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

Hongwei Long[1] 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 classification. 60G52, 62G20, 62M05, 65C30

Keywords: central limit theorem; consistency; geometrically strong mixing; kernel estimator; Lévy motion; Nadaraya-Watson estimator; rate of convergence; stable stochastic integrals

1 Introduction

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

$$dX_t = b(X_{t-})dt + \sigma(X_{t-})dZ_t, X_0 = \eta, \tag{1}$$

where $b : \mathbb{R} \to \mathbb{R}$ is an unknown measurable function, $\sigma : \mathbb{R} \to \mathbb{R}_+$ 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 $(\Omega, \mathcal{F}, P)$ 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 [38], Janicki and Weron [25], and Sato [39]. We assume that the stochastic process $\{X_t\}$ is observed at discrete time points $\{t_i = i\Delta, i =$

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

$0, 1, \ldots, n\}$, 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 [41], Kyprianou, Schoutens and Wilmott [29]), network traffic (see Mikosch et al. [31]), physics (see, e.g. Schertzer et al. [40]), and climate dynamics (see Ditlevsen [10], [11]). The existence and uniqueness of solutions to (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). Most recently, Masuda [30] 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) 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]. 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) using nonparametric smoothing approach. Nonparametric smooth approach is a data driven method and has many benefits. It 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 [32] and Waston [44]. 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 (α, β and ϕ)-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 [45] and [46].

Many authors have investigated nonparametric estimation for the drift function b in the setting of diffusions driven by Brownian motions. Pham [34] and Prakasa Rao [35] 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 [43], Dalalyan and Kutoyants [8], [9], Dalalyan [7] (see also the monograph by Kutoyants [28] and references therein). Locally linear (or polynomial) estimators have been proposed by Fan [12] and further discussed in Fan and Gijbels [14], Spokoiny [42] 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 [33] 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. For convenience, we shall discuss the Nadaraya-Waston 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 $Y_i := X_{t_{i+1}} - X_{t_i}$, $\Delta = t_{i+1} - t_i, i = 0, 1, \ldots, n-1$. The weight function is given by

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

where $K_h(\cdot) = K(\cdot/h)/h$, K is a kernel density function (with compact support) 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

$$\hat{b}_n(x) = \frac{\sum_{i=0}^{n-1} Y_i K_h(X_{t_i} - x)}{\Delta \sum_{i=0}^{n-1} K_h(X_i - x)}. \tag{2}$$

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. $\Delta \to 0$) in this paper.

The paper is organized as follows. In Section 2, we obtain consistency ($1 < \alpha < 2$) or inconsistency ($0 < \alpha \leq 1$) of the N-W estimator $\hat{b}_n(x)$. A central limit theorem for stable stochastic integrals is also established in Section 2, which is the crucial tool for proofs presented in Sections 2 and 3. Then, in Section 3 we derive the rate of convergence and pointwise asymptotic distribution of $\hat{b}_n(x)$. Finally, in Section 4, we conduct a simulation study to confirm the finite sample property. 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 "$\to_P$" to denote "convergence in probability" and notation "$\Rightarrow$" to denote "convergence in distribution".

2 Consistency of the Nadaraya-Watson Estimator

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

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=0}^{n-1} K_h(X_{t_i} - x). \tag{3}$$

Denote

$$\hat{g}_n(x) = \frac{1}{n\Delta} \sum_{i=0}^{n-1} Y_i K_h(X_{t_i} - x). \tag{4}$$

Then, the N-W estimator given in (2) can be represented as

$$\hat{b}_n(x) = \hat{g}_n(x)/\hat{f}_n(x). \tag{5}$$

Define the strong mixing coefficient of X 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_s and X_{s+t}, respectively.

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

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

(A.2). The dispersion function $\sigma(\cdot)$ satisfies the following bounded condition: there exist positive constants σ_0 and σ_1 such that $0 < \sigma_0 \le \sigma(x) \le \sigma_1$ 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 $\alpha_X(t) \le c_0\rho^t,\ t \ge 0$. Consequently, X_t is ergodic and 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 (with compact support) satisfying

$$\int_{-\infty}^{\infty} u^2 K(u) du < \infty,\ \int_{-\infty}^{\infty} K^2(u) du < \infty.$$

(A.6). As $n \to \infty$, $h \to 0$, $\Delta \to 0$ and $n\Delta h \to \infty$.

Our main results of this section are stated in the following theorems. Basically we discuss 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 2.1. *Assume that (A.1)-(A.6) hold. If $f(x) > 0$ and $\alpha \in (1,2)$, then $\hat{b}_n(x) \to_P b(x)$ as $n \to \infty$.*

Theorem 2.2. *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 $nh\Delta^{\frac{r}{\alpha(1-r)}} \to 0$, then $\hat{b}_n(x)$ is not consistent.*

Remark 2.3. The Lipschitz condition (A.1) is a typical condition which ensures that SDE (1) admits a unique non-explosive cádlág adpated solution. For some sufficient conditions which guarantee (A.3), we refer to Theorem 2.2 and Lemma 2.4 in Masuda [30].

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

Lemma 2.4. *Under the conditions (A.1)-(A.6), we have*

$$\hat{f}_n(x) \to_P f(x) \quad \text{as } n \to \infty. \tag{6}$$

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 $n \to \infty$ by Lebesgue dominated convergence theorem. Thus, it suffices to prove that $\hat{f}_n(x) - \mathbb{E}[\hat{f}_n(x)] \to_P 0$. 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 $\eta_{n,i}(x) = K_h(X_{t_{i-1}} - x) - \mathbb{E}K_h(X_{t_{i-1}} - x), i = 1, 2, \cdots, n$. Note that $\sup_{1 \le i \le n} |\eta_{n,i}(x)| \le C_0 h^{-1}$ *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 $\varepsilon > 0$

$$\mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n}\eta_{n,i}(x)\right| > \varepsilon\right) \leq 4\exp\left(-\frac{\varepsilon^2 q}{8v^2(q)}\right) + 22\left(1+\frac{4C_0h^{-1}}{\varepsilon}\right)^{1/2} q\alpha_X\left([p]\Delta\right), \tag{7}$$

where

$$v^2(q) = \frac{2}{p^2}s(q) + \frac{C_0h^{-1}\varepsilon}{2}$$

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

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

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 $s(q) = O(p^2h^{-1})$. By choosing $q = [\sqrt{n\Delta}/\sqrt{h}]$ and $p = \frac{n}{2q} = O(\sqrt{nh}/\sqrt{\Delta})$, we obtain

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

By the GSM property of X_t and some basic calculations, we find

$$22\left(1+\frac{4C_0h^{-1}}{\varepsilon}\right)^{1/2} q\alpha_X\left([p]\Delta\right) \leq C(\varepsilon)\exp(-O(\sqrt{n\Delta h})). \tag{9}$$

Combining (7), (8) and (9), we have

$$\mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n}\eta_{n,i}(x)\right| > \varepsilon\right) \leq C(\varepsilon)\exp\{-O(\varepsilon^2\sqrt{n\Delta h})\}. \tag{10}$$

Therefore, the desired convergence result (6) follows from given conditions. □

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 [37], Kallenberg [27]). We assume that either $\phi(t)$ is nonnegative or Z is symmetric. Then, we have the following version of Lenglart's inequality in the stable setting.

Lemma 2.5. *For any given $\varepsilon > 0$ and $\delta > 0$, there is some constant $c > 0$ such that*

$$P\left(\sup_{0\le t\le T}\left|\int_0^t \phi(s)dZ_s\right| > \varepsilon\right) \le \frac{c\delta}{\varepsilon^\alpha} + P\left(\int_0^T |\phi(t)|^\alpha dt > \delta\right). \tag{11}$$

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 [37] 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 $c > 0$

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

This completes the proof. □

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

Lemma 2.6. *Suppose that there is a deterministic and nonnegative function Φ such that*

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

Then, we have

$$\Phi(T)\int_0^T \phi(t)dZ_t \Rightarrow S_\alpha(1, \beta, 0). \tag{12}$$

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 $\tau_T = \inf\{t \ge 0 : R_t > 1\}$. 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 2.5 and following exactly the same arguments as in the proof of Theorem 1.19 in Kutoyants [28], 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 $T \to \infty$. Therefore, by the continuity theorem (see Theorem 26.3 of Billingsley [5]), it immediately follows that (12) holds. □

We say that a continuous function $F : [0,\infty) \to [0,\infty)$ grows more slowly than u^α ($\alpha > 0$) if there exist positive constants c, λ_0 and $\alpha_0 < \alpha$ such that $F(\lambda u) \le c\lambda^{\alpha_0} F(u)$ for all $u > 0$ and all $\lambda \ge \lambda_0$. Now we state the moment inequalities for stable stochastic integrals in the following lemma, which can be regarded as a generalization of Theorem 3.2 of Rosinski and Woyczynski [36] where the symmetric case is dealt with. This lemma will be a crucial tool in the proofs of our main results.

Lemma 2.7. *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 $T > 0$*

$$c_1 \mathbb{E}[F((\int_0^T |\phi(t)|^\alpha dt)^{1/\alpha})] \le \mathbb{E}[F(\sup_{t\le T} |\int_0^t \phi(s) dZ_s|)] \le c_2 \mathbb{E}[F((\int_0^T |\phi(t)|^\alpha dt)^{1/\alpha})]. \quad (13)$$

Proof. When Z is symmetric α-stable, moment inequalities (13) for stable stochastic integrals have been established in Theorem 3.1 and Theorem 3.2 of Rosinski and Woyczynski [36]. We claim that the moment inequalities in Theorem 3.1 and Theorem 3.2 of [36] are still true for non-symmetric (strictly) stable Lévy processes and stable stochastic integrals. In this case, we assume that the integrand process $\phi(\cdot)$ is non-negative and predictable so that the random time change property (or inner clock property) of stable stochastic integrals is applicable (see Kallenberg [27]). In the proof of Theorem 3.1 of [36], there is only one place in the probability estimate of part I where the authors have used the symmetric property via Lévy inequality. However, we can replace this estimate by the following probability estimate with some constant $C > 0$

$$\sup_{\lambda > 0} \lambda^\alpha P\left(\sup_{0\le s\le 1} |Z_s| \ge \lambda\right) \le C,$$

provided in Proposition 10.2 of Fristedt [16] (see also Joulin [26]). All the remaining arguments in the proof of Theorem 3.1 of [36] work throughout. Consequently, the moment inequalities in Theorem 3.2 of [36] are also true for non-symmetric case as stated in (13) when the integrand process is non-negative, predictable and L^α-integrable. □

Now we are in a position to prove Theorem 2.1.

Proof of Theorem 2.1. By (5) and Lemma 2.4, it suffices to prove that

$$\hat{g}_n(x) \to_P f(x)b(x) \text{ as } \quad n \to \infty. \quad (14)$$

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), it follows that

$$\begin{aligned} \hat{g}_n(x) &= \frac{1}{n}\sum_{i=0}^{n-1} b(X_{t_i})K_h(X_{t_i} - x) \\ &\quad + \frac{1}{n\Delta}\sum_{i=0}^{n-1} K_h(X_{t_i} - x)\int_{t_i}^{t_{i+1}} (b(X_{s-}) - b(X_{t_i}))ds \\ &\quad + \frac{1}{n\Delta}\sum_{i=0}^{n-1} K_h(X_{t_i} - x)\int_{t_i}^{t_{i+1}} \sigma(X_{s-})dZ_s \\ &=: g_{n,1}(x) + g_{n,2}(x) + g_{n,3}(x). \end{aligned} \tag{15}$$

We have the following claims:
(i) $g_{n,1}(x) \to_P f(x)b(x)$ as $n \to \infty$;
(ii) $g_{n,2}(x) \to_P 0$ as $n \to \infty$;
(iii) $g_{n,3}(x) \to_P 0$ as $n \to \infty$.
These three claims guarantee that (14) holds.

Proof of Claim (i). Note that

$$\begin{aligned} g_{n,1}(x) &= b(x)\frac{1}{n}\sum_{i=0}^{n-1} K_h(X_{t_i} - x) + \frac{1}{n}\sum_{i=0}^{n-1}(b(X_{t_i}) - b(x))K_h(X_{t_i} - x) \\ &=: B_{n,1}(x) + B_{n,2}(x). \end{aligned} \tag{16}$$

By Lemma 2.4, it is clear that $B_{n,1}(x) \to_P b(x)f(x)$ when $n \to \infty$. For $B_{n,2}(x)$, by the Lipschitz property of $b(\cdot)$ and stationarity of X_t, we have

$$\begin{aligned} |B_{n,2}(x)| &\leq \frac{1}{n}\sum_{i=0}^{n-1} L|X_{t_i} - x|K_h(X_{t_i} - x) \\ &\leq L\frac{1}{n}\sum_{i=0}^{n-1} (|X_{t_i} - x|K_h(X_{t_i} - x) - \mathbb{E}[|X_{t_i} - x|K_h(X_{t_i} - x)]) \\ &\quad + L \cdot \mathbb{E}[|X_0 - x|K_h(X_0 - x)]. \end{aligned} \tag{17}$$

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 (since $K(\cdot)$ has a compact and bounded support). By slightly modifying the proof of Lemma 2.4, we can show that

$$L\frac{1}{n}\sum_{i=0}^{n-1} (|X_{t_i} - x|K_h(X_{t_i} - x) - \mathbb{E}[|X_{t_i} - x|K_h(X_{t_i} - x)]) \to_P 0 \tag{18}$$

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

$$\begin{aligned}\lim_{h\to 0}\frac{\mathbb{E}[|X_0-x|K_h(X_0-x)]}{h} &= \lim_{h\to 0}\frac{1}{h}\int_{-\infty}^{\infty}|y-x|K_h(y-x)f(y)dy \\ &= \lim_{h\to 0}\int_{-\infty}^{\infty}|u|K(u)f(x+uh)du \\ &= f(x)\int_{-\infty}^{\infty}|u|K(u)du. \end{aligned} \tag{19}$$

Combining (17), (18) and (19), it follows that $B_{n,2}(x) \to_P 0$ when $n \to \infty$. Hence the claim (i) holds. $\square$

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

$$\begin{aligned}|g_{n,2}(x)| &\le \frac{1}{n\Delta}\sum_{i=0}^{n-1}K_h(X_{t_i}-x)\int_{t_i}^{t_{i+1}}|b(X_{s-})-b(X_{t_i})|ds \\ &\le \frac{L}{n\Delta}\sum_{i=0}^{n-1}K_h(X_{t_i}-x)\int_{t_i}^{t_{i+1}}|X_{s-}-X_{t_i}|ds \\ &\le \frac{L}{n}\sum_{i=0}^{n-1}K_h(X_{t_i}-x)\sup_{t_i\le t\le t_{i+1}}|X_t-X_{t_i}|. \end{aligned} \tag{20}$$

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

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

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

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

By Gronwall's inequality, we have

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

It follows that

$$\sup_{t_i\le t\le t_{i+1}}|X_t-X_{t_i}| \le e^{L\Delta}\left(|b(X_{t_i})|\Delta + \sup_{t_i\le t\le t_{i+1}}\left|\int_{t_i}^{t}\sigma(X_{s-})dZ_s\right|\right). \tag{21}$$

By (20) and (21), we find

$$
\begin{aligned}
|g_{n,2}(x)| &\leq \frac{L}{n}\sum_{i=0}^{n-1} K_h(X_{t_i}-x)e^{L\Delta}\left(|b(X_{t_i})|\Delta + \sup_{t_i\leq t\leq t_{i+1}}\left|\int_{t_i}^{t}\sigma(X_{s-})dZ_s\right|\right) \\
&\leq L\Delta e^{L\Delta}\cdot\frac{1}{n}\sum_{i=0}^{n-1} K_h(X_{t_i}-x)|b(X_{t_i})| \\
&\quad + Le^{L\Delta}\cdot\frac{1}{n}\sum_{i=0}^{n-1} K_h(X_{t_i}-x)\sup_{t_i\leq t\leq t_{i+1}}\left|\int_{t_i}^{t}\sigma(X_{s-})dZ_s\right| \\
&=: B_{n,3}(x)+B_{n,4}(x). \qquad (22)
\end{aligned}
$$

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

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

when $n\to\infty$. This implies that $B_{n,3}(x)\to_P 0$ when $n\to\infty$. By Markov inequality and Lemma 2.7, we have

$$
\begin{aligned}
& P\left(\frac{1}{n}\sum_{i=0}^{n-1} K_h(X_{t_i}-x)\sup_{t_i\leq t\leq t_{i+1}}\left|\int_{t_i}^{t}\sigma(X_{s-})dZ_s\right| > \varepsilon\right) \\
\leq\ & \frac{1}{n\varepsilon}\sum_{i=0}^{n-1}\mathbb{E}\left[\sup_{t_i\leq t\leq t_{i+1}}\left|\int_{t_i}^{t} K_h(X_{t_i}-x)\sigma(X_{s-})dZ_s\right|\right] \\
\leq\ & \frac{C_1}{n\varepsilon}\sum_{i=0}^{n-1}\mathbb{E}\left[\left(\int_{t_i}^{t_{i+1}} K_h^{\alpha}(X_{t_i}-x)\sigma^{\alpha}(X_{s-})ds\right)^{1/\alpha}\right] \\
\leq\ & \frac{C_1}{n\varepsilon}\sum_{i=0}^{n-1}\mathbb{E}[K_h(X_{t_i}-x)\sigma_1\Delta^{1/\alpha}] \\
\leq\ & O(\Delta^{1/\alpha}). \qquad (23)
\end{aligned}
$$

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

Proof of Claim (iii). We define

$$
\phi_n(t,x) = \sum_{i=0}^{n-1}\frac{1}{h^{1/\alpha}}K\left(\frac{X_{t_i}-x}{h}\right)\sigma(X_{t-})1_{(t_i,t_{i+1}]}(t), \qquad (24)
$$

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 2.7, boundedness of $\sigma(X_{t-})$, and stationarity of X_t, we find for some constant $C_2 > 0$

$$\begin{aligned}
& P(|g_{n,3}(x)| > \varepsilon) \\
\leq\ & \frac{1}{n\Delta h^{\frac{\alpha-1}{\alpha}}\varepsilon}\mathbb{E}\left|\int_0^{t_n}\phi_n(t,x)dZ_t\right| \\
\leq\ & \frac{C_2}{n\Delta h^{\frac{\alpha-1}{\alpha}}\varepsilon}\mathbb{E}\left[\left(\int_0^{t_n}|\phi_n(t,x)|^\alpha dt\right)^{1/\alpha}\right] \\
\leq\ & \frac{C_2}{n\Delta h^{\frac{\alpha-1}{\alpha}}\varepsilon}\left(\mathbb{E}\left[\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\frac{1}{h}K^\alpha\left(\frac{X_{t_i}-x}{h}\right)\sigma^\alpha(X_{t-})dt\right]\right)^{1/\alpha} \\
\leq\ & \frac{C_2}{n\Delta h^{\frac{\alpha-1}{\alpha}}\varepsilon}\left(n\Delta\sigma_1^\alpha\mathbb{E}\left[\frac{1}{h}K^\alpha\left(\frac{X_0-x}{h}\right)\right]\right)^{1/\alpha} \\
\leq\ & \frac{C_2\sigma_1}{n\Delta h^{\frac{\alpha-1}{\alpha}}\varepsilon}(n\Delta)^{\frac{1}{\alpha}}O(1) \\
=\ & O((n\Delta h)^{\frac{1-\alpha}{\alpha}}),
\end{aligned} \tag{25}$$

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

Proof of Theorem 2.2. The claims (i) in the proof of Theorem 2.1 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 $B_{n,3}(x) \to_P 0$ under (A.1)-(A.6). For $B_{n,4}(x)$, by Markov inequality, Lemma 2.7, condition (A.2), and stationarity of X_t, we have for $\frac{\alpha}{1+\alpha} < r < \alpha \leq 1$

$$\begin{aligned}
& P\left(\frac{1}{n}\sum_{i=0}^{n-1}K_h(X_{t_i}-x)\sup_{t_i\leq t\leq t_{i+1}}\left|\int_{t_i}^t\sigma(X_{s-})dZ_s\right| > \varepsilon\right) \\
\leq\ & \frac{1}{n^r\varepsilon^r}\sum_{i=0}^{n-1}\mathbb{E}\left[\sup_{t_i\leq t\leq t_{i+1}}\left|\int_{t_i}^t K_h(X_{t_i}-x)\sigma(X_{s-})dZ_s\right|^r\right] \\
\leq\ & \frac{C_1}{n^r\varepsilon^r}\sum_{i=0}^{n-1}\mathbb{E}\left[\left(\int_{t_i}^{t_{i+1}}K_h^\alpha(X_{t_i}-x)\sigma^\alpha(X_{s-})ds\right)^{r/\alpha}\right] \\
\leq\ & \frac{C_1}{n^r\varepsilon^r}\sum_{i=0}^{n-1}\mathbb{E}[K_h^r(X_{t_i}-x)\sigma_1^r\Delta^{r/\alpha}] \\
\leq\ & O((nh\Delta^{\frac{r}{\alpha(1-r)}})^{1-r}).
\end{aligned} \tag{26}$$

which tends to zero when $nh\Delta^{\frac{r}{\alpha(1-r)}} \to 0$. 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 $n \to \infty$ 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

$$\Phi_{t_n} \int_0^{t_n} \phi_n(t,x) dZ_t = \frac{(n\Delta h)^{1-\frac{1}{\alpha}}}{\sigma(x) f(x)^{\frac{1}{\alpha}} \left(\int_{-\infty}^{\infty} K^\alpha(u)du\right)^{\frac{1}{\alpha}}} g_{n,3}(x). \tag{27}$$

Note that

$$\begin{aligned} \Phi_{t_n}^\alpha \cdot \int_0^{t_n} \phi_n^\alpha(t,x) dt &= \Phi_{t_n}^\alpha \cdot \int_0^{t_n} \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha\left(\frac{X_{t_i}-x}{h}\right) \sigma^\alpha(X_{t-}) 1_{(t_i,t_{i+1}]}(t) dt \\ &= \Phi_{t_n}^\alpha \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha\left(\frac{X_{t_i}-x}{h}\right) \int_{t_i}^{t_{i+1}} \sigma^\alpha(X_{t-}) dt \\ &= \Phi_{t_n}^\alpha \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha\left(\frac{X_{t_i}-x}{h}\right) \sigma^\alpha(X_{t_i}) \Delta \\ &\quad + \Phi_{t_n}^\alpha \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha\left(\frac{X_{t_i}-x}{h}\right) \int_{t_i}^{t_{i+1}} (\sigma^\alpha(X_{t-}) - \sigma^\alpha(X_{t_i})) dt \\ &=: I + J. \end{aligned} \tag{28}$$

Similar to Lemma 2.4, it is easy to prove that

$$\frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha\left(\frac{X_{t_i}-x}{h}\right) \sigma^\alpha(X_{t_i}) \to_P \sigma^\alpha(x) f(x) \int_{-\infty}^{\infty} K^\alpha(u) du. \tag{29}$$

Therefore, we have

$$I = \frac{1}{\sigma^\alpha(x) f(x) \int_{-\infty}^{\infty} K^\alpha(u)du} \cdot \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{h} K^\alpha\left(\frac{X_{t_i}-x}{h}\right) \sigma^\alpha(X_{t_i}) \to_P 1. \tag{30}$$

Next we deal with the second term J. By Lipschitz condition (A.1) on $\sigma(\cdot)$, (21) and basic inequality $||u+v|^q - |v|^q| \le |u|^q$ for $u, v \in \mathbb{R}$ and $q \in (0,1]$, we have

$$\begin{aligned}
|J| &= \Phi_{t_n}^{\alpha} \left| \sum_{i=0}^{n-1} \frac{1}{h} K^{\alpha}\left(\frac{X_{t_i}-x}{h}\right) \int_{t_i}^{t_{i+1}} (\sigma^{\alpha}(X_{t-}) - \sigma^{\alpha}(X_{t_i}))dt \right| \\
&\le \Phi_{t_n}^{\alpha} \sum_{i=0}^{n-1} \frac{1}{h} K^{\alpha}\left(\frac{X_{t_i}-x}{h}\right) \int_{t_i}^{t_{i+1}} \left|\sigma^{\alpha}(X_{t-}) - \sigma^{\alpha}(X_{t_i})\right| dt \\
&\le \Phi_{t_n}^{\alpha} \sum_{i=0}^{n-1} \frac{1}{h} K^{\alpha}\left(\frac{X_{t_i}-x}{h}\right) \int_{t_i}^{t_{i+1}} \left|\sigma(X_{t-}) - \sigma(X_{t_i})\right|^{\alpha} dt \\
&\le \Phi_{t_n}^{\alpha} \sum_{i=0}^{n-1} \frac{1}{h} K^{\alpha}\left(\frac{X_{t_i}-x}{h}\right) L^{\alpha}\Delta \sup_{t_i \le t \le t_{i+1}} |X_t - X_{t_i}|^{\alpha} \\
&\le \frac{L^{\alpha} e^{\alpha L\Delta}\Delta^{\alpha}}{\sigma^{\alpha}(x) f(x) \int_{-\infty}^{\infty} K^{\alpha}(u)du} \cdot \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{h} K^{\alpha}\left(\frac{X_{t_i}-x}{h}\right) |b(X_{t_i})|^{\alpha} \\
&\quad + \frac{L^{\alpha} e^{\alpha L\Delta}}{\sigma^{\alpha}(x) f(x) \int_{-\infty}^{\infty} K^{\alpha}(u)du} \cdot \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{h} K^{\alpha}\left(\frac{X_{t_i}-x}{h}\right) \sup_{t_i \le t \le t_{i+1}} \left| \int_{t_i}^{t} \sigma(X_{s-}) dZ_s \right|^{\alpha} \\
&= J_1 + J_2. \qquad (31)
\end{aligned}$$

It is clear that $J_1 \to_P 0$ under given conditions since (similar to Lemma 2.4)

$$\frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{h} K^{\alpha}\left(\frac{X_{t_i}-x}{h}\right) |b(X_{t_i})|^{\alpha} \to_P |b(x)|^{\alpha} f(x) \int_{-\infty}^{\infty} K^{\alpha}(u)du. \qquad (32)$$

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

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

By using Markov inequality, Lemma 2.7, and condition (A.2), we have for $q < 1$

$$
\begin{aligned}
& P\left(\frac{1}{n}\sum_{i=0}^{n-1}\frac{1}{h}K^{\alpha}\left(\frac{X_{t_i}-x}{h}\right)\sup_{t_i\le t\le t_{i+1}}\left|\int_{t_i}^{t}\sigma(X_{s-})dZ_s\right|^{\alpha} > \varepsilon\right) \\
\le\ & \frac{1}{(nh\varepsilon)^q}\mathbb{E}\left|\sum_{i=0}^{n-1}K^{\alpha}\left(\frac{X_{t_i}-x}{h}\right)\sup_{t_i\le t\le t_{i+1}}\left|\int_{t_i}^{t}\sigma(X_{s-})dZ_s\right|^{\alpha}\right|^{q} \\
\le\ & \frac{1}{(nh\varepsilon)^q}\sum_{i=0}^{n-1}\mathbb{E}\left[\left(\sup_{t_i\le t\le t_{i+1}}\left|\int_{t_i}^{t}K\left(\frac{X_{t_i}-x}{h}\right)\sigma(X_{s-})dZ_s\right|\right)^{q\alpha}\right] \\
\le\ & \frac{C}{(nh\varepsilon)^q}\sum_{i=0}^{n-1}\mathbb{E}\left[\left(\int_{t_i}^{t_{i+1}}K^{\alpha}\left(\frac{X_{t_i}-x}{h}\right)\sigma^{\alpha}(X_{s-})ds\right)^{q}\right] \\
\le\ & \frac{C}{(nh\varepsilon)^q}\sum_{i=0}^{n-1}\mathbb{E}\left[K^{q\alpha}\left(\frac{X_{t_i}-x}{h}\right)\sigma_1^{q\alpha}\Delta^q\right] \\
=\ & \frac{nC\sigma_1^{q\alpha}\Delta^q}{(nh\varepsilon)^q}\int_{-\infty}^{\infty}K^{q\alpha}\left(\frac{y-x}{h}\right)f(y)dy \\
=\ & O((nh\Delta^{\frac{q}{1-q}})^{1-q}), \qquad (33)
\end{aligned}
$$

which goes to zero if we choose $\frac{r}{\alpha+(1-\alpha)r} \le q < 1$ so that $\frac{q}{1-q} \ge \frac{r}{\alpha(1-r)}$. Thus, combining (28), (30)-(33), we find that

$$
\Phi_{t_n}^{\alpha}\cdot\int_0^{t_n}\phi_n^{\alpha}(t,x)dt \to_P 1. \qquad (34)
$$

Hence, by Lemma 2.6, we have

$$
\Phi_{t_n}\int_0^{t_n}\phi_n(t,x)dZ_t \Rightarrow S_{\alpha}(1,\beta,0). \qquad (35)
$$

If $g_{n,3}(x) \to_P 0$ as $n \to \infty$, then the right hand side of (27) converges to zero in probability as $n \to \infty$ since $(n\Delta h)^{1-\frac{1}{\alpha}} \to 0$ when $0 < \alpha < 1$ and $(n\Delta h)^{1-\frac{1}{\alpha}} = 1$ when $\alpha = 1$ under condition (A.6). This contradicts (35). Therefore, we conclude that $g_{n,3}(x)$ does not converge to zero in probability. This completes the proof. □

3 Asymptotic Properties of the Nadaraya-Watson Estimator

In this section, we study the asymptotic behavior 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

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

The main result of this section is stated in the following theorem.

Theorem 3.1 *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 $(n\Delta h)^{1-\frac{1}{\alpha}}\Delta^{1/\kappa} = O(1)$ for some $\kappa > \alpha$, then

$$(n\Delta h)^{1-\frac{1}{\alpha}}\Lambda(x)(\hat{b}_n(x) - b(x)) \Rightarrow S_{\alpha}(1, \beta, 0). \tag{36}$$

(ii) If $(n\Delta h)^{1-\frac{1}{\alpha}}h^2 = O(1)$ and $(n\Delta h)^{1-\frac{1}{\alpha}}\Delta^{1/\kappa} = O(1)$ for some $\kappa > \alpha$, then

$$(n\Delta h)^{1-\frac{1}{\alpha}}\Lambda(x)(\hat{b}_n(x) - b(x) - h^2\Gamma_b(x)) \Rightarrow S_{\alpha}(1, \beta, 0). \tag{37}$$

Remark 3.2. (1) Since the set of conditions (A.1)-(A.3) and (A.5)-(A.8) in Theorem 3.1 is stronger than the set of conditions (A.1)-(A.6), all the results in section 2 are valid under (A.1)-(A.3) and (A.5)-(A.8).

(2) Let us establish some sufficient conditions on admissible bandwidth h and time frequency Δ. We first consider the conditions in part (i) of Theorem 3.1. When $1 < \alpha < 2$, to ensure $(n\Delta h)^{1-\frac{1}{\alpha}}h^2 = o(1)$, we may choose $h = (n\Delta \log(n\Delta))^{-\frac{\alpha-1}{3\alpha-1}}$ or $h = (n\Delta)^{-\frac{(1+\varepsilon)(\alpha-1)}{3\alpha-1}}$ for any $\varepsilon > 0$ which is compatible with (A.6) when $0 < \varepsilon < 2\alpha/(\alpha-1)$. By some basic calculation, we find that if $\Delta = O\left((\log n)^{\delta} n^{-\frac{2\kappa(\alpha-1)}{2\kappa(\alpha-1)+3\alpha-1}}\right)$ with

$$\delta = \frac{\kappa(\alpha-1)^2}{\alpha[2\kappa(\alpha-1) + 3\alpha - 1]}$$

or $\Delta = O(n^{-\gamma})$ with

$$\gamma = \frac{\frac{2(\alpha-1)}{3\alpha-1} - \frac{(\alpha-1)^2\varepsilon}{\alpha(3\alpha-1)}}{\frac{2(\alpha-1)}{3\alpha-1} + \frac{1}{\kappa} - \frac{(\alpha-1)^2\varepsilon}{\alpha(3\alpha-1)}},$$

then the condition $(n\Delta h)^{1-\frac{1}{\alpha}}\Delta^{1/\kappa} = O(1)$ is satisfied. By letting $\varepsilon = 0$ in the previous arguments, we obtain the admissible h and Δ for part (ii) of Theorem 3.1.
(3) When $\alpha = 1$, it is clear that $h^2 = o(1)$ and $\Delta^{1/\kappa} = o(1)$ under the condition (A.6). Thus, we have

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

This also shows that $\hat{b}_n(x)$ is not consistent when $\alpha = 1$ (see Theorem 2.2).

Now we are in a position to prove Theorem 3.1.

Proof of Theorem 3.1-(i). Note that

$$\begin{aligned}(n\Delta h)^{1-\frac{1}{\alpha}}\Lambda(x)(\hat{b}_n(x) - b(x)) &= \frac{(n\Delta h)^{1-\frac{1}{\alpha}}\Lambda(x)[\hat{g}_n(x) - b(x)\hat{f}_n(x)]}{\hat{f}_n(x)} \\ &:= \frac{V_n(x)}{\hat{f}_n(x)}. \qquad (38)\end{aligned}$$

Since $\hat{f}_n(x) \to f(x)$ in probability as $n \to \infty$, it is enough to study the asymptotic behavior of $V_n(x)$. By (15), we find

$$\begin{aligned}V_n(x) &= (n\Delta h)^{1-\frac{1}{\alpha}}\Lambda(x)[g_{n,1}(x) - b(x)\hat{f}_n(x)] \\ &\quad + (n\Delta h)^{1-\frac{1}{\alpha}}\Lambda(x)g_{n,2}(x) \\ &\quad + (n\Delta h)^{1-\frac{1}{\alpha}}\Lambda(x)g_{n,3}(x) \\ &:= V_{n,1}(x) + V_{n,2}(x) + V_{n,3}(x). \qquad (39)\end{aligned}$$

We have the following claims:
Claim 1.

$$\begin{aligned}& V_{n,1}(x) \\ =\ & o_P(1) + o_P(1)\cdot O((n\Delta h)^{1-\frac{1}{\alpha}}h^2) \\ & + \Lambda(x)[b'(x)f'(x) + \frac{1}{2}b''(x)f(x)]\int_{-\infty}^{\infty} u^2K(u)du\cdot(n\Delta h)^{1-\frac{1}{\alpha}}h^2(1+o(1)). \qquad (40)\end{aligned}$$

Claim 2.

$$V_{n,2}(x) = O_P(1)\cdot(n\Delta h)^{1-\frac{1}{\alpha}}\Delta + o_P(1)\cdot(n\Delta h)^{1-\frac{1}{\alpha}}\Delta^{1/\kappa}. \qquad (41)$$

Claim 3.

$$V_{n,3}(x) \Rightarrow f(x)S_\alpha(1, \beta, 0). \qquad (42)$$

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

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

By Taylor's expansion, we have

$$b(X_{t_i}) - b(x) = b^{'}(x)(X_{t_i} - x) + \frac{1}{2}b^{''}(x + \theta_i(X_{t_i} - x))(X_{t_i} - x)^2,$$

where θ_i is some random variable satisfying $\theta_i \in [0, 1]$. Thus, it follows that

$$\begin{aligned}
& V_{n,1}(x) \\
= \ & (n\Delta h)^{1-\frac{1}{\alpha}}\Lambda(x)b^{'}(x)\frac{1}{n}\sum_{i=0}^{n-1}(X_{t_i} - x)K_h(X_{t_i} - x) \\
& +(n\Delta h)^{1-\frac{1}{\alpha}}\Lambda(x)\frac{1}{2}b^{''}(x)\frac{1}{n}\sum_{i=0}^{n-1}(X_{t_i} - x)^2K_h(X_{t_i} - x) \\
& +(n\Delta h)^{1-\frac{1}{\alpha}}\Lambda(x)\frac{1}{n}\sum_{i=0}^{n-1}\frac{1}{2}[b^{''}(x + \theta_i(X_{t_i} - x)) - b^{''}(x)](X_{t_i} - x)^2K_h(X_{t_i} - x) \\
:= \ & V_{n,1}^{(1)}(x) + V_{n,1}^{(2)}(x) + V_{n,1}^{(3)}(x).
\end{aligned}$$

We have the following three claims:
Claim 1-(i):

$$V_{n,1}^{(1)}(x) = o_P(1) + \Lambda(x)b^{'}(x)f^{'}(x)\int_{-\infty}^{\infty}u^2K(u)du \cdot (n\Delta h)^{1-\frac{1}{\alpha}}h^2(1 + o(1)). \tag{43}$$

Claim 1-(ii):

$$V_{n,1}^{(2)}(x) = o_P(1) + \frac{1}{2}\Lambda(x)b^{''}(x)f(x)\int_{-\infty}^{\infty}u^2K(u)du \cdot (n\Delta h)^{1-\frac{1}{\alpha}}h^2(1 + o(1)). \tag{44}$$

Claim 1-(iii):

$$V_{n,1}^{(3)}(x) = o_P(1) \cdot O((n\Delta h)^{1-\frac{1}{\alpha}}h^2). \tag{45}$$

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

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

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

By the stationarity of X_t, we have

$$\begin{aligned}
V_{n,1}^{(1)}(x) &= \Lambda(x)b^{'}(x) \cdot \frac{1}{n}\sum_{i=1}^{n}\xi_{n,i}(x) \\
&\quad +\Lambda(x)b^{'}(x)(n\Delta h)^{1-\frac{1}{\alpha}}\mathbb{E}[(X_0 - x)K_h(X_0 - x)] \\
&:= D_{n,1}(x) + D_{n,2}(x).
\end{aligned} \tag{46}$$

We apply the inequality (1.26) in Theorem 1.3 of Bosq [6] to prove that $D_{n,1}(x) = o_P(1)$. It suffices to show that $\frac{1}{n}\sum_{i=1}^{n} \xi_{n,i}(x) = o_P(1)$. Note that

$$\sup_{1\le i\le n} |\xi_{n,i}(x)| \le M_0(n\Delta h)^{1-\frac{1}{\alpha}} \ a.s.$$

for some positive constant $M_0 < \infty$. Applying Theorem 1.3 of Bosq [6], we have for each integer $q \in [1, \frac{n}{2}]$ and each $\varepsilon > 0$

$$\begin{aligned}\mathbb{P}\left(\frac{1}{n}\left|\sum_{i=0}^{n-1} \xi_{n,i}(x)\right| > \varepsilon\right) &\le 4\exp\left(-\frac{\varepsilon^2 q}{8v^2(q)}\right) \\ &\quad +22\left(1+\frac{4M_0(n\Delta h)^{1-\frac{1}{\alpha}}}{\varepsilon}\right)^{1/2} q\alpha_X\left([p]\Delta\right),\end{aligned}$$

where

$$v^2(q) = \frac{2}{p^2}s(q) + \frac{M_0(n\Delta h)^{1-\frac{1}{\alpha}}\varepsilon}{2}$$

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

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

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

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

Therefore, we conclude that $\frac{1}{n}\sum_{i=1}^{n}\xi_{n,i}(x) = o_P(1)$. Now, we turn to $D_{n,2}(x)$. Note that

$$\begin{aligned}
& \mathbb{E}[(X_0-x)K_h(X_0-x)] \\
= & \int_{-\infty}^{\infty} \frac{y-x}{h} K\left(\frac{y-x}{h}\right) f(y)dy \\
= & \int_{-\infty}^{\infty} uK(u)f(x+uh)du \cdot h \\
= & hf(x)\int_{-\infty}^{\infty} uK(u)du + h^2 \int_{-\infty}^{\infty} u^2K(u)f^{'}(x+\theta uh)du \\
= & f^{'}(x)\int_{-\infty}^{\infty} u^2K(u)du \cdot h^2(1+o(1)).
\end{aligned}$$

Thus, we have

$$D_{n,2}(x) = \Lambda(x)b^{'}(x)f^{'}(x)\int_{-\infty}^{\infty} u^2K(u)du \cdot (n\Delta h)^{1-\frac{1}{\alpha}} h^2(1+o(1)). \tag{47}$$

This completes the proof of claim 1-(i).

Proof of Claim 1-(ii). The proof ideas are the same as above. For $i = 0, \cdots, n-1$, let

$$\zeta_{n,i+1}(x) = (n\Delta h)^{1-\frac{1}{\alpha}} \left((X_{t_i}-x)^2K_h(X_{t_i}-x) - \mathbb{E}[(X_{t_i}-x)^2K_h(X_{t_i}-x)]\right).$$

By the stationarity of X_t, we have

$$\begin{aligned}
V_{n,1}^{(2)}(x) &= \frac{1}{2}\Lambda(x)b^{''}(x)\cdot\frac{1}{n}\sum_{i=0}^{n-1}\zeta_{n,i}(x) \\
& \quad +\frac{1}{2}\Lambda(x)b^{'}(x)(n\Delta h)^{1-\frac{1}{\alpha}}E[(X_0-x)^2K_h(X_0-x)] \\
&:= D_{n,3}(x) + D_{n,4}(x).
\end{aligned} \tag{48}$$

Note that

$$\sup_{1\le i\le n}|\zeta_{n,i}(x)| \le M_1(n\Delta h)^{1-\frac{1}{\alpha}}h \ a.s.$$

for some positive constant $M_1 < \infty$. Applying Theorem 1.3 of Bosq [6], we have for each integer $q \in [1, \frac{n}{2}]$ and each $\varepsilon > 0$

$$\begin{aligned}
\mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n}\zeta_{n,i}(x)\right| > \varepsilon\right) &\le 4\exp\left(-\frac{\varepsilon^2 q}{8\tilde{v}^2(q)}\right) \\
& \quad +22\left(1+\frac{4M_1(n\Delta h)^{1-\frac{1}{\alpha}}h}{\varepsilon}\right)^{1/2} q\alpha_X([p]\Delta),
\end{aligned}$$

where

$$\tilde{v}^2(q) = \frac{2}{p^2}\tilde{s}(q) + \frac{M_1(n\Delta h)^{1-\frac{1}{\alpha}}h\varepsilon}{2}$$

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

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

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

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

Then, we have

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

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

$$22\left(1+\frac{4M_1(n\Delta h)^{1-\frac{1}{\alpha}}h}{\varepsilon}\right)^{1/2} q\alpha_X\left([p]\Delta\right) \to 0.$$

Therefore, we have $D_{n,3}(x) = o_P(1)$. Now, we consider $D_{n,4}(x)$. Basic calculation yields that

$$\mathbb{E}[(X_0-x)^2K_h(X_0-x)] = f(x)\int_{-\infty}^{\infty} u^2K(u)du\cdot h^2(1+o(1)).$$

Hence, it follows that

$$D_{n,4}(x) = \frac{1}{2}\Lambda(x)b^{'}(x)f(x)\int_{-\infty}^{\infty} u^2K(u)du\cdot(n\Delta h)^{1-\frac{1}{\alpha}}h^2(1+o(1)). \tag{49}$$

Thus, the claim 1-(ii) holds.

Proof of Claim 1-(iii). By the uniform continuity of $b^{''}(\cdot)$ and the bounded support of the kernel function $K(\cdot)$ (assuming that $K(x) = 0$ if $|x| > M$ for some finite positive number M), we have

$$\begin{aligned}|V_{n,1}^{(3)}(x)| &\le \frac{1}{2}\Lambda(x)\sup_{|x-y|\le Mh}|b^{''}(x)-b^{''}(y)|(n\Delta h)^{1-\frac{1}{\alpha}}\frac{1}{n}\sum_{i=0}^{n-1}(X_{t_i}-x)^2K_h(X_{t_i}-x) \\ &= o(1)\cdot(n\Delta h)^{1-\frac{1}{\alpha}}h^2\frac{1}{n}\sum_{i=0}^{n-1}K_h(X_{t_i}-x) \\ &= o_P(1)\cdot(n\Delta h)^{1-\frac{1}{\alpha}}h^2, \qquad (50)\end{aligned}$$

since $\frac{1}{n}\sum_{i=0}^{n-1}K_h(X_{t_i}-x) \to f(x)$ in probability. Hence, claim 1-(iii) holds.

Proof of Claim 2. By (22), we have

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

Similar to claim (i) in the proof of Theorem 2.1, under given conditions, we have

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

As in the proof of Claim (ii) of Theorem 2.1 via Lemma 2.7, we can show that

$$P\left(\frac{1}{n\Delta^{\frac{1}{\kappa}}}\sum_{i=0}^{n-1}K_h(X_{t_i}-x)\sup_{t_i\leq t\leq t_{i+1}}\left|\int_{t_i}^{t}\sigma(X_{s-})dZ_s\right| > \varepsilon\right) \leq O(\Delta^{\frac{1}{\alpha}-\frac{1}{\kappa}}).$$

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 (35), it follows that

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

i.e., claim 3 holds.

Finally, by (39) and the claims 1-3, we obtain

$$V_n(x) \Rightarrow f(x)S_\alpha(1,\beta,0) \tag{51}$$

under the given conditions. By using (38), (51), Slutsky's theorem and Lemma 2.4, 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 part (i) of Theorem 3.1.

Proof of Theorem 3.1-(ii). If $(n\Delta h)^{1-\frac{1}{\alpha}} h^2 = O(1)$, by (40), 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 (41) and (42), 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 $(n\Delta h)^{1-\frac{1}{\alpha}}\Delta^{\frac{1}{\kappa}} = O(1)$. Then, we use (39) 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 $f(x)/\hat{f}_n(x) \to 1$ in probability, we find

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

This completes the proof. □

4 A Simulation Study

In this section, we conduct a simulation study to confirm the finite sample properties of the asymptotic results developed in sections 2 and 3. 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]):

$$X_{t_{i+1}} = X_{t_i} + b(X_{t_i})\Delta + \sigma \cdot \Delta Z_{t_i}, \tag{52}$$

where $t_i = i\Delta$ and $\Delta Z_{t_i} = Z_{t_{i+1}} - Z_{t_i}$, $i = 0, 1, \cdots, n-1$. 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:

- 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. 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

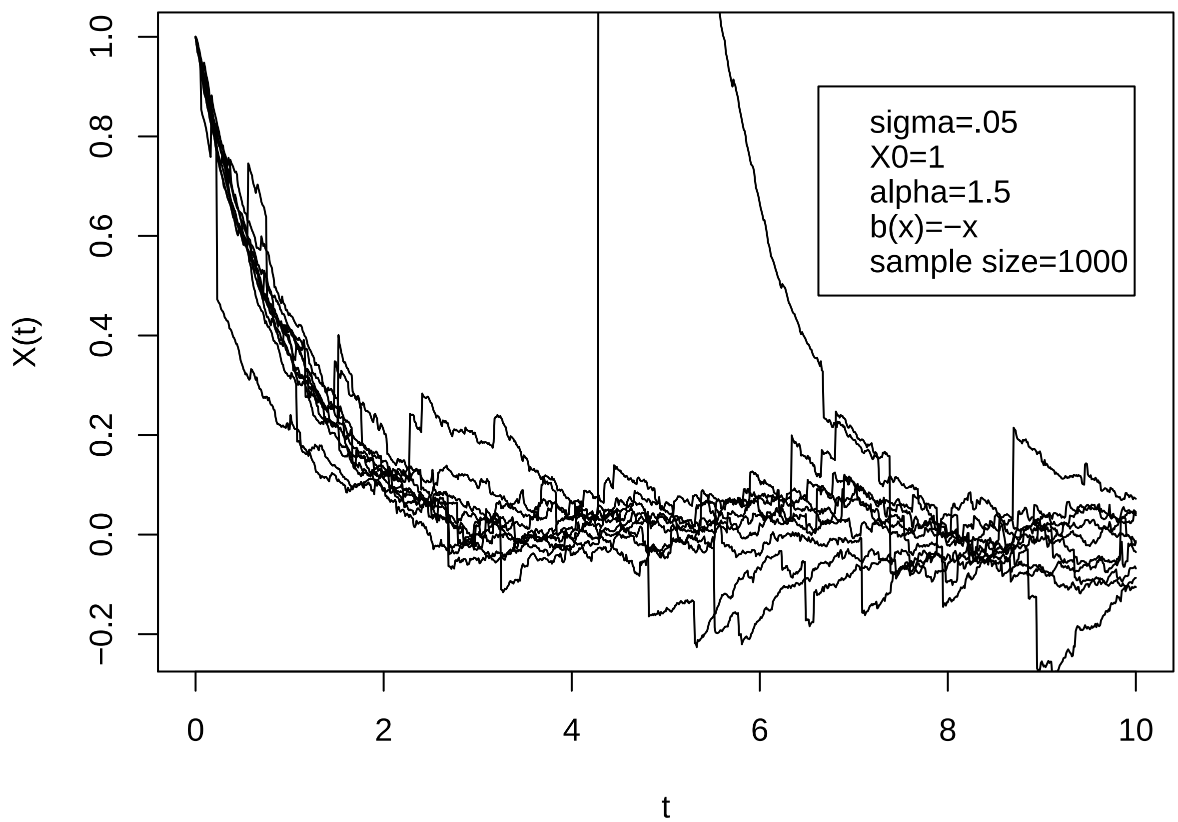

Figure 1: The ten sample paths of the process X_t

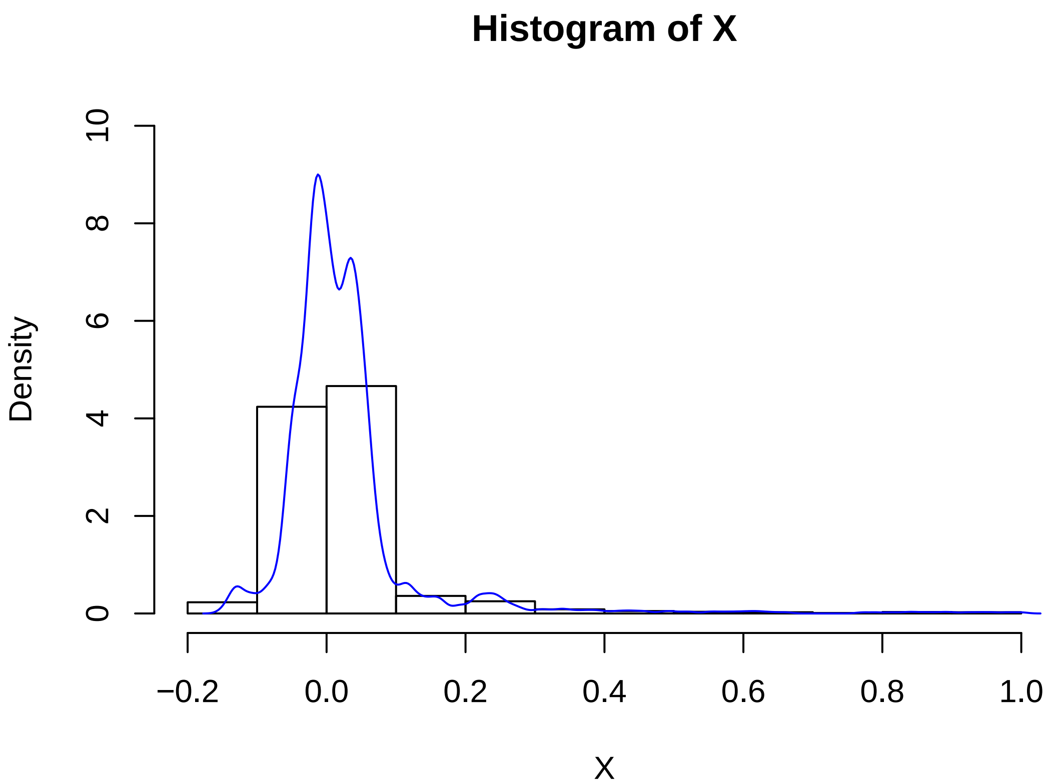

Figure 2: The estimated density function of the process X_t.

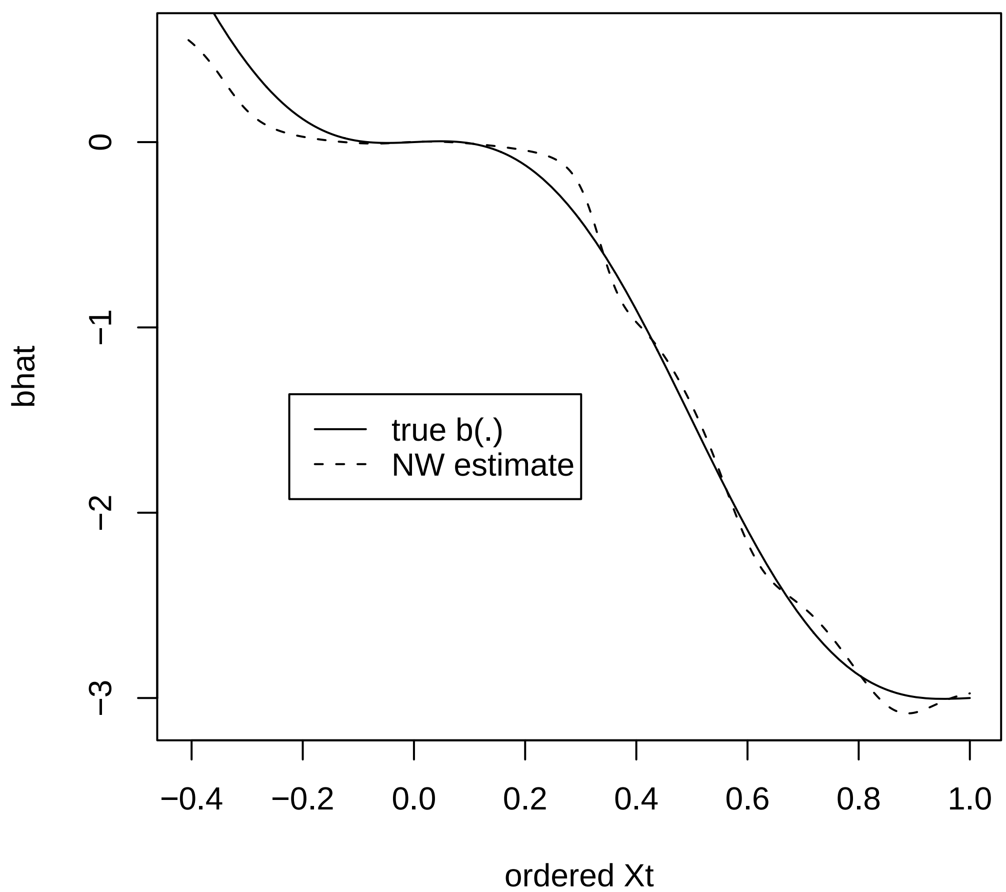

Figure 3: The N-W estimate of $b(.)$ for $\sigma = 0.05, c = 3, d = 0.5$ and $b(x) = -cx + d\sin(2\pi x)$ for $T = 100, n = 1000$.

on 1000 replicates. In each cell, the first, second and third numbers represent the mean, standard deviation and median of $RASE$ for three sample sizes $n = 1000$, $n = 5000$ and $n = 10000$, respectively.

Table 1: Simulation results on RASE for three lengths of time interval and three sample sizes for six parameter settings. mean=sample mean, median=sample median, sd=sample standard deviation over the 1000 replicates.

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

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 sensitive to the linearity assumption on the drift function. As the σ increases, the summary statistics of RASE confirm the results in Theorem 3.1. 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, 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.

References

[1] D. Applebaum, Lévy Processes and Stochastic Calculus, Cambridge Studies in Advanced Mathematics, Vol. 93, Cambridge, 2004.

[2] M. Arfi, Non-parametric drift estimation from ergodic samples, *Scand. J. Statist.* **25** (1995) 225-234.

[3] F.M. Bandi, P.C.B. Phillips, Fully nonparametric estimation of scalar diffusion models, *Econometrica* **71** (2003) 241-283.

[4] G. Banon, Nonparametric identification for diffusion processes, *SIAM J. Control Optim.* **16** (1978) 380-395.

[5] P. Billingsley, Probability and Measure, Third Edition, Wiley, New York, 1995.

[6] D. Bosq, Nonparametric Statistics for Stochastic Processes, Lecture Notes in Statistics, Vol. 110, Springer-Verlag, New York, 1996.

[7] A. Dalalyan, Sharp adaptive estimation of the drift function for ergodic diffusions, *Ann. Statist.* **33** (2005) 2507-2528.

[8] A. Dalalyan, Yu.A. Kutoyants, Asymptotically efficient trend coefficient estimation for ergodic diffusion, *Math. Methods Statist.* **11** (2002) 402-427.

[9] A. Dalalyan, Yu.A. Kutoyants, Asymptotically efficient estimation of the derivative of invariant density, *Stat. Inference Stoch. Process.* **6** (2003) 89-107.

[10] P.D. Ditlevsen, Observation of α-stable noise induced millennial climate changes from an ice-core record, *Geophysical Research Letter* **26** (1999) 1441-1444.

[11] P.D. Ditlevsen, Anomalous jumping in a double-well potential, *Phys. Rev. E* **60** (1999) 172-179.

[12] J. Fan, Design-adaptive nonparametric regression, *J. Amer. Statist. Assoc.* **87** (1992) 998-1004.

[13] J. Fan, A selection overview of nonparametric methods in financial econometrics, *Statist. Sci.* **20** (2005) 317-337.

[14] J. Fan, I. Gijbels, Local Polynomial Modelling and Its Applications, Chapman and Hall, London, 1996.

[15] J. Fan, C. Zhang, A reexamination of diffusion estimators with applications to financial model validation, *J. Amer. Statist. Assoc.* **98** (2003) 118-134.

[16] B. Fristedt, 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, 1974.

[17] E. Gobet, M. Hoffmann, M. Reiss, Nonparametric estimation of scalar diffusions based on low frequency data, *Ann. Statist.* **32** (2004) 2223-2253.

[18] P. Hall, L. Peng, Q. Yao, Prediction and nonparametric estimation for time series with heavy tails, *J. Time Ser. Anal.* **23** (2002) 313-331.

[19] W. Härdle, Applied nonparametric regression, Cambridge University Press, Cambridge, 1990.

[20] M. Hoffmann, Adaptive estimation in diffusion processes, *Stochastic Process. Appl.* **79** (1999) 135-163.

[21] Y. Hu, H. Long, Parameter estimation for Ornstein-Uhlenbeck processes driven by α-stable Lévy motions, *Communications on Stochastic Analysis* **1** (2007) 175-192.

[22] Y. Hu, H. Long, Least square estimator for Ornstein-Uhlenbeck processes driven by α-stable motions, Submitted, 2007.

[23] J. Jacod, The Euler scheme for Lévy driven stochastic differential equations: limit theorems, *Ann. Probab.* **32** (2004) 1830-1872.

[24] J. Jacod, T.G. Kurtz, S. Méléard, P. Protter, The approximate Euler method for Lévy driven stochastic differential equations, *Ann. Inst. H. Poincaré Probab. Statist.* **41** (2005) 523-558.

[25] A. Janicki, A. Weron, Simulation and Chaotic Behavior of α-stable Stochastic Processes, Marcel Dekker, New York, 1994.

[26] A. Joulin, On maximal inequalities for stable stochastic integrals, *Potential Anal.* **26** (2007) 57-78.

[27] O. Kallenberg, Some time change representations of stable integrals, via predictable transformations of local martingales, *Stochastic Process. Appl.* **40** (1992) 199-223.

[28] Yu. A. Kutoyants, Statistical Inference for Ergodic Diffusion Processes, Springer-Verlag, London, Berlin, Heidelberg, 2004.

[29] A. Kyprianou, W. Schoutens, P. Wilmott (Eds.), Exotic Option Pricing and Advanced Lévy Models, Wiley, 2005.

[30] H. Masuda, Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps, *Stochastic Process. Appl.* **117** (2007) 35-56.

[31] T. Mikosch, S. Resnick, H. Rootzen, A. Stegeman, Is network traffic approximated stable Lévy motion or fractional Brownian motion? *Ann. Appl. Probab.* **12** (2002) 23-68.

[32] E.A. Nadaraya, On estimating regression, *Theory Probab. Appl.* **9** (1964) 141-142.

[33] L. Peng, Q. Yao, Nonparametric regression under dependent errors with infinite variance, *Ann. Inst. Statist. Math.* **56** (2004) 73-86.

[34] D.T. Pham, Nonparametric estimation of the drift coefficient in the diffusion equation, *Math. Operationsforsch. Statist. Ser. Statist.* **12** (1981) 61-73.

[35] B.L.S. Prakasa Rao, Estimation of the drift for diffusion process, *Statistics* **16** (1985) 263-275.

[36] J. Rosinski, W.A. Woyczynski, 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, 1985.

[37] J. Rosinski, W.A. Woyczynski, On Ito stochastic integration with respect to p-stable motion: inner clock, integrability of sample paths, double and multiple integrals, *Ann. Probab.* **14** (1986) 271-286.

[38] G. Samorodnitsky, M.S. Taqqu, Stable non-Gaussian Random Processes: Stochastic Models with Infinite Variance, Chapman & Hall, New York, London, 1994.

[39] K.I. Sato, Lévy Processes and Infinitely Divisible Distributions, Cambridge University Press, Cambridge, 1999.

[40] D. Schertzer, M. Larcheveque, J. Duan, V.V. Yanovsky, S. Lovejoy, Fractional Fokker-Planck equation for nonlinear stochastic differential equations driven by non-Gaussian Lévy stable noises, *J. Math. Phys.* **42** (2001) 200-212.

[41] W. Schoutens, Lévy Processes in Finance: Pricing Financial Derivatives, Wiley, Hoboken, NJ, USA, 2003.

[42] V.G. Spokoiny, Adaptive drift estimation for nonparametric diffusion model, *Ann. Statist.* **28** (2000) 815-836.

[43] J.H. van Zanten, Rates of convergence and asymptotic normality of curve estimators for ergodic diffusion processes, Report PNA-R0011, CWI, Amsterdam, 2000.

[44] G.S. Watson, Smooth regression analysis, *Sankhya, Ser. A.* **26** (1964) 359-372.

[45] W.B. Wu, Nonparametric estimation for stationary processes, Technical Report No. 536, Department of Statistics, The University of Chicago, 2003.

[46] W.B. Wu, Nonlinear system theory: another look at dependence, *Proc. Natl. Acad. Sci. USA* **102** (2005) 14150-14154.