) = o_P(1). \tag{4.38}$$

Claim 3.

$$V_{n,3}(x) \Rightarrow f(x)S_\alpha(1, \beta, 0). \tag{4.39}$$

Here the notation $o_P(1)$ (or $O_P(1)$) means a sequence of random variables converging to zero (or a finite constant) in probability.

Proof of Claim 1. We can express $V_{n,1}(x)$ as

$$V_{n,1}(x) = (n\Delta h)^{1-\frac{1}{\alpha}}\Lambda(x) \cdot \frac{1}{n} \sum_{i=0}^{n-1} (b(X_{t_i}) - b(x))K_h(X_{t_i} - x).$$

Note that $V_{n,1}(x) = \mathbb{E}[V_{n,1}(x)] + (V_{n,1}(x) - \mathbb{E}[V_{n,1}(x)])$. We have the following two claims:

Claim 1-(i):

$$\begin{aligned} & \mathbb{E}[V_{n,1}(x)] \\ &= (n\Delta h)^{1-\frac{1}{\alpha}} h^2 \Lambda(x) \left[(b'(x)f'(x) + \frac{1}{2}b''(x)f(x)) \int_{-\infty}^{\infty} u^2 K(u) du + o(1) \right]. \end{aligned} \quad (4.40)$$

Claim 1-(ii):

$$V_{n,1}(x) - \mathbb{E}[V_{n,1}(x)] = o_P(1). \quad (4.41)$$

Then, claim 1 follows immediately from claims 1-(i) and 1-(ii).

Proof of Claim 1-(i). By stationarity of X_t , condition (A.5) for the kernel function $K(\cdot)$, Taylor's formula and Lebesgue dominated convergence, we find that for some $\theta, \tilde{\theta} \in [0, 1]$

$$\begin{aligned} & \mathbb{E}[V_{n,1}(x)] \\ &= (n\Delta h)^{1-\frac{1}{\alpha}} \Lambda(x) \cdot \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}[(b(X_{t_i}) - b(x))K_h(X_{t_i} - x)] \\ &= (n\Delta h)^{1-\frac{1}{\alpha}} \Lambda(x) \mathbb{E}[(b(X_0) - b(x))K_h(X_0 - x)] \\ &= (n\Delta h)^{1-\frac{1}{\alpha}} \Lambda(x) \int_{-\infty}^{\infty} (b(y) - b(x)) \frac{1}{h} K\left(\frac{y-x}{h}\right) f(y) dy \\ &= (n\Delta h)^{1-\frac{1}{\alpha}} \Lambda(x) \int_{-\infty}^{\infty} (b(x+hu) - b(x)) K(u) f(x+hu) du \\ &= (n\Delta h)^{1-\frac{1}{\alpha}} \Lambda(x) \int_{-\infty}^{\infty} \left[b'(x)hu + \frac{1}{2}b''(x+\theta hu)h^2u^2 \right] \\ & \quad \times (f(x) + f'(x+\tilde{\theta}hu)hu) K(u) du \\ &= (n\Delta h)^{1-\frac{1}{\alpha}} \Lambda(x) \left[hb'(x)f(x) \int_{-\infty}^{\infty} uK(u) du \right. \\ & \quad + h^2b'(x) \int_{-\infty}^{\infty} f'(x+\tilde{\theta}hu)u^2K(u) du \\ & \quad + \frac{1}{2}h^2f(x) \int_{-\infty}^{\infty} b''(x+\theta hu)u^2K(u) du \\ & \quad \left. + \frac{1}{2}h^3 \int_{-\infty}^{\infty} b''(x+\theta hu)f'(x+\tilde{\theta}hu)u^3K(u) du \right] \\ &= (n\Delta h)^{1-\frac{1}{\alpha}} h^2 \Lambda(x) \left[(b'(x)f'(x) + \frac{1}{2}b''(x)f(x)) \int_{-\infty}^{\infty} u^2 K(u) du + o(1) \right], \end{aligned}$$

where we have used the facts that $\int_{-\infty}^{\infty} uK(u) du = 0$ and $\int_{-\infty}^{\infty} u^3 K(u) du = 0$.

Proof of Claim 1-(ii). For $i = 1, 2, \dots, n$, set

$$\begin{aligned} \xi_{n,i}(x) &= (n\Delta h)^{1-\frac{1}{\alpha}} \left((b(X_{t_{i-1}}) - b(x))K_h(X_{t_{i-1}} - x) \right. \\ &\quad \left. - \mathbb{E}[(b(X_{t_{i-1}}) - b(x))K_h(X_{t_{i-1}} - x)] \right). \end{aligned}$$

Then, we have

$$V_{n,1}(x) - \mathbb{E}[V_{n,1}(x)] = \Lambda(x) \cdot \frac{1}{n} \sum_{i=1}^n \xi_{n,i}(x).$$

It suffices to show that $\frac{1}{n} \sum_{i=1}^n \xi_{n,i}(x) = o_P(1)$. By Lipschitz condition on $b(\cdot)$ and condition (A.5), we find

$$\sup_{1 \leq i \leq n} |\xi_{n,i}(x)| \leq LM_0(n\Delta h)^{1-\frac{1}{\alpha}} \text{ a.s.}$$

Applying Theorem 1.3 of Bosq [6], we have for each integer $q \in [1, \frac{n}{2}]$ and each $\varepsilon > 0$

$$\begin{aligned} \mathbb{P} \left(\frac{1}{n} \left| \sum_{i=1}^n \xi_{n,i}(x) \right| > \varepsilon \right) &\leq 4 \exp \left(-\frac{\varepsilon^2 q}{8v^2(q)} \right) \\ &\quad + 22 \left(1 + \frac{4LM_0(n\Delta h)^{1-\frac{1}{\alpha}}}{\varepsilon} \right)^{1/2} q\alpha_X([p]\Delta), \end{aligned}$$

where

$$v^2(q) = \frac{2}{p^2} s(q) + \frac{LM_0(n\Delta h)^{1-\frac{1}{\alpha}} \varepsilon}{2}$$

with $p = \frac{n}{2q}$ and

$$\begin{aligned} s(q) &= \max_{0 \leq j \leq 2q-1} \mathbb{E}[(j)p + 1 - jp] \xi_{n,[jp]+1}(x) + \xi_{n,[jp]+2}(x) \\ &\quad + \dots + \xi_{n,[(j+1)p]}(x) + ((j+1)p - [(j+1)p]) \xi_{n,[(j+1)p]+1}(x)^2. \end{aligned}$$

By using Billingsley's inequality (see Corollary 1.1 of Bosq [6]) and stationarity of $\xi_{n,i}(x)$, it is easy to find that

$$s(q) = O(p(n\Delta h)^{2(1-\frac{1}{\alpha})} \Delta^{-1}),$$

here we have used the fact that $\sum_{k=0}^{[p]} \alpha_X(k\Delta) = O(\Delta^{-1})$ under the GSM condition on X_t . Then, we have

$$\frac{\varepsilon^2 q}{8v^2(q)} = \frac{\varepsilon^2 n}{O((n\Delta h)^{2(1-\frac{1}{\alpha})} \Delta^{-1}) + O(\varepsilon p(n\Delta h)^{1-\frac{1}{\alpha}})},$$

which goes to ∞ by choosing $q = \lceil \sqrt{n\Delta}/\sqrt{h} \rceil$ and $p = \frac{n}{2q} = O(\sqrt{nh}/\sqrt{\Delta})$. It is also easy to see that (by GSM property of X_t again)

$$22 \left(1 + \frac{4M_0(n\Delta h)^{1-\frac{1}{\alpha}}}{\varepsilon} \right)^{1/2} q\alpha_X([p]\Delta) \rightarrow 0.$$

Therefore, we conclude that $\frac{1}{n} \sum_{i=1}^n \xi_{n,i}(x) = o_P(1)$.

Proof of Claim 2. By (4.21), we have

$$\begin{aligned} |V_{n,2}(x)| &= (n\Delta h)^{1-\frac{1}{\alpha}} \Lambda(x) |g_{n,2}(x)| \\ &\leq (n\Delta h)^{1-\frac{1}{\alpha}} \Lambda(x) \left[L\Delta e^{L\Delta} \cdot \frac{1}{n} \sum_{i=0}^{n-1} K_h(X_{t_i} - x) |b(X_{t_i})| \right. \\ &\quad \left. + L e^{L\Delta} \cdot \frac{1}{n} \sum_{i=0}^{n-1} K_h(X_{t_i} - x) \sup_{t_i \leq t \leq t_{i+1}} \left| \int_{t_i}^t \sigma(X_{s-}) dZ_s \right| \right]. \\ &= (n\Delta h)^{1-\frac{1}{\alpha}} \Lambda(x) B_{n,3}(x) + (n\Delta h)^{1-\frac{1}{\alpha}} \Lambda(x) B_{n,4}(x). \end{aligned}$$

As in the proof of Claim (ii) of Theorem 2.3 via Lemma 4.4, we can show that

$$(n\Delta h)^{1-\frac{1}{\alpha}} \Lambda(x) B_{n,3}(x) = O_P(1) \cdot O((n\Delta h)^{1-\frac{1}{\alpha}} \Delta),$$

which goes to zero in probability since $(n\Delta h)^{1-\frac{1}{\alpha}} \Delta = (n\Delta h)^{1-\frac{1}{\alpha}} \Delta^{1/\alpha} \cdot \Delta^{(\alpha-1)/\alpha} = o(1)$, and

$$P((n\Delta h)^{1-\frac{1}{\alpha}} \Lambda(x) B_{n,4}(x) > \varepsilon) \leq O((n\Delta h)^{1-\frac{1}{\alpha}} \Delta^{1/\alpha}),$$

which goes to zero under the condition $(n\Delta h)^{1-\frac{1}{\alpha}} \Delta^{1/\alpha} = o(1)$. Thus, it follows that claim 2 holds.

Proof of Claim 3. Recall that

$$\phi_n(t, x) = \sum_{i=0}^{n-1} \frac{1}{h^{1/\alpha}} K\left(\frac{X_{t_i} - x}{h}\right) \sigma(X_{t-}) 1_{(t_i, t_{i+1}]}(t).$$

and

$$\Phi_{t_n} = \left(t_n \sigma^\alpha(x) f(x) \int_{-\infty}^{\infty} K^\alpha(u) du \right)^{-\frac{1}{\alpha}}.$$

Then, we have

$$\begin{aligned} V_{n,3}(x) &= (n\Delta h)^{1-\frac{1}{\alpha}} \Lambda(x) g_{n,3}(x) \\ &= (n\Delta h)^{1-\frac{1}{\alpha}} \Lambda(x) \frac{1}{n\Delta} \sum_{i=0}^{n-1} K_h(X_{t_i} - x) \int_{t_i}^{t_{i+1}} \sigma(X_{s-}) dZ_s \\ &= f(x) \cdot \Phi_{t_n} \int_0^{t_n} \phi_n(t, x) dZ_t. \end{aligned}$$

By Lemma 4.3, it suffices to prove that (4.33) holds for $1 < \alpha < 2$. Since the proof is similar to that of (4.33) for $0 < \alpha \leq 1$, we sketch only the main steps. Recall that (4.27) gives

$$\begin{aligned} &\Phi_{t_n}^\alpha \cdot \int_0^{t_n} \phi_n^\alpha(t, x) dt \\ &= \Phi_{t_n}^\alpha \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha\left(\frac{X_{t_i} - x}{h}\right) \sigma^\alpha(X_{t_i}) \Delta \end{aligned}$$

$$\begin{aligned}
 & + \Phi_{t_n}^\alpha \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha \left(\frac{X_{t_i} - x}{h} \right) \int_{t_i}^{t_{i+1}} (\sigma^\alpha(X_{t-}) - \sigma^\alpha(X_{t_i})) dt \\
 =: & I + J.
 \end{aligned}$$

From (4.28) and (4.29), we have $I \rightarrow_P 1$ under the given conditions since (4.28) is true for any $\alpha \in (0, 2)$. Next, it remains to prove that $J \rightarrow_P 0$. Similar to (4.30), for $1 < \alpha < 2$, by using the basic inequality $\| |u + v|^p - |u|^p \| \leq |u| \cdot |v|^{p-1} + |u|^{p-1} \cdot |v| + |v|^p$ for $1 < p \leq 2$ and the Lipschitz condition on $\sigma(\cdot)$, we find

$$\begin{aligned}
 |J| & \leq \Phi_{t_n}^\alpha \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha \left(\frac{X_{t_i} - x}{h} \right) \int_{t_i}^{t_{i+1}} |\sigma^\alpha(X_{t-}) - \sigma^\alpha(X_{t_i})| dt \\
 & \leq \Phi_{t_n}^\alpha \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha \left(\frac{X_{t_i} - x}{h} \right) \int_{t_i}^{t_{i+1}} (|\sigma(X_{t-}) - \sigma(X_{t_i})|^\alpha \\
 & \quad + |\sigma(X_{t_i})| \cdot |\sigma(X_{t-}) - \sigma(X_{t_i})|^{\alpha-1} \\
 & \quad + |\sigma(X_{t_i})|^{\alpha-1} \cdot |\sigma(X_{t-}) - \sigma(X_{t_i})|) dt \\
 & \leq \Phi_{t_n}^\alpha \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha \left(\frac{X_{t_i} - x}{h} \right) L^\alpha \Delta \sup_{t_i \leq t \leq t_{i+1}} |X_t - X_{t_i}|^\alpha \\
 & \quad + \sigma_1 \Phi_{t_n}^\alpha \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha \left(\frac{X_{t_i} - x}{h} \right) L^{\alpha-1} \Delta \sup_{t_i \leq t \leq t_{i+1}} |X_t - X_{t_i}|^{\alpha-1} \\
 & \quad + \sigma_1^{\alpha-1} \Phi_{t_n}^\alpha \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha \left(\frac{X_{t_i} - x}{h} \right) L \Delta \sup_{t_i \leq t \leq t_{i+1}} |X_t - X_{t_i}| \\
 = & H_1 + H_2 + H_3.
 \end{aligned}$$

For H_1 , we can get essentially the same decomposition as (4.30) by just multiplying a constant $2^{\alpha-1}$ on the right hand side of (4.30). The same derivation from (4.31) to (4.32) yields $H_1 \rightarrow_P 0$ under the given conditions since when $1 < \alpha < 2$ we can choose $q = \frac{\alpha}{2\alpha-1}$ such that

$$nh\Delta^{\frac{q}{1-q}} = nh\Delta^{\frac{\alpha}{\alpha-1}} = [(n\Delta h)^{1-\frac{1}{\alpha}} \Delta^{1/\alpha}]^{\alpha/(\alpha-1)} = o(1).$$

For H_2 , using (4.20), we have

$$\begin{aligned}
 |H_2| & \leq \frac{\sigma_1 L^{\alpha-1} e^{(\alpha-1)L\Delta} \Delta^{\alpha-1}}{\sigma^\alpha(x) f(x) \int_{-\infty}^\infty K^\alpha(u) du} \cdot \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha \left(\frac{X_{t_i} - x}{h} \right) |b(X_{t_i})|^{\alpha-1} \\
 & \quad + \frac{\sigma_1 L^{\alpha-1} e^{(\alpha-1)L\Delta}}{\sigma^\alpha(x) f(x) \int_{-\infty}^\infty K^\alpha(u) du} \cdot \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha \left(\frac{X_{t_i} - x}{h} \right) \\
 & \quad \times \sup_{t_i \leq t \leq t_{i+1}} \left| \int_{t_i}^t \sigma(X_{s-}) dZ_s \right|^{\alpha-1} \\
 = & H_{2,1} + H_{2,2}.
 \end{aligned}$$

It is easy to see that $H_{2,1} \rightarrow_P 0$ under the given conditions (similar to Lemma 4.1). Similar to (4.22), we can use Markov inequality and Lemma 4.4 to show that

$$P(H_{2,2} > \varepsilon) \leq O(\Delta^{\frac{\alpha-1}{\alpha}}).$$

Hence, we have $H_2 \rightarrow_P 0$ under the given conditions. Similarly, we can prove that $H_3 \rightarrow_P 0$ under the given conditions. Combining the above arguments, we have $J \rightarrow_P 0$. Finally, by (4.36) and the claims 1-3, we obtain

$$V_n(x) \Rightarrow f(x)S_\alpha(1, \beta, 0) \tag{4.42}$$

under the given conditions. By using (4.35), (4.42), Slutsky's theorem and Lemma 4.1, we conclude that

$$\Lambda(x)(n\Delta h)^{1-\frac{1}{\alpha}}(\hat{b}_n(x) - b(x)) = \frac{V_n(x)}{\hat{f}_n(x)} \Rightarrow S_\alpha(1, \beta, 0).$$

This completes the proof of Theorem 2.5(a)-(i). □

Proof of Theorem 2.5(a)-(ii). If $(n\Delta h)^{1-\frac{1}{\alpha}}h^2 = O(1)$, by (4.37), we have

$$V_{n,1}(x) = (n\Delta h)^{1-\frac{1}{\alpha}}h^2\Lambda(x)\Gamma_b(x)f(x) + o_P(1).$$

By (4.38) and (4.39), we know that $V_{n,2}(x) = o_P(1)$ and $V_{n,3}(x) \Rightarrow f(x)S_\alpha(1, \beta, 0)$ under the condition $(n\Delta h)^{1-\frac{1}{\alpha}}\Delta^{\frac{1}{\alpha}} = o(1)$. Then, we use (4.36) to obtain

$$V_n(x) - (n\Delta h)^{1-\frac{1}{\alpha}}h^2\Lambda(x)\Gamma_b(x)f(x) \Rightarrow f(x)S_\alpha(1, \beta, 0).$$

By Slutsky's theorem and the fact that $f(x)/\hat{f}_n(x) \rightarrow 1$ in probability, we find

$$\begin{aligned} & (n\Delta h)^{1-\frac{1}{\alpha}}\Lambda(x)(\hat{b}_n(x) - b(x) - h^2\Gamma_b(x)) \\ &= \frac{V_n(x)}{\hat{f}_n(x)} - (n\Delta h)^{1-\frac{1}{\alpha}}h^2\Lambda(x)\Gamma_b(x) \\ &= \frac{V_n(x) - (n\Delta h)^{1-\frac{1}{\alpha}}h^2\Lambda(x)\Gamma_b(x)f(x)}{\hat{f}_n(x)} \\ & \quad + (n\Delta h)^{1-\frac{1}{\alpha}}h^2\Lambda(x)\Gamma_b(x) \left(\frac{f(x)}{\hat{f}_n(x)} - 1 \right) \\ & \Rightarrow S_\alpha(1, \beta, 0). \end{aligned}$$

This completes the proof. □

Proof of Theorem 2.5 (b). Note that Claims 1-2 in the proof of Theorem 2.5(a)-(i) are still valid for $\alpha = 1$. When $\alpha = 1$, by (4.34), it follows that

$$V_{n,3}(x) \Rightarrow f(x)S_\alpha(1, \beta, 0),$$

i.e., Claim 3 holds in this case. Here we remark that the condition $nh\Delta^{\frac{q}{1-q}} = o(1)$ for some $\frac{1}{2} < q < 1$ is required for (4.34) via the derivation of (4.32). These three claims along with Slutsky's theorem and Lemma 4.1 immediately follow that $\frac{1}{\sigma(x)}(\hat{b}_n(x) - b(x)) \Rightarrow S_\alpha(1, \beta, 0)$. This completes the proof. □

Acknowledgements

The authors are very grateful to the referees, the associate editor and the editor for their insightful and valuable comments which have greatly improved the presentation of the paper.

References

- [1] APPLEBAUM, D. (2004). *Lévy Processes and Stochastic Calculus*. Cambridge Studies in Advanced Mathematics, Vol. 93, Cambridge.
- [2] ARFI, M. (1995). Non-parametric drift estimation from ergodic samples. *J. Nonpara. Statist.* **5** 381-389. [MR1379537](#)
- [3] BANDI, F.M. and PHILLIPS, P.C.B. (2003). Fully nonparametric estimation of scalar diffusion models. *Econometrica* **71** 241-283. [MR1956859](#)
- [4] BANON, G. (1978). Nonparametric identification for diffusion processes. *SIAM J. Control Optim.* **16** 380-395. [MR0492159](#)
- [5] BILLINGSLEY, P. (1995). *Probability and Measure*. Third Edition, Wiley, New York. [MR1324786](#)
- [6] BOSQ, D. (1996). *Nonparametric Statistics for Stochastic Processes*. Lecture Notes in Statistics, Vol. 110, Springer-Verlag, New York. [MR1441072](#)
- [7] DALALYAN, A. (2005). Sharp adaptive estimation of the drift function for ergodic diffusions. *Ann. Statist.* **33** 2507-2528. [MR2253093](#)
- [8] DALALYAN, A. and KUTOYANTS, YU.A. (2002). Asymptotically efficient trend coefficient estimation for ergodic diffusion. *Math. Methods Statist.* **11** 402-427. [MR1979742](#)
- [9] DALALYAN, A. and KUTOYANTS, YU.A. (2003). Asymptotically efficient estimation of the derivative of invariant density. *Stat. Inference Stoch. Process.* **6** 89-107. [MR1965186](#)
- [10] DITLEVSEN, P.D. (1999). Observation of α -stable noise induced millennial climate changes from an ice-core record. *Geophysical Research Letter* **26** 1441-1444.
- [11] DITLEVSEN, P.D. (1999). Anomalous jumping in a double-well potential. *Phys. Rev. E* **60** 172-179.
- [12] FAN, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.* **87** 998-1004. [MR1209561](#)
- [13] FAN, J. (2005). A selection overview of nonparametric methods in financial econometrics. *Statist. Sci.* **20** 317-337. [MR2210224](#)
- [14] FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London. [MR1383587](#)
- [15] FAN, J. and ZHANG, C. (2003). A reexamination of diffusion estimators with applications to financial model validation. *J. Amer. Statist. Assoc.* **98** 118-134. [MR1965679](#)
- [16] FRISTEDT, B. (1974). Sample functions of stochastic processes with stationary, independent increments. *Advances in Probability and Related Topics*, Vol. 3, Eds. P. Ney and S. Port, pp. 241-396, Marcel Dekker, New York. [MR0400406](#)

- [17] GOBET, E., HOFFMANN, M. and REISS, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.* **32** 2223-2253. [MR2102509](#)
- [18] HALL, P., PENG, L. and YAO, Q. (2002). Prediction and nonparametric estimation for time series with heavy tails. *J. Time Ser. Anal.* **23** 313-331. [MR1908594](#)
- [19] HÄRDLE, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press, Cambridge. [MR1161622](#)
- [20] HOFFMANN, M. (1999). Adaptive estimation in diffusion processes. *Stochastic Process. Appl.* **79** 135-163. [MR1670522](#)
- [21] HU, Y. and LONG, H. (2007). Parameter estimation for Ornstein-Uhlenbeck processes driven by α -stable Lévy motions. *Communications on Stochastic Analysis* **1** 175-192. [MR2397392](#)
- [22] HU, Y. and LONG, H. (2009). Least square estimator for Ornstein-Uhlenbeck processes driven by α -stable motions. *Stochastic Process. Appl.* **119** 2465-2480. [MR2532208](#)
- [23] JACOD, J. (2004). The Euler scheme for Lévy driven stochastic differential equations: limit theorems. *Ann. Probab.* **32** 1830-1872. [MR2073179](#)
- [24] JACOD, J., KURTZ, T.G., MÉLÉARD, S. and PROTTER, P. (2005). The approximate Euler method for Lévy driven stochastic differential equations. *Ann. Inst. H. Poincaré Probab. Statist.* **41** 523-558. [MR2139032](#)
- [25] JANICKI, A. and WERON, A. (1994). *Simulation and Chaotic Behavior of α -stable Stochastic Processes*. Marcel Dekker, New York. [MR1306279](#)
- [26] JOULIN, A. (2007). On maximal inequalities for stable stochastic integrals. *Potential Anal.* **26** 57-78. [MR2276525](#)
- [27] KALLENBERG, O. (1992). Some time change representations of stable integrals, via predictable transformations of local martingales. *Stochastic Process. Appl.* **40** 199-223. [MR1158024](#)
- [28] KOUTROUVELIS, I. (1980). Regression-type estimation of the parameters of stable laws. *J. Am. Stat. Assoc.* **69** 108-113. [MR0600977](#)
- [29] KOUTROUVELIS, I. (1981). An iterative procedure for the estimation of the parameters of stable laws. *Commun. Stat. Simul. Comput.* **10** 17-28. [MR0608513](#)
- [30] KUTOYANTS, YU. A. (2004). *Statistical Inference for Ergodic Diffusion Processes*. Springer-Verlag, London, Berlin, Heidelberg. [MR2144185](#)
- [31] KYPRIANOU, A., SCHOUTENS, W. and WILMOTT, P. (Eds.) (2005). *Exotic Option Pricing and Advanced Lévy Models*. Wiley, Chichester, UK. [MR2343205](#)
- [32] LONG, H. (2009). Least squares estimator for discretely observed Ornstein-Uhlenbeck processes with small Lévy noises. *Statistics and Probability Letters* **79** 2076-2085. [MR2571771](#)
- [33] LONG, H. (2010). Parameter estimation for a class of stochastic differential equations driven by small stable noises from discrete observations. *Acta Mathematica Scientia* **30B** 645-663. [MR2675775](#)

- [34] MA, C. (2010). A note on “Least squares estimator for discretely observed Ornstein-Uhlenbeck processes with small Lévy noises”. *Statistics and Probability Letters* **80** 1528-1531. [MR2669756](#)
- [35] MASUDA, H. (2007). Ergodicity and exponential β -mixing bounds for multidimensional diffusions with jumps. *Stochastic Process. Appl.* **117** 35-56. [MR2287102](#)
- [36] MASUDA, H. (2010). Approximate self-weighted LAD estimation of discretely observed ergodic Ornstein-Uhlenbeck processes. *Electronic Journal of Statistics* **4** 525-565. [MR2660532](#)
- [37] MIKOSCH, T., RESNICK, S., ROOTZEN, H. and STEGEMAN, A. (2002). Is network traffic approximated stable Lévy motion or fractional Brownian motion? *Ann. Appl. Probab.* **12** 23-68. [MR1890056](#)
- [38] NADARAYA, E.A. (1964). On estimating regression. *Theory Probab. Appl.* **9** 141-142.
- [39] PENG, L. and YAO, Q. (2004). Nonparametric regression under dependent errors with infinite variance. *Ann. Inst. Statist. Math.* **56** 73-86. [MR2053729](#)
- [40] PHAM, D.T. (1981). Nonparametric estimation of the drift coefficient in the diffusion equation. *Math. Operationsforsch. Statist. Ser. Statist.* **12** 61-73. [MR0607955](#)
- [41] PRAKASA RAO, B.L.S. (1985). Estimation of the drift for diffusion process. *Statistics* **16** 263-275. [MR0789070](#)
- [42] PRESS, S.J. (1972). Estimation in univariate and multivariate stable distributions. *J. Am. Stat. Assoc.* **67** 842-846. [MR0362666](#)
- [43] PROTTER, P. (1990). *Stochastic Integration and Differential Equations*. Springer-Verlag, Berlin, Heidelberg, New York. [MR1037262](#)
- [44] ROSINSKI, J. and WOYCZYNSKI, W.A. (1985). Moment inequalities for real and vector p -stable stochastic integrals. *Probability in Banach Spaces V*, Lecture Notes in Math., **1153**, pp. 369-386, Springer, Berlin. [MR0821992](#)
- [45] ROSINSKI, J. and WOYCZYNSKI, W.A. (1986). On Ito stochastic integration with respect to p -stable motion: inner clock, integrability of sample paths, double and multiple integrals. *Ann. Probab.* **14** 271-286. [MR0815970](#)
- [46] SAMORODNITSKY, G. and TAQQU, M.S. (1994). *Stable non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall, New York, London. [MR1280932](#)
- [47] SATO, K.I. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge. [MR1739520](#)
- [48] SCHERTZER, D., LARCHEVEQUE, M., DUAN, J., YANOVSKY, V.V. and LOVEJOY, S. (2001). Fractional Fokker-Planck equation for nonlinear stochastic differential equations driven by non-Gaussian Lévy stable noises. *J. Math. Phys.* **42** 200-212. [MR1808774](#)
- [49] SCHOUTENS, W. (2003). *Lévy Processes in Finance: Pricing Financial Derivatives*. Wiley, Hoboken, NJ, USA.
- [50] SPOKOINY, V.G. (2000). Adaptive drift estimation for nonparametric diffusion model. *Ann. Statist.* **28** 815-836. [MR1792788](#)

- [51] VAN ZANTEN, J.H. (2000). Rates of convergence and asymptotic normality of curve estimators for ergodic diffusion processes. Report PNA-R0011, CWI, Amsterdam.
- [52] WATSON, G.S. (1964). Smooth regression analysis. *Sankhya, Ser. A.* **26** 359-372. [MR0185765](#)
- [53] WU, W.B. (2003). Nonparametric estimation for stationary processes. Technical Report No. 536, Department of Statistics, The University of Chicago.
- [54] WU, W.B. (2005). Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150-14154. [MR2172215](#)