

# Approximate self-weighted LAD estimation of discretely observed ergodic Ornstein-Uhlenbeck processes

Masuda, Hiroki  
Faculty of Mathematics, Kyushu University

<https://hdl.handle.net/2324/16247>

---

出版情報 : MI Preprint Series. 2010-1, 2010-01-11. 九州大学大学院数理学研究院  
バージョン :  
権利関係 :



# MI Preprint Series

Kyushu University  
The Global COE Program  
Math-for-Industry Education & Research Hub

## Approximate self-weighted LAD estimation of discretely observed ergodic Ornstein-Uhlenbeck processes

Hiroki Masuda

MI 2010-1

( Received January 11, 2010 )

Faculty of Mathematics  
Kyushu University  
Fukuoka, JAPAN

# Approximate Self-Weighted LAD Estimation of Discretely Observed Ergodic Ornstein-Uhlenbeck Processes

Hiroki Masuda

*Graduate School of Mathematics, Kyushu University,  
744, Motoooka, Nishi-ku, Fukuoka 819-0395, Japan  
Email: hiroki@math.kyushu-u.ac.jp*

This version: January 11, 2010

## Abstract

We consider drift estimation of a discretely observed Ornstein-Uhlenbeck process driven by a possibly heavy-tailed symmetric Lévy process with positive Blumenthal-Gettoor activity index  $\beta$ . Under an infill and large-time sampling design, we first establish an asymptotic normality of a self-weighted least absolute deviation estimator with the rate of convergence being  $\sqrt{nh_n^{1-1/\beta}}$ , where  $n$  denotes sample size and  $h_n > 0$  the sampling mesh satisfying that  $h_n \rightarrow 0$  and  $nh_n \rightarrow \infty$ . This implies that the rate of convergence is determined by the most active part of the driving Lévy process; the presence of a driving Wiener part leads to  $\sqrt{nh_n}$ , which is familiar in the context of asymptotically efficient estimation of diffusions with compound Poisson jumps, while a pure-jump driving Lévy process leads to a faster one. Also discussed is how to construct corresponding asymptotic confidence regions without full specification of the driving Lévy process. Second, by means of a polynomial type large deviation inequality we derive convergence of moments of our estimator under additional conditions.

## 1 Introduction

Estimation of discretely observed stochastic processes with jumps has received growing interest from both theoreticians and practitioners. Among others, Markovian Ornstein-Uhlenbeck (OU for short) process has several attractive features mainly stemming from its continuous-time first-order autoregressive structure. Let  $X = (X_t)_{t \in \mathbb{R}_+}$  be the univariate OU process given by the stochastic differential equation

$$dX_t = (\gamma - \lambda X_t)dt + dZ_t, \quad (1)$$

where  $Z$  is a nontrivial symmetric Lévy process independent of  $X_0$ . In this paper, we are concerned with estimation of the true value  $\theta_0 := (\lambda_0, \gamma_0)$  of the unknown parameter

$$\theta := (\lambda, \gamma) \in \Theta \subset (0, \infty) \times \mathbb{R}$$

based on a discrete-time data  $(X_{t_i})_{i=0}^n$ , without full specification of  $Z$ 's Lévy measure. Here  $t_i = t_i^n = ih$  with  $h = h_n > 0$  such that  $h \rightarrow 0$  and  $nh \rightarrow \infty$  as  $n \rightarrow \infty$ , that is, we consider infill and large-time asymptotics for sampling design. Analysis of non-Gaussian OU processes was initiated by Doob [4] for symmetric stable  $Z$ . Stochastic modeling of several physical phenomena has been supported by the non-Gaussianity through realistic experimentations; see Garbaczewski and Olkiewicz [5] and the references therein.

As we target drift estimation, the most naive but practical way would be to use the approximate least-squares estimator (LSE), which minimizes the contrast function

$$\theta \mapsto \sum_{i=1}^n \{X_{t_i} - X_{t_{i-1}} - h(\gamma - \lambda X_{t_{i-1}})\}^2.$$

Indeed, the LSE fulfils an asymptotic normality when  $Z$  is centered with finite moments and  $nh^2 \rightarrow 0$  as  $nh \rightarrow \infty$ , the resulting rate of convergence being necessarily  $\sqrt{nh}$ ; see Masuda [14] as well as Section 2.2.2 below. Although the rate  $\sqrt{nh}$  is well known to be optimal in the context of drift estimation of diffusions with compound Poisson jumps, our main result says that this is no longer the case as soon as  $Z$  is of pure-jump type. Hu and Long [7] recently studied the LSE for  $\lambda > 0$  when  $Z$  is symmetric  $\beta$ -stable, with supposing that  $\gamma_0 = 0$  from the beginning. Our results are completely different from theirs in view of the rate of convergence and the limit distribution. See Section 2.2.2 for some theoretical comparisons between their result and ours.

Instead, motivated by Ling [11], in this paper we introduce an approximate least absolute deviation (LAD) type estimator and study its asymptotic behavior. The LAD estimation has a long history and is one of popular estimation procedures robust to outlying observations. The LAD estimator is based on the “Laplacian”  $L^1$ -loss, while the LSE on the “Gaussian”  $L^2$ -one. We refer to, among others, Knight [8], Koenker [9], and Portnoy and Koenker [16] as well as the references therein for a detailed account and historical backgrounds of the LAD estimation. The LAD type estimation has been deeply investigated also in the time-series literature, e.g., Davis and Dunsmuir [2], Davis *et al.* [3], and so on. Just for illustrative purposes, suppose that observed time-series data stems from the ergodic first-order autoregressive model  $X_k = \theta_0 X_{k-1} + \epsilon_k$ ,  $k \leq n$ , where  $|\theta_0| < 1$  and  $(\epsilon_k)$  is an i.i.d. noise sequence with common median 0. Then, the *unweighted* LAD estimator  $\hat{\theta}_n$  of  $\theta_0$  is defined to be a minimizer of the contrast function  $\theta \mapsto \sum_{k=1}^n |X_k - \theta X_{k-1}|$ . If  $\epsilon_1$  admits finite absolute moments of sufficiently high order,  $\hat{\theta}_n$  is known to be asymptotically normally distributed at rate  $\sqrt{n}$ . On the other hand, in case where  $\epsilon_k$  has infinite-variance, it is known that the maximum likelihood and the LAD estimators have a faster rate of convergence than  $\sqrt{n}$ , while both of them lead to intractable limit distributions; see Andrews *et al.* [1] and Davis *et al.* [3] for details in this direction. On the other hand, Ling [11] introduced a *self-weighted* LAD (SLAD) contrast function for infinite-variance autoregressive models, which entails asymptotically normally distributed estimators at rate  $\sqrt{n}$ . That is to say, Ling’s result means that we may derive a conventional asymptotic normality result in compensation for the slower rate of convergence than the maximum likelihood and the unweighted LAD estimators. It can be expected that the SLAD estimation can be employed on a robust drift estimation for discretely observed continuous-time stochastic processes as well, in which large jumps may deteriorate finite-sample performance of the LSE or, more generally, the quasi-likelihood estimator.

Our SLAD estimator is defined as a minimizer  $\hat{\theta}_n$  of the contrast function

$$\theta \mapsto \sum_{i=1}^n w(X_{t_{i-1}}) |X_{t_i} - X_{t_{i-1}} - h(\gamma - \lambda X_{t_{i-1}})|,$$

for an appropriate weight function  $w$ ; the unweighted LAD estimation corresponds to the case where  $w \equiv 1$ . Under regularity conditions, we first derive an *asymptotic normality of  $\hat{\theta}_n$  at rate  $\sqrt{nh}^{1-1/\beta}$*  (see Theorem 2.1), where  $\beta$  stands for the Blumenthal-Gettoor activity index of the driving Lévy process (see (5) below for the definition). As a result, when  $Z$  is of pure-jump type, we have a faster rate of convergence than the familiar  $\sqrt{nh}$ . It is interesting that we could get faster rate of convergence only by changing the type of loss from  $L^2$  to  $L^1$ . Although the corresponding asymptotic covariance matrix as well as the rate of convergence inevitably depends on the unknown index  $\beta$ , we clarify that it is possible to formulate a feasible construction of asymptotic confidence interval; specifically, we can construct explicit

statistics  $\hat{T}_n$  such that  $\hat{T}_n(\hat{\theta}_n - \theta_0)$  tends to the standard normal distribution (see Theorem 2.3). Due to robustness of LAD type estimates to “outlying” data, our SLAD estimator should be robust to “big” jumps caused by the driving process  $Z$  without individual detection of them, making the estimation procedure more practical. Also obtained under additional conditions is *the convergence of moments of the normalized quantities*  $\sqrt{nh}^{1-1/\beta}(\hat{\theta}_n - \theta_0)$  (see Theorem 2.2). This much stronger mode of convergence is obtained as a byproduct of the polynomial type large deviation inequality (32), which we prove by applying a general result due to Yoshida [25]. Finally, we remark that convergence of moments as well as a large deviation inequality is a crucial tool for investigating: asymptotic behavior of expected values of statistics depending on estimators; also, error estimate appearing in higher-order theoretical statistics. For smooth statistical random fields associated with a stochastic process, large deviation inequalities have been investigated and applied, e.g., to the information criteria in model selection, the validity of higher-order asymptotic statistical theory, and moment convergence for quasi-likelihood and Bayes estimators of multidimensional ergodic diffusion processes; among others, see Uchida and Yoshida [23, 24], Sakamoto and Yoshida [19], and Yoshida [25] for details in these directions. To the best of author’s knowledge, our Theorem 2.2 is the first result providing a large deviation inequality and convergence of moments associated with a non-differentiable LAD type statistical random fields for dependent data.

In the rest of this paper, we state our main results in Section 2, and then Section 3 is devoted to the proofs.

## 2 Asymptotic behavior of the SLAD estimator

### 2.1 Statement of main results

Let  $X$  be given by (1), and denote by  $\eta$  the initial distribution of  $X$ . Throughout this paper we assume that:

$$\Theta \text{ is a bounded convex domain such that its closure } \Theta^- \subset (0, \infty) \times \mathbb{R}; \quad (2)$$

$$h \rightarrow 0 \text{ and } nh \rightarrow \infty; \quad (3)$$

$$\text{There exists a constant } q > 0 \text{ such that } \int |x|^q \eta(dx) < \infty. \quad (4)$$

Here and in what follows, asymptotic symbols are used for  $n \rightarrow \infty$  unless otherwise mentioned.

We denote by  $\nu$  and  $\sigma^2$  the Lévy measure and Gaussian variance of  $Z$ , respectively; we implicitly presuppose that either  $\sigma^2 > 0$  or  $\nu(\mathbb{R}) > 0$ , excluding the trivial case. We refer to Sato [20] for a systematic account of Lévy processes. The Blumenthal-Gettoor activity index  $\beta$  of  $Z$ , which measures the degree of small-jump fluctuations, is defined as follows:

$$\beta := \begin{cases} 2, & \text{if } \sigma^2 > 0, \\ \inf \left\{ r \geq 0 : \int_{|z| \leq 1} |z|^r \nu(dz) < \infty \right\}, & \text{if } \sigma^2 = 0. \end{cases} \quad (5)$$

As  $\beta \leq 2$ , we always have  $\sqrt{nh}^{1-1/\beta} \rightarrow \infty$  under (3).

#### 2.1.1 Asymptotic normality

We impose some structural assumptions on  $Z$ .

**Assumption 1.** (*Structure of  $Z$* )

1.  $\nu$  is symmetric around the origin,

$$\int_{|z|>1} |z|^q \nu(dz) < \infty$$

( $q$  is the same one as in (4)), and the characteristic function of  $Z_t$  is given by

$$\varphi_{Z_t}(u) := \exp \left\{ t \left( -\frac{\sigma^2}{2} u^2 + \int (\cos(uz) - 1) \nu(dz) \right) \right\}, \quad u \in \mathbb{R}.$$

2. If  $\beta < 2$  (hence  $\sigma^2 = 0$ ), then  $\nu$  is decomposed as  $\nu = \nu' + \nu''$  for two Lévy measures  $\nu'$  and  $\nu''$  satisfying the following (a) and (b).

(a)  $\nu'$  admits a (symmetric) density  $g$  on  $U \setminus \{0\}$  for an open neighborhood  $U$  of the origin, where  $g$  fulfils that

$$g(z) = c|z|^{-1-\beta} \{1 + \tilde{g}(z)\}$$

for some  $\beta \in (0, 2)$ ,  $c > 0$ , and a function  $\tilde{g}$  such that  $\tilde{g}(z) = O(|z|^\delta)$  as  $|z| \rightarrow 0$  for some  $\delta > 0$ .

(b)  $\beta'' := \inf \left\{ r \geq 0 : \int_{|z| \leq 1} |z|^r \nu''(dz) \right\} \in [0, \beta)$ .

We may ignore Assumption 1.2 if  $\sigma^2 > 0$ . Roughly speaking, Assumption 1.2 entails that small fluctuations of  $Z$  should be like that of a  $\beta$ -stable Lévy process; in particular, the measure  $\nu''$  and the function  $\tilde{g}$  are identically null for the symmetric  $\beta$ -stable  $Z$ . Especially if  $g(z) = |z|^{-1-\beta} v(z)$  with  $v$  being positive, bounded, and smooth on  $U \setminus \{0\}$ , then we have  $\delta = 1$  in Assumption 1.2(a). Typical such examples are the generalized hyperbolic (except for the variance gamma) and the exponentially tempered stable cases, the corresponding Lévy densities (on the whole  $\mathbb{R} \setminus \{0\}$ ) of which are given by

$$z \mapsto a_0 |z|^{-2} + a_1 |z|^{-1} + o(|z|^{-1})$$

for  $|z| \rightarrow 0$  and

$$z \mapsto b_0 |z|^{-1-\beta} e^{-b_1 |z|},$$

respectively, with some positive constants  $a_k$  and  $b_k$ ; see Raible [17, pp.39–40] for the former, and Rosiński [18] as well as the references therein for the latter.

Before proceeding, we point out some facts concerning the OU processes. Denote by  $P_0$  the distribution of  $X$  associated with  $\theta_0$ , and by  $E_0$  the corresponding expectation operator. Then we know the following, both of which are essential in our forthcoming results.

- $X$  admits a unique invariant distribution, say  $\pi_0$  (depending on  $\theta_0$ ), and is exponentially absolute regular (hence exponentially strong-mixing) under  $P_0$ .
- $\int |x|^q \pi_0(dx) < \infty$  and  $\sup_{t \in \mathbb{R}_+} E_0[|X_t|^q] < \infty$ , with  $q$  being the same one as in Assumption 1.

See Masuda [13, 15] for more details. We also note that  $\pi_0$  is necessarily selfdecomposable, hence admits a density with respect to the Lebesgue measure. The characteristic function of  $\pi_0$  is given by

$$u \mapsto \exp \left\{ i \left( \frac{\gamma_0}{\lambda_0} \right) u - \frac{1}{2} \left( \frac{\sigma^2}{2\lambda_0} \right) u^2 + \int (\cos(uz) - 1) \int_0^\infty \nu(e^{\lambda_0 s} dz) dz \right\}. \quad (6)$$

Trivially, the density of  $\pi_0$  is symmetric around  $\gamma_0/\lambda_0$ .

Now we introduce our contrast function

$$M_n(\theta) = \sum_{i=1}^n w(X_{t_{i-1}}) |\Delta_i X - (\gamma - \lambda X_{t_{i-1}}) h|, \quad (7)$$

where  $\Delta_i X := X_{t_i} - X_{t_{i-1}}$ ,  $i \leq n$ , and  $w : \mathbb{R} \rightarrow \mathbb{R}_+$  (supposed to be free of  $\theta$ ). Then the SLAD estimator is defined to be any measurable mapping  $\hat{\theta}_n$  such that  $M_n(\hat{\theta}_n) = \inf_{\theta \in \Theta^-} M_n(\theta)$ . We impose the following technical conditions on the “weight” function  $w$ .

**Assumption 2.** (Regularity of  $w$ )

1.  $w$  is bounded and uniformly continuous.
2.  $\limsup_{|x| \rightarrow \infty} w(x)|x|^{4-q} < \infty$  ( $q$  is the same one as in Assumption 1). Especially if  $\int_{|z|>1} |z|^{q'} \nu(dz) = \infty$  for any  $q' > 2$ , then the function  $x \mapsto w(x)x^2$  is uniformly continuous.

In analogy with Ling [11], in order to deduce an asymptotic normality result, Assumption 2.2 is indispensable for  $\nu$  having heavy tails. Note that Assumption 2.2 is redundant if we can pick  $q \geq 4$ , and in this case  $w$  does not need to be tapering at infinity (in particular, we may take  $w \equiv 1$ , arriving back to the unweighted LAD estimation). See also Section 2.2.1 for some related remarks.

We need the following condition on the decreasing rate of  $h = h_n$  in connection with the value  $\beta$ .

**Assumption 3.**  $nh^{4-2/\beta} \rightarrow 0$ .

In view of (3), Assumption 3 entails that  $\beta > 2/3$ . Note that it suffices that  $nh^3 \rightarrow 0$  when  $\sigma^2 > 0$ , while a faster decay of  $h$  is required when  $\sigma^2 = 0$ . The so-called “rapidly increasing experimental design”, i.e.  $nh^2 \rightarrow 0$ , is sufficient as soon as  $\beta \geq 1$ .

We denote by  $\mathcal{N}_r(\mu, V)$  the  $r$ -dimensional normal distribution with mean vector  $\mu$  and covariance matrix  $V$ , and by  $\phi_\beta$  the symmetric  $\beta$ -stable density corresponding to  $\mathcal{N}_1(0, \sigma^2)$  for  $\beta = 2$ , and to the Lévy density  $z \mapsto c|z|^{-1-\beta}$  for  $\beta < 2$ . This implies that

$$\phi_\beta(0) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}}, & \text{for } \beta = 2, \\ \frac{1}{\pi} \Gamma\left(1 + \frac{1}{\beta}\right) \left\{ \frac{2c}{\beta} \Gamma(1 - \beta) \cos \frac{\beta\pi}{2} \right\}^{-1/\beta}, & \text{for } \beta < 2. \end{cases} \quad (8)$$

Here, for  $\beta < 2$ , we used the fact  $\phi_\beta(0) = \pi^{-1}(\sigma')^{-1/\beta} \Gamma(1 + 1/\beta)$  with

$$\sigma' = \frac{2c}{\beta} \Gamma(1 - \beta) \cos \frac{\beta\pi}{2},$$

which comes from the relation

$$-\sigma'|u|^\beta = \int (\cos(uz) - 1) \frac{c}{|z|^{1+\beta}} dz, \quad u \in \mathbb{R}.$$

Next, let  $\Gamma_0$  and  $\Sigma_0$  be symmetric matrices defined by

$$\Sigma_0 := \int w(x)^2 \begin{pmatrix} x^2 & -x \\ -x & 1 \end{pmatrix} \pi_0(dx), \quad (9)$$

$$\Gamma_0 := 2\phi_\beta(0) \int w(x) \begin{pmatrix} x^2 & -x \\ -x & 1 \end{pmatrix} \pi_0(dx). \quad (10)$$

In view of Schwarz’s inequality, both  $\Sigma_0$  and  $\Gamma_0$  are always positive definite. Let

$$V_0 := \Gamma_0^{-1} \Sigma_0 \Gamma_0^{-1},$$

and let  $\rightarrow^d$  denote the weak convergence under  $P_0$ .

Now we are ready to state our asymptotic normality result. See Section 3.1 for the proof.

**Theorem 2.1.** Suppose (2), (3), and (4), and Assumptions 1, 2, and 3. Then

$$\sqrt{nh}^{1-1/\beta} (\hat{\theta}_n - \theta_0) \rightarrow^d \mathcal{N}_2(0, V_0).$$

Thus, unlike with the case of diffusions with compound Poisson jumps, for pure-jump  $Z$  we have no longer the typical rate  $\sqrt{nh}$  corresponding to  $\sqrt{T}$  in case where we have a continuous-time record  $(X_t)_{t \in [0, T]}$  with the asymptotics  $T \rightarrow \infty$ ; see, e.g., Luschgy [12, Section 5] and Sørensen [22]. The rate  $\sqrt{nh^{1-1/\beta}}$  reflects the degree of small-jump fluctuation of  $Z$  in conjunction with the sampling frequency  $1/h$ . It is worth mentioning that the rate of convergence becomes free of the sampling frequency for  $\beta = 1$ ; as mentioned before, this is the case for, e.g., any symmetric generalized hyperbolic  $Z$  with positive scale parameter.

For construction of asymptotic confidence intervals of  $\theta$ , we have to derive a consistent estimator of  $V_0$ . We return to this issue in Section 2.1.3 shortly after stating the moment convergence result.

### 2.1.2 Convergence of moments

Here we introduce

**Assumption 4.** (*Additional conditions for the convergence of moments*)

1. *There exists a constant  $\epsilon_0 \in (0, 1)$  such that  $nh \gtrsim n^{\epsilon_0}$ .*
2.  *$\limsup_{|x| \rightarrow \infty} \{w(x)|x|^4\}^k / |x|^q < \infty$  for any  $k > 0$ .*
3.  *$\beta \in (1, 2]$  and  $nh^{4(1-1/\beta)} \lesssim 1$ , but the latter can be removed if either  $\gamma_0$  or  $\lambda_0$  is known, so that the dimension of  $\theta$  is one.*

For Assumption 4.1, we may set  $h = n^{-\tau}$  with an appropriate  $\tau$  (see Section 2.2.3), while, for example, the choice  $h = n^{-1} \log n$  is not enough. Trivially, as in Assumption 2.2 we can remove Assumption 4.2, if we can make  $q$  in Assumption 1 arbitrarily large; otherwise, it suffices to take, e.g., any uniformly continuous function with compact support, or any  $w$  subgeometrically decreasing for  $|x| \rightarrow \infty$ . We need “provisory” Assumption 4.3 to handle asymptotically negligible martingale terms uniformly in the parameter, when proving the polynomial type large deviation inequality (see (32) below), which is of substantial importance in the proof of our moment convergence result. The proviso says that the dimension of unknown  $\theta$  comes into play in the uniform estimates of the martingale terms; such a phenomenon does not arise when the contrast function is smooth in  $\theta$  and sufficiently integrable.

We can get the convergence of moments in compensation for the additional assumptions. See Section 3.2 for the proof.

**Theorem 2.2.** *Suppose Assumption 4 in addition to the conditions imposed in Theorem 2.1. Then, for any continuous function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  at most polynomial growth,*

$$E_0 \left[ f \left( \sqrt{nh^{1-1/\beta}} (\hat{\theta}_n - \theta_0) \right) \right] \rightarrow \int f(u) \phi_2(u; 0, V_0) du,$$

where  $\phi_2(\cdot; 0, V_0)$  stand for the Gaussian density corresponding to  $\mathcal{N}_2(0, V_0)$ .

### 2.1.3 Interval estimation

Now we look at how to implement an interval estimation based on Theorem 2.1. The asymptotic covariance matrix  $V_0$  depends on the quantities

$$M(k, l) := \int x^k w(x)^l \pi_0(dx), \quad k \in \{0, 1, 2\}, \quad l \in \{1, 2\},$$

and  $(\beta, c)$  through  $\phi_\beta(0)$ . In order to make Theorem 2.1 usable in practice, we in principle have to estimate these quantities.



Let  $\check{\Gamma}_0 := \{2\phi_\beta(0)\}^{-1}\Gamma_0$  and  $U_0 := \check{\Gamma}_0^{-1}\Sigma_0\check{\Gamma}_0^{-1}$ , so that

$$V_0 = \{2\phi_\beta(0)\}^{-2}U_0.$$

Since  $U_0$  is written only by  $\{M(k, l)\}$ , we can readily get a consistent estimator  $\hat{U}_n$  of  $U_0$  by means of Lemma 3.5:

$$\hat{M}_n(k, l) := \frac{1}{n} \sum_{i=1}^n w_{i-1}^k X_{t_{i-1}}^l \rightarrow^p M(k, l).$$

On the other hand, as specified by (8), the remaining quantity  $\phi_\beta(0)$  depends only on the two parameters  $\beta$  and  $c$  (this point is completely different from Ling [11]). Nevertheless, direct consistent estimation of  $(\beta, c)$  seems rather difficult in general, for the full form of  $\nu$  here is not specified. In addition, even if we could get some consistent estimators  $(\hat{\beta}_n, \hat{c}_n)$ , we actually need to specify the rate of convergence of  $\hat{\beta}_n$  in order to successfully replace  $\sqrt{nh}^{1-1/\beta}$  with  $\sqrt{nh}^{1-1/\hat{\beta}_n}$  to get the desired asymptotically standard normal version  $2\phi_{\hat{\beta}_n}(0)\hat{U}_n^{-1/2}\sqrt{nh}^{1-1/\hat{\beta}_n}(\hat{\theta}_n - \theta_0) \rightarrow^d \mathcal{N}_2(0, I_2)$ , where  $I_2$  denotes the two-dimensional identity matrix; namely, we need to have  $h^{1/\beta-1/\hat{\beta}_n} \rightarrow^p 1$ . In Theorem 2.3 given below, we show that an appropriate use of kernel estimator, which was also used by Ling [11], enables us to overcome this annoying aspects. Specifically, we show how to provide a consistent estimator of  $\phi_\beta(0)$  and then formulate a converted distributional result with asymptotic standard normal distribution, which can be used without direct estimate of  $(\beta, c)$ .

**Theorem 2.3.** *Suppose the conditions imposed in Theorem 2.1. Let  $K : \mathbb{R} \mapsto \mathbb{R}_+$  be a globally Lipschitz function such that*

$$\int K(z)dz = 1 \quad \text{and} \quad \int |z|K(z)dz \vee \int \{K(z)\}^p dz < \infty$$

for some constant  $p > 2$ , let  $(B_n)$  be a positive sequence such that

$$B_n = o(1), \quad nB_n \rightarrow \infty, \quad \liminf_{n \rightarrow \infty} nB_n^2 > 0, \quad (11)$$

and define

$$\hat{\phi}_\beta(0)_n := \left(B_n \sum_{i=1}^n w_{i-1}\right)^{-1} \sum_{i=1}^n w_{i-1} K\left(\frac{\Delta_i X - h(\hat{\gamma}_n - \hat{\lambda}_n X_{t_{i-1}})}{h^{1/\beta} B_n}\right).$$

Then  $\hat{\phi}_\beta(0)_n \rightarrow^p \phi_\beta(0)$ . In particular, additionally supposing

$$\liminf_{n \rightarrow \infty} nh^{2-1/\beta} > 0, \quad (12)$$

we have the  $(\beta, c)$ -free version of the asymptotic normality:

$$\frac{2 \sum_{i=1}^n w_{i-1} K\left(\frac{\Delta_i X - h(\hat{\gamma}_n - \hat{\lambda}_n X_{t_{i-1}})}{\sqrt{nh}^2}\right)}{h \sum_{i=1}^n w_{i-1}} \hat{U}_n^{-1/2}(\hat{\theta}_n - \theta_0) \rightarrow^d \mathcal{N}_2(0, I_2). \quad (13)$$

See Section 3.3 for the proof of Theorem 2.3. For the kernel function  $K$ , we can adopt the standard-Gaussian kernel  $K(z) = (2\pi)^{-1/2}e^{-z^2/2}$  or, as used by Ling [11], the logistic kernel  $K(z) = e^{-z}/(1 + e^{-z})^2$ . See Section 2.2.3 for some discussions on the conditions we have imposed so far on the sampling design (asymptotic behavior of  $h$ ).

### 2.1.4 Case of nonnull Gaussian part

Just for reference, we single out the nonnull Gaussian-part case as a corollary of the previous results.

**Corollary 2.4.** *Suppose that  $\sigma^2 > 0$ , that  $nh^3 \rightarrow 0$ , and let (2), (3), (4), Assumption 1.1, and Assumption 2 be in force. Then the following holds true.*

1. *We have*

$$\sqrt{nh}(\hat{\theta}_n - \theta_0) \rightarrow^d \mathcal{N}_2\left(0, \frac{\pi\sigma^2}{2}U_0\right).$$

2. *Suppose that  $\limsup_{n \rightarrow \infty} nh^2 < \infty$  if both  $\lambda_0$  and  $\gamma_0$  are actually unknown; no additional condition is made especially if either  $\lambda_0$  or  $\gamma_0$  is known from the beginning. Then, for any continuous function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  at most polynomial growth,*

$$E_0 \left[ f \left( \sqrt{nh}(\hat{\theta}_n - \theta_0) \right) \right] \rightarrow \int f(u) \phi_2 \left( u; 0, \frac{\pi\sigma^2}{2}U_0 \right) du.$$

3. *We have (13) if  $\liminf_{n \rightarrow \infty} nh^{3/2} > 0$ .*

Corollary 2.4 directly follows from the previous claims.

### 2.1.5 Some future issues

In the literature, we could find no previous work concerning LAD type estimation for discretely observed processes with infill asymptotics. Here we mention some future issues worth being considered, together with some conjectures.

- “*What is a proper definition of asymptotic efficiency in the present framework?*”  
Concerning the statistical model in question, we want to derive the local asymptotic normality in principle. Nevertheless, we could see that the LAN cannot hold true at least for the non-Gaussian stable  $Z$ , as in the case of infinite-variance autoregressive time series models (see Davis *et al.* [3] for details); for the stable case we conjecture that the best attainable one for estimating  $\lambda$  is the much faster  $n^{1/\beta}h^{1-1/\beta}$  than  $\sqrt{nh}^{1-1/\beta}$  of our SLAD estimators. On the other hand, we conjecture that  $\sqrt{nh}^{1-1/\beta}$  is the best attainable rate of convergence for estimating  $\gamma$ .
- “*Is it possible to relax the ergodicity and the long-term asymptotics?*”  
We have set  $nh \rightarrow \infty$  and focused on the ergodic case. Nevertheless, our SLAD estimator does seem to work even when  $X$  is non-recurrent with or without imposing that  $nh \rightarrow \infty$  (e.g.,  $nh$  is a fixed positive constant, but in this case we need  $\beta < 2$ ); needless to say, the limit distribution may be then no longer normal. For specific derivation of the limit distribution when  $nh$  fixed, we would need a more sophisticated weak limit theorem than the martingale central limit theorem used in the present proof; it would be nice if we could derive a tailor-made stable convergence in law, leading to a mixed normal limit distribution with specified limit random covariance matrix.
- “*What is occurring in the higher order part?*”  
For example, Knight [8, Section 4.2] (see also the references therein) discussed this issue in case of linear regression models. It would be interesting to investigate this point in the framework of discretely observed OU processes.
- “*What will occur for small  $\beta$ ?*”  
Overall, our assumptions require that the index  $\beta$  is large. Concerning Theorem 2.1,

it is expected from the proof (see Section 3.1) that we may relax the sampling-design condition Assumption 3 by targeting the “genuine” SLAD estimator defined to be

$$\arg \inf_{\theta=(\lambda,\gamma)\in\Theta} \left\{ \sum_{i=1}^n w(X_{t_{i-1}}) \left| X_{t_i^n} - e^{-\lambda h_n} X_{t_{i-1}^n} - \frac{\gamma}{\lambda} (1 - e^{-\lambda h_n}) \right| \right\},$$

which is a little bit more involved than ours, but should be appropriate in view of the expression (17). Furthermore, even if this estimating function works properly, it still excludes the case of  $\beta = 0$ , e.g., the variance gamma and the bilateral gamma  $Z$  as well as the purely compound Poisson  $Z$ : in such cases the local-limit result given by Lemma 3.3 below, which is essential in our proofs, breaks down. Therefore we have to resort to a whole other kind of estimating procedure. See also Section 2.2.4.

We leave answering such questions to future works.

## 2.2 Remarks on the results

Here we gather some technical remarks concerning the results given in Section 2.1.

### 2.2.1 On the asymptotic covariance matrix

In general, we do not know which  $w$  optimizes the asymptotic covariance matrix  $V_0$ . Nevertheless,  $V_0$  can be simplified and actually optimized in some instances. Let  $m_k := \int x^k \pi_0(dx)$  and denote by  $v^2$  the variance of  $\pi_0$  if they exist; specifically, by (6), we have  $m_1 = \gamma_0/\lambda_0$  and  $v^2 = \sigma^2/(2\lambda_0) + \int z^2 \int_0^\infty \nu(e^{\lambda_0 s} dz) dz$ .

1. We can take  $w \equiv 1$  if  $q \geq 4$ , so that

$$V_0 = \frac{1}{\{2\phi_\beta(0)\}^2 v^2} \begin{pmatrix} 1 & m_1 \\ m_1 & m_2 \end{pmatrix}. \quad (14)$$

If additionally  $\gamma_0 = 0$ , then  $V_0$  becomes diagonal, hence the SLAD estimation of  $\lambda_0$  and  $\gamma_0$  are asymptotically independent, as in the case of OU diffusions.

2. On the other hand, suppose that  $\gamma_0 = 0$  and  $w$  is symmetric around zero, while we now do not assume  $q \geq 4$ . Then Assumption 2 entails that  $\int w(x)x\pi_0(dx) = \int w(x)^2 x \pi_0(dx) = 0$ , hence

$$V_0 = \frac{1}{\{2\phi_\beta(0)\}^2} \text{diag} \left\{ \frac{\int w(x)^2 x \pi_0(dx)}{\{\int w(x)x\pi_0(dx)\}^2}, \frac{\int w(x)^2 \pi_0(dx)}{\{\int w(x)\pi_0(dx)\}^2} \right\}.$$

If further  $m_2 < \infty$ , then Schwarz’s inequality readily gives

$$V_0 \geq \frac{1}{\{2\phi_\beta(0)\}^2} \text{diag}(1/m_2, 1),$$

the lower bound attained for (any positive) constant  $w$ .

### 2.2.2 Comparisons with respect to the LSE

When  $nh^2 \rightarrow 0$  and  $q > 0$  in Assumption 1 can be taken large enough, we can deduce the asymptotic normality of the LSE  $\tilde{\theta}_n$ , namely,  $\sqrt{nh}(\tilde{\theta}_n - \theta_0) \rightarrow^d \mathcal{N}_2(0, \tilde{V}_0)$  with

$$\tilde{V}_0 := \frac{\text{Var}[Z_1]}{v^2} \begin{pmatrix} 1 & m_1 \\ m_1 & m_2 \end{pmatrix}.$$

(See Masuda [14] for details.) Now suppose that  $w \equiv 1$  and  $\beta = 2$ , so that, as in the LSE, the SLAD estimator is asymptotically normal at rate  $\sqrt{nh}$ . Then, the  $\tilde{V}_0$  compared with

(14) implies that the asymptotic relative efficiency of the SLAD estimator with respect to the LSE can be measured by the quantity

$$r(\sigma^2, \nu) := \{2\phi_2(0)\}^2 \text{Var}[Z_1] = \frac{2}{\pi} \left( 1 + \frac{\int z^2 \nu(dz)}{\sigma^2} \right).$$

From this aspect, the SLAD estimator turns out to be asymptotically superior to the LSE if the Gaussian variance is not so large compared with the jump-part variance:

$$r(\sigma^2, \nu) > 1 \iff \frac{2}{\pi - 2} \int z^2 \nu(dz) > \sigma^2.$$

In other words, the LSE is asymptotically superior to the SLAD estimator if  $\sigma^2$  is dominant in the sense that the last inequality sign is reversed.

The SLAD estimation is formally new even for the Gaussian case. Suppose  $X$  is given by

$$dX_t = -\lambda X_t dt + \sigma dw_t,$$

where  $w$  is a standard Wiener process. As is well known in the literature, or as can be seen from direct computations, the exact maximum likelihood estimator of  $\lambda_0$  is asymptotically normal and efficient with asymptotic variance being  $2\lambda_0$ . On the other hand, building on Section 2.2.1 we see that the unweighted SLAD estimator  $\hat{\lambda}_n$  leads to the asymptotic variance  $\pi\lambda_0$ , hence the asymptotic efficiency of  $\hat{\lambda}_n$  relative to the maximum likelihood estimator is  $2/\pi$ ; this asymptotic relative efficiency is the same as in the one for the sample median over the sample mean in estimating the mean of i.i.d. normal samples. Moreover, for the asymptotic normality of  $\hat{\lambda}_n$  we do not need the rapidly increasing experimental design  $nh^2 \rightarrow 0$ , which is quite often inevitable when adopting a contrast function based on the naive “Euler-type approximation”; for the SLAD estimator, the weaker sampling design  $nh^3 \rightarrow 0$  is sufficient in view of Corollary 2.4.

Recently, Hu and Long [7] derived an asymptotic distributional result concerning the (approximate) LSE of  $\lambda_0 > 0$ , presupposing that  $\gamma_0 = 0$  and letting  $Z$  be a symmetric  $\beta$ -stable Lévy process with  $\beta \in (1, 2)$ ; in this setting, we can ignore Assumption 4.3 for our SLAD estimator  $\hat{\lambda}_n$  from the very beginning. The LSE is given by

$$\tilde{\lambda}_n := \arg \inf_{\lambda > 0} \sum_{i=1}^n (\Delta_i X + \lambda X_{t_{i-1}} h)^2 = -\frac{\sum_{i=1}^n X_{t_{i-1}} \Delta_i X}{h \sum_{i=1}^n X_{t_{i-1}}^2}.$$

Let us make some comparisons between  $\tilde{\lambda}_n$  with  $\hat{\lambda}_n$ . The primary point is the differences in the rates of convergence and the limit distributions: their  $\tilde{\lambda}_n$  fulfil that

$$\left( \frac{nh}{\log n} \right)^{1/\beta} (\tilde{\lambda}_n - \lambda_0) \rightarrow^d \frac{S}{S_+},$$

where  $nh \rightarrow \infty$ , and  $S \in \mathbb{R}$  and  $S_+ > 0$  are two independent strictly stable random variables with specified parameters. In order to obtain the explicit asymptotic distribution, Hu and Long [7] also imposed some technical conditions on the decreasing rate of  $h \rightarrow 0$  in connection with the value of  $\beta$ , while they are not necessary for the (strong) consistency; see their (A1). As

$$\frac{(nh/\log n)^{1/\beta}}{\sqrt{nh}^{1-1/\beta}} = \frac{(nh^2)^{(2-\beta)/(2\beta)}}{(\log n)^{1/\beta}},$$

we see that the SLAD estimator  $\hat{\lambda}_n$  with appropriate  $w$  (tapering at infinity) converges more rapidly than  $\tilde{\lambda}_n$  as soon as

$$h = o\left(\frac{1}{\sqrt{n}}(\log n)^{1/(2-\beta)}\right).$$

In particular, this is the case if  $nh^2 \lesssim 1$ . Moreover, in case where  $h = n^{-a}$  for some constant  $a > 0$ , Hu and Long [7, Remark 3.3] mentioned that the choice  $a = (1 + \beta)^{-1}$  is optimal. This choice implies that  $nh^2 \rightarrow \infty$ , and we can see that  $\tilde{\lambda}_n$  now converges more rapidly than  $\hat{\lambda}_n$ . Nevertheless, our  $\hat{\lambda}_n$  would be more convenient to use because of its asymptotic normality. Moreover, contrary to our  $\hat{\lambda}_n$ , the convergence of moments seems impossible for  $\tilde{\lambda}_n$ .

### 2.2.3 On the sampling-design conditions in Theorem 2.3

In the statement of Theorem 2.3 we have several conditions on the decreasing rate of  $h$ . The conditions can be more specific when  $h = n^{-\tau}$  with the choice  $l_n = \sqrt{nh^2}$ , where  $\tau \in (0, 1)$  for (3) to be fulfilled. In this case Assumption 3 is equivalent to  $\beta/\{2(2\beta - 1)\} < \tau$ , while the condition  $\liminf_{n \rightarrow \infty} nh^{2-1/\beta} > 0$  to  $\tau \leq \beta/(2\beta - 1)$ . Thus, the admissible region of  $\tau$  for the “interval estimation (13)” turns out to be

$$\begin{cases} \frac{\beta}{2(2\beta-1)} < \tau \leq \frac{\beta}{2\beta-1}, & \text{if } \beta > 1, \\ \frac{\beta}{2(2\beta-1)} < \tau < 1, & \text{if } \beta \in (2/3, 1]. \end{cases} \quad (15)$$

For the convergence of moments, we additionally need  $\beta > 1$  and also the condition  $nh^{4(1-1/\beta)} \lesssim 1$  if both  $\lambda$  and  $\gamma$  are actually unknown (see Assumption 4). The condition  $nh^{4(1-1/\beta)} \lesssim 1$  is equivalent to  $\tau \geq \beta/\{4(\beta - 1)\}$  ( $> \beta/\{2(2\beta - 1)\}$ , always); in view of (15), we need  $\beta/\{4(\beta - 1)\} \leq \beta/\{2\beta - 1\} \iff \beta \geq 3/2$  in order to make the admissible region non-empty. Thus, the admissible region of  $\tau$  for the “convergence of moments” is

$$\begin{cases} \frac{\beta}{2(2\beta-1)} < \tau \leq \frac{\beta}{2\beta-1}, & \text{if either } \lambda \text{ or } \gamma \text{ is known and } \beta > 1, \\ \frac{\beta}{4(\beta-1)} \leq \tau \leq \frac{\beta}{2\beta-1}, & \text{if both } \lambda \text{ and } \gamma \text{ unknown and } \beta \geq 3/2. \end{cases} \quad (16)$$

In particular, the “boundary” case in (16), where  $\beta = 3/2$  with both  $\lambda$  and  $\gamma$  being unknown, leads to the only possibility  $\tau = 3/4$  making both the convergence (13) and the convergence of moments valid.

### 2.2.4 Model extension

We have used some inherent property of the OU processes in our proofs, hence, unfortunately, it is not clear whether or not a similar type of contrast function actually works for more general nonlinear-drift stochastic differential equation models with jumps. Nevertheless, a more general statement is formally possible so that we can provide a set of conditions for the convergence of moments and so on to hold true for a broader class of statistical experiments of dependent data as well as independent ones (not necessarily a stochastic differential equation). This may be done by setting the contrast function as

$$\theta \mapsto \sum_{i=1}^n \bar{w}_{n,i-1} |\bar{X}_{ni} - \theta^\top \bar{Y}_{n,i-1}|,$$

where  $\bar{w}_{n,i-1}$  and  $\bar{Y}_{n,i-1}$  are  $\mathcal{G}_{n,i-1}$ -measurable while  $\bar{X}_{ni}$  is  $\mathcal{G}_{ni}$ -measurable with respect to some underlying filtration  $(\mathcal{G}_{ni})_{i \leq n}$ ; all of  $\bar{w}_{n,i-1}$ ,  $\bar{Y}_{n,i-1}$  and  $\bar{X}_{n,i}$  should be observable in order to follow a similar line to the proofs given in this paper. This setting might allow us to deal with, for instance, discretely observed Lévy process, general i.i.d. regression model, and also the autoregressive time-series in a unified way. Of course, the rate of convergence as well as the limit distribution should depend on the specific structure of the underlying statistical model.

### 3 Proofs

We denote by  $(\Omega, \mathcal{F}, P)$  the underlying probability space on which  $(X_0, Z)$  is defined, and by  $E$  the corresponding expectation operator. Throughout this section, we use some basic notation:  $C$  denotes a positive generic constant, possibly varying from line to line;  $A^\top$  denotes the transpose of a matrix  $A$ ;  $A^{\otimes 2} := AA^\top$  for any matrix  $A$ ; we write  $a'_n \lesssim a''_n$  for nonnegative sequences  $a'_n$  and  $a''_n$ , if  $a'_n \leq Ca''_n$  a.s. for every  $n$  large enough;  $w_{i-1} := w(X_{t_{i-1}})$ ;  $P_0^{i-1}[\cdot] := P_0[\cdot | \mathcal{F}_{t_{i-1}}]$ , where  $\mathcal{F}_t := \sigma(X_0, Z_s; s \leq t)$ ; finally,  $\rightarrow^p$  denotes the convergence in probability under  $P_0$ .

#### 3.1 Proof of Theorem 2.1

Under  $P_0$  we have the autoregressive representation

$$X_{t_i} = \frac{\gamma_0}{\lambda_0}(1 - e^{-\lambda_0 h}) + e^{-\lambda_0 h} X_{t_{i-1}} + \int_{t_{i-1}}^{t_i} e^{-\lambda_0(t_i-s)} dZ_s \quad (17)$$

for  $n \in \mathbb{N}$  and  $i \leq n$ . For convenience we write

$$\begin{aligned} \epsilon_{ni} &= h^{-1/\beta} \int_{t_{i-1}}^{t_i} e^{-\lambda_0(t_i-s)} dZ_s, \\ \epsilon'_{n,i-1} &= h^{-1/\beta} \left( X_{t_{i-1}} - \frac{\gamma_0}{\lambda_0} \right) (e^{-\lambda_0 h} - 1 + \lambda_0 h), \\ x_{i-1} &= (-X_{t_{i-1}}, 1)^\top, \end{aligned}$$

so that

$$h^{-1/\beta} \{ \Delta_i X - (\gamma - \lambda X_{t_{i-1}})h \} = (\epsilon'_{n,i-1} + \epsilon_{ni}) - (\theta - \theta_0)^\top x_{i-1} h^{1-1/\beta}. \quad (18)$$

It follows from the definition (7) that, for each  $\theta \in \Theta$ ,

$$\begin{aligned} & h^{-1/\beta} \{ M_n(\theta) - M_n(\theta_0) \} \\ &= \sum_{i=1}^n w_{i-1} \left\{ |(\epsilon'_{n,i-1} + \epsilon_{ni}) - (\theta - \theta_0)^\top x_{i-1} h^{1-1/\beta}| - |\epsilon'_{n,i-1} + \epsilon_{ni}| \right\}. \end{aligned} \quad (19)$$

In our proof, it is crucial that  $\epsilon'_{n,i-1}$  is  $\mathcal{F}_{t_{i-1}}$ -measurable.

Let  $\mathbb{U}_n(\theta_0) := \{u \in \mathbb{R}^2 : \theta_0 + a_n u \in \Theta\}$  with

$$a_n = a_n(\beta) := (\sqrt{n} h^{1-1/\beta})^{-1}.$$

Then we define the random fields  $\mathbb{Z}_n(\cdot; \theta_0) : \mathbb{U}_n(\theta_0) \times \Omega \rightarrow (0, \infty)$  by

$$\log \mathbb{Z}_n(u; \theta_0) = -h^{-1/\beta} \{ M_n(\theta_0 + a_n u) - M_n(\theta_0) \}, \quad (20)$$

its maximizer equalling  $\hat{u}_n := a_n^{-1}(\hat{\theta}_n - \theta_0)$ . To achieve the proof, we are going to derive the following asymptotically locally quadratic structure of  $\log \mathbb{Z}_n(u; \theta_0)$  for each  $u \in \mathbb{U}_n(\theta_0)$ :

$$\log \mathbb{Z}_n(u; \theta_0) = u^\top \Delta_n - \frac{1}{2} u^\top \Gamma_n u + o_p(1), \quad (21)$$

where  $\Delta_n \rightarrow^d \mathcal{N}(0, \Sigma_0)$  and  $\Gamma_n \rightarrow^p \Gamma_0$  for positive definite nonrandom matrices  $\Sigma_0$  and  $\Gamma_0$  given by (9) and (10), respectively. Then, in view of the convexity of  $u \mapsto -\log \mathbb{Z}_n(u; \theta_0)$ , Theorem 2.1 follows on applying Hjrt and Pollard [6, BASIC COROLLARY].

First we rewrite  $\log \mathbb{Z}_n(u; \theta_0)$ . For any function  $K$  of the form  $K(x) = \int_0^x k(y)dy$ , we have  $K(x-y) - K(x) = -yk(x) + \int_0^y \{k(x) - k(x-s)\}ds$ . As in Knight [8], taking  $k(y) = I(y \geq 0) - I(y \leq 0)$  so that  $K(x) = |x|$ , we make use of the following identity valid for any  $x \neq 0$  and  $y \in \mathbb{R}$ :

$$|x-y| - |x| = -y\{I(x > 0) - I(x < 0)\} + 2 \int_0^y \{I(x \leq s) - I(x \leq 0)\}ds. \quad (22)$$

From Lemma 3.3 below, we have  $P[\epsilon_{ni} \neq 0] = 1$  for each  $n \in \mathbb{N}$  and  $i \leq n$ . Combining (19), (20), and (22) yields  $\log \mathbb{Z}_n(u; \theta_0) = \mathbb{L}_n(u) + \mathbb{Q}_n(u)$ ,  $P_0$ -a.s., where

$$\begin{aligned} \mathbb{L}_n(u) &:= u^\top \sum_{i=1}^n \frac{1}{\sqrt{n}} w_{i-1} x_{i-1} \{I(\epsilon'_{n,i-1} + \epsilon_{ni} > 0) - I(\epsilon'_{n,i-1} + \epsilon_{ni} < 0)\}, \\ \mathbb{Q}_n(u) &:= -2 \sum_{i=1}^n w_{i-1} \int_0^{u^\top x_{i-1}/\sqrt{n}} \{I(\epsilon'_{n,i-1} + \epsilon_{ni} \leq s) - I(\epsilon'_{n,i-1} + \epsilon_{ni} \leq 0)\}ds. \end{aligned}$$

Write  $\mathbb{L}_n(u) = u^\top \sum_{i=1}^n l_{ni}$  and  $\mathbb{Q}_n(u) = \sum_{i=1}^n q_{ni}(u)$ . Both  $\mathbb{L}_n(u)$  and  $\mathbb{Q}_n(u)$  entail a leading term plus some remainder terms. We look at them separately.

*Asymptotic behavior of  $\mathbb{L}_n(u)$ .* We decompose  $\mathbb{L}_n(u)$  as

$$\mathbb{L}_n(u) = u^\top \sum_{i=1}^n (l_{ni} - E_0^{i-1}[l_{ni}]) + u^\top \sum_{i=1}^n E_0^{i-1}[l_{ni}] =: u^\top \sum_{i=1}^n \Delta_{ni} + R_n^1(u). \quad (23)$$

Denote the distribution of  $\epsilon_{ni}$  by

$$F_h(dz) = p_h(z)dz,$$

which is independent of  $i$  (see Lemma 3.3). Since  $p_h$  is symmetric around 0,

$$\begin{aligned} E_0^{i-1}[\Delta_{ni}^{\otimes 2}] &= w_{i-1}^2 x_{i-1}^{\otimes 2} \left\{ P_0^{i-1}[\epsilon'_{n,i-1} + \epsilon_{ni} \neq 0] \right. \\ &\quad \left. - (P_0^{i-1}[\epsilon'_{n,i-1} + \epsilon_{ni} > 0] - P_0^{i-1}[\epsilon'_{n,i-1} + \epsilon_{ni} < 0])^2 \right\} \\ &= w_{i-1}^2 x_{i-1}^{\otimes 2} \left\{ 1 - \left( 2 \int_0^{|\epsilon'_{n,i-1}|} p_h(z)dz \right)^2 \right\}. \end{aligned}$$

Moreover, since

$$|\epsilon'_{n,i-1}| \lesssim h^{2-1/\beta} (1 + |X_{t_{i-1}}|) \quad (24)$$

and  $p_h$  is bounded, we have  $|\int_0^{|\epsilon'_{n,i-1}|} p_h(z)dz|^2 \lesssim h^{2(2-1/\beta)} (1 + |X_{t_{i-1}}|)^2$ . Thus

$$\begin{aligned} \left| \sum_{i=1}^n E_0^{i-1}[\Delta_{ni}^{\otimes 2}] - \frac{1}{n} \sum_{i=1}^n w_{i-1}^2 x_{i-1}^{\otimes 2} \right| &\lesssim h^{2(2-1/\beta)} \frac{1}{n} \sum_{i=1}^n w_{i-1}^2 (1 + |X_{t_{i-1}}|)^4 \\ &= O_p(h^{2(2-1/\beta)}) = o_p(1), \end{aligned}$$

so that Lemma 3.5 and ergodic theorem yield

$$\sum_{i=1}^n E_0^{i-1}[\Delta_{ni}^{\otimes 2}] \rightarrow^p \Sigma_0. \quad (25)$$

Also, under the assumptions it is easy to see that for  $a \in (0, 2]$

$$\begin{aligned} \sum_{i=1}^n E_0[|\Delta_{ni}|^{2+a}] &\lesssim n^{-a/2} \sup_t E_0[\{w(X_t)(1 + |X_t|)\}^{2+a}] \\ &\lesssim n^{-a/2} \sup_t E_0[w(X_t)(1 + |X_t|)^4] = O(n^{-a/2}) = o(1). \end{aligned} \quad (26)$$

From (25) and (26) we can apply the martingale central limit theorem for

$$\Delta_n := \sum_{i=1}^n \Delta_{ni}$$

to obtain  $\Delta_n \rightarrow^d \mathcal{N}_1(0, \Sigma_0)$ . As for  $R_n^1(u)$ ,

$$\begin{aligned} |R_n^1(u)| &= \left| \frac{u^\top}{n} \sum_{i=1}^n w_{i-1} x_{i-1} \sqrt{n} \{1 - 2F_h(-\epsilon'_{n,i-1})\} \right| \\ &= \left| \frac{2u^\top}{n} \sum_{i=1}^n w_{i-1} x_{i-1} \sqrt{n} \epsilon'_{n,i-1} \int_0^1 p_h(-y \epsilon'_{n,i-1}) dy \right| \\ &\lesssim |u| \sqrt{n} h^{2-1/\beta} \frac{1}{n} \sum_{i=1}^n w_{i-1} (1 + |X_{t_{i-1}}|)^2 = |u| O(\sqrt{n} h^{2-1/\beta}) = o(1), \end{aligned} \quad (27)$$

where we used  $1 - 2F_h(0) = 0$  for the second equality. Thus  $\mathbb{L}_n(u) = u^\top \Delta_n + o_p(1)$  with  $\Delta_n \rightarrow^d \mathcal{N}_1(0, \Sigma_0)$  for each  $u$ , as was desired.

*Asymptotic behavior of  $\mathbb{Q}_n(u)$ .* Again we separate the martingale term:

$$\mathbb{Q}_n(u) = \sum_{i=1}^n E_0^{i-1}[q_{ni}(u)] + \sum_{i=1}^n \{q_{ni}(u) - E_0^{i-1}[q_{ni}(u)]\}.$$

For the first term of the right-hand side, Taylor's formula gives

$$\begin{aligned} \sum_{i=1}^n E_0^{i-1}[q_{ni}(u)] &= -2 \sum_{i=1}^n w_{i-1} \int_0^{u^\top x_{i-1}/\sqrt{n}} \{F_h(s - \epsilon'_{n,i-1}) - F_h(-\epsilon'_{n,i-1})\} ds \\ &= -\frac{1}{2} u^\top \left\{ 2p_h(0) \frac{1}{n} \sum_{i=1}^n w_{i-1} x_{i-1}^{\otimes 2} \right\} u \\ &\quad + \left[ -\frac{1}{n} \sum_{i=1}^n w_{i-1} (u^\top x_{i-1})^2 \{p_h(-\epsilon'_{n,i-1}) - p_h(0)\} \right. \\ &\quad \left. - 2 \sum_{i=1}^n w_{i-1} \int_0^{u^\top x_{i-1}/\sqrt{n}} s^2 \int_0^1 (1-y) \partial p_h(sy - \epsilon'_{n,i-1}) dy ds \right] \\ &=: -\frac{1}{2} u^\top \Gamma_n u + R_n^2(u), \quad \text{say.} \end{aligned} \quad (28)$$

We have  $\Gamma_n \rightarrow^p \Gamma_0$  by means of Lemmas 3.3 and 3.5. To deal with  $R_n^2(u)$ , we note that: first,  $|\partial p_h(z)| = |\partial p_h(z) - \partial p_h(0)| \lesssim |z|$ , which follows from the first half of the proof of Lemma 3.3; second,  $|\int_0^x g(y) dy| \leq \int_0^{|x|} \{|g(y)| \vee |g(-y)|\} dy$  for any  $x \in \mathbb{R}$  and  $g : \mathbb{R} \rightarrow \mathbb{R}$ . Using these facts, we derive the following estimates:

$$\begin{aligned} |R_n^2(u)| &\lesssim |u|^2 \frac{1}{n} \sum_{i=1}^n w_{i-1} |x_{i-1}|^2 |\epsilon'_{n,i-1}|^2 + \sum_{i=1}^n w_{i-1} \int_0^{|u||x_{i-1}|/\sqrt{n}} s^2 (s + |\epsilon'_{n,i-1}|) ds \\ &\lesssim \left( h^{2(2-1/\beta)} |u|^2 + \frac{h^{2-1/\beta} |u|^3}{\sqrt{n}} + \frac{|u|^4}{n} \right) \frac{1}{n} \sum_{i=1}^n w_{i-1} (1 + |X_{t_{i-1}}|)^4 \\ &= O_p \left( h^{2(2-1/\beta)} |u|^2 + \frac{h^{2-1/\beta} |u|^3}{\sqrt{n}} + \frac{|u|^4}{n} \right) = o_p(1). \end{aligned} \quad (29)$$

Thus  $\sum_{i=1}^n E_0^{i-1}[q_{ni}(u)] \rightarrow^p -u^\top \Gamma_0 u/2$  for each  $u$ . It remains to show that the martingale part

$$R_n^3(u) := \sum_{i=1}^n \{q_{ni}(u) - E_0^{i-1}[q_{ni}(u)]\} \quad (30)$$



is  $o_p(1)$  for each  $u$ . This readily follows on applying Burkholder's and Schwarz's inequalities:

$$\begin{aligned}
E_0[\{R_n^3(u)\}^2] &\lesssim \sum_{i=1}^n E_0[\{q_{ni}(u) - E_0^{i-1}[q_{ni}(u)]\}^2] \\
&= \sum_{i=1}^n E_0 \left[ w_{i-1}^2 \left\{ \int_0^{u^\top x_{i-1}/\sqrt{n}} (I(\epsilon'_{n,i-1} + \epsilon_{ni} \leq s) - I(\epsilon'_{n,i-1} + \epsilon_{ni} \leq 0)) ds \right\}^2 \right] \\
&\leq \sum_{i=1}^n E_0 \left[ w_{i-1}^2 \left\{ \int_0^{|u||x_{i-1}|/\sqrt{n}} I(|\epsilon'_{n,i-1} + \epsilon_{ni}| \leq s) ds \right\}^2 \right] \\
&\leq \sum_{i=1}^n E_0 \left[ w_{i-1}^2 \frac{|u||x_{i-1}|}{\sqrt{n}} \int_0^{|u||x_{i-1}|/\sqrt{n}} I(-s - \epsilon'_{n,i-1} \leq \epsilon_{ni} \leq s - \epsilon'_{n,i-1}) ds \right] \\
&\leq \sum_{i=1}^n E_0 \left[ w_{i-1}^2 \frac{|u||x_{i-1}|}{\sqrt{n}} \int_0^{|u||x_{i-1}|/\sqrt{n}} \left( \int_{-s-\epsilon'_{n,i-1}}^{s-\epsilon'_{n,i-1}} p_h(z) dz \right) ds \right] \\
&\lesssim \frac{|u|^3}{\sqrt{n}} \sup_{t \in \mathbb{R}_+} E_0[w(X_t)^2(1 + |X_t|)^3] \lesssim \frac{|u|^3}{\sqrt{n}} \rightarrow 0.
\end{aligned} \tag{31}$$

Summarizing the above yields that  $\mathbb{Q}_n(u) = -u^\top \Gamma_n u/2 + o_p(1)$  for each  $u$ , with  $\Gamma_n \rightarrow^p \Gamma_0$ .

Combining the two steps leads to (21), hence the claim of Theorem 2.1.

### 3.2 Proof of Theorem 2.2

We keep using the notation introduced in the proof of Theorem 2.1. As we have already derived the asymptotic normality  $\hat{u}_n \rightarrow^d \mathcal{N}_2(0, V_0)$ , it suffices to ensure the  $L^p(P_0)$ -boundedness of  $(\hat{u}_n)_{n \in \mathbb{N}}$  for every  $p > 0$ . Suppose that, given any  $L > 0$  there exist constant  $C_L > 0$  and such that for every  $r > 0$  large enough

$$\sup_{n \in \mathbb{N}} P_0 \left[ \sup_{u \in \mathbb{U}_n(\theta_0): r \leq |u|} \mathbb{Z}_n(u; \theta_0) \geq e^{-r} \right] \leq \frac{C_L}{r^L}. \tag{32}$$

Then the desired  $L^p(P_0)$ -boundedness follows: for every  $r > 0$  large enough,

$$\begin{aligned}
\sup_{n \in \mathbb{N}} P_0[|\hat{u}_n| > r] &\leq \sup_{n \in \mathbb{N}} P_0 \left[ \sup_{|u| > r} \mathbb{Z}_n(u; \theta_0) \geq \mathbb{Z}_n(0) \right] \\
&= \sup_{n \in \mathbb{N}} P_0 \left[ \sup_{|u| > r} \mathbb{Z}_n(u; \theta_0) \geq 1 \right] \leq \frac{C_L}{r^L}.
\end{aligned}$$

We are going to derive the polynomial type large deviation inequality (32) by applying Yoshida [25].

From the proof of Theorem 2.1, we know that  $\log \mathbb{Z}_n(u; \theta_0) = u^\top \Delta_n - u^\top \Gamma_n u/2 + \sum_{j=1}^3 R_n^j(u)$ ; see (23), (28) and (30). Rearrange this as

$$\log \mathbb{Z}_n(u; \theta_0) = u^\top \Delta_n - \frac{1}{2} u^\top \Gamma_0 u + \sum_{j=1}^5 R_n^j(u), \tag{33}$$

where

$$\begin{aligned}
R_n^4(u) &:= -u^\top \{p_h(0) - p_0(0)\} \frac{1}{n} \sum_{i=1}^n w_{i-1} x_{i-1}^{\otimes 2} u, \\
R_n^5(u) &:= -u^\top p_0(0) \left\{ \frac{1}{n} \sum_{i=1}^n w_{i-1} x_{i-1}^{\otimes 2} - \int w(x) \begin{pmatrix} x^2 & -x \\ -x & 1 \end{pmatrix} \pi_0(dx) \right\} u.
\end{aligned}$$

The martingale part  $\Delta_n$  is  $L^M(P_0)$ -bounded for any  $M > 0$ :

$$\begin{aligned} \sup_{n \in \mathbb{N}} E_0[|\Delta_n|^M] &\lesssim \sup_{n \in \mathbb{N}} \frac{1}{n} \sum_{i=1}^n E_0[\{w_{i-1}(1 + |X_{t_{i-1}}|)\}^M] \\ &\lesssim \sup_{t \in \mathbb{R}_+} E_0[\{w(X_t)(1 + |X_t|)\}^M] < \infty \end{aligned} \quad (34)$$

by means of Burkholder's inequality and Assumption 4.2. Let

$$b_n := a_n^{-2} = nh^{2(1-1/\beta)},$$

and

$$\begin{aligned} \mathbb{Y}_n(\theta) &:= \frac{1}{b_n} \log \mathbb{Z}_n(a_n^{-1}(\theta - \theta_0); \theta_0), \\ \mathbb{Y}(\theta) &:= -\frac{1}{2}(\theta - \theta_0)^\top \Gamma_0(\theta - \theta_0). \end{aligned} \quad (35)$$

Since  $\Gamma_0$  is positive definite, we can find a constant  $\chi > 0$  such that for each  $\theta \in \Theta$

$$\mathbb{Y}(\theta) = \mathbb{Y}(\theta) - \mathbb{Y}(\theta_0) \leq -\chi|\theta - \theta_0|^2. \quad (36)$$

Let  $U_n(r) := \{u \in \mathbb{U}_n(\theta_0) : r \leq |u| \leq b_n^{(1-\alpha)/2}\}$ , where  $\alpha \in (0, 1)$  is a constant. Having (34) and (36) in hand, in order to ensure (32) we are left to proving the following lemmas (see Yoshida [25] for details).

**Lemma 3.1.** *For any  $L > 0$ , we can find (sufficiently small) constants  $\alpha \in (0, 1)$  and  $\rho_1 \in (0, 1)$  in such a way that there exists a constant  $C_L > 0$  such that for every  $r > 0$*

$$\max_{1 \leq j \leq 5} \sup_{n \in \mathbb{N}} P_0 \left[ \sup_{u \in U_n(r)} \frac{|R_n^j(u)|}{1 + |u|^2} \geq r^{-\rho_1} \right] \leq \frac{C_L}{r^L}.$$

**Lemma 3.2.** *For any  $L > 0$ , we can find a  $\beta_2 \in [0, 1/2)$  (sufficiently close to  $1/2$ ) such that*

$$\sup_{n \in \mathbb{N}} E_0 \left[ \left( \sup_{\theta \in \Theta} b_n^{1/2-\beta_2} |\mathbb{Y}_n(\theta) - \mathbb{Y}(\theta)| \right)^L \right] < \infty.$$

For convenience we write

$$\mathcal{I}_n(k) = \frac{1}{n} \sum_{i=1}^n w_{i-1}(1 + |X_{t_{i-1}}|)^k$$

in what follows. Then,  $\sup_{n \in \mathbb{N}} E_0[|\mathcal{I}_n(k)|^M] < \infty$  for every  $k \in \mathbb{N}$  and  $M > 0$ .

### 3.2.1 Proof of Lemma 3.1

Fix any  $L > 0$  and let

$$H_j(r) := \sup_{n \in \mathbb{N}} P_0 \left[ \sup_{u \in U_n(r)} \frac{|R_n^j(u)|}{1 + |u|^2} \geq r^{-\rho_1} \right], \quad r > 0, \quad j \in \{1, 2, \dots, 5\}.$$

We are going to show  $H_j(r) \leq C_L/r^L$  individually. In the sequel,  $K_j \geq 2$  denote arbitrarily large reals, and the constant  $C_L$  may vary from step to step.

*Estimate of  $H_1(r)$ .* We have

$$\begin{aligned} H_1(r) &\leq \sup_{n \in \mathbb{N}} P_0 \left[ \sup_{u \in U_n(r)} \frac{|u|}{1 + |u|^2} \sqrt{n} h^{2-1/\beta} \mathcal{I}_n(2) \geq C r^{-\rho_1} \right] \\ &\leq \sup_{n \in \mathbb{N}} P_0 [\sqrt{n} h^{2-1/\beta} \mathcal{I}_n(2) \geq C r^{1-\rho_1}] \lesssim r^{-K_1(1-\rho_1)}, \end{aligned}$$

hence we get  $H_1(r) \leq C_L/r^L$ .

*Estimate of  $H_2(r)$ .* Let

$$c'_n := b_n^{(1-\alpha)/2}.$$

On  $U_n(r)$ , we have for any  $\kappa_2 > \rho_1$

$$\sup_{u \in U_n(r)} \frac{|R_n^2(u)|}{1 + |u|^2} \leq r^{-\kappa_2} (c'_n)^{\kappa_2} \sup_{u \in U_n(r)} \frac{|R_n^2(u)|}{1 + |u|^2},$$

hence

$$H_2(r) \leq \sup_{n \in \mathbb{N}} P_0 \left[ (c'_n)^{\kappa_2} \sup_{u \in U_n(r)} \frac{|R_n^2(u)|}{1 + |u|^2} \geq C r^{\kappa_2 - \rho_1} \right] \quad (37)$$

On the other hand, using (29),

$$\sup_{u \in U_n(r)} \frac{|R_n^2(u)|}{1 + |u|^2} \lesssim \left( h^{2(2-1/\beta)} + \frac{h^{2-1/\beta} c'_n}{\sqrt{n}} + \frac{(c'_n)^2}{n} \right) \mathcal{I}_n(4). \quad (38)$$

From (37) and (38), we get

$$H_2(r) \lesssim \frac{1}{r^{K_2(\kappa_2 - \rho_1)}} \sup_{n \in \mathbb{N}} \left\{ (c'_n)^{\kappa_2} \left( h^{2(2-1/\beta)} + \frac{h^{2-1/\beta} c'_n}{\sqrt{n}} + \frac{(c'_n)^2}{n} \right) \right\}^{K_2} =: \frac{\sup_{n \in \mathbb{N}} B_n^{K_2}}{r^{K_2(\kappa_2 - \rho_1)}}.$$

Set  $\kappa_2 = \alpha/(1 - \alpha)$  (given any  $\alpha \in (0, 1)$ ), and

$$0 < \rho_1 < \frac{\alpha}{1 - \alpha}. \quad (39)$$

Note that Assumption 3 entails  $\beta > 2/3$ , hence  $2(2 - 1/\beta) > 1 > \alpha$ . Now,

$$\begin{aligned} B_n &= \left\{ (n h^{4-2/\beta}) h^{\frac{4(2-1/\beta)}{\kappa_2(1-\alpha)} - 2} \right\}^{\kappa_2(1-\alpha)/2} + \frac{h^{2-1/\beta}}{\sqrt{n}} b_n^{(1-\alpha)(1+\kappa_2)/2} + \frac{1}{n} b_n^{(1-\alpha)(1+\kappa_2)/2} b_n^{(1-\alpha)/2} \\ &\lesssim o(1) + \frac{h^{2-1/\beta}}{\sqrt{n}} \sqrt{b_n} + \frac{1}{n} b_n^{1-\alpha/2} \lesssim o(1) + h^{3-2/\beta} + \frac{h^{2(1-1/\beta)}}{b_n^{\alpha/2}} \lesssim 1, \end{aligned}$$

yielding that  $H_2(r) \leq C_L/r^L$ . The condition (39) continues to be in force in the sequel.

*Estimate of  $H_3(r)$ .* To derive the uniform convergence of the martingale residual terms, we make use of Lemma 3.7. Picking any constant  $\epsilon \in (0, 1/2]$ , we strengthen (39) to

$$0 < \rho_1 < \frac{\alpha}{1 - \alpha} \wedge \epsilon.$$

Then, in a similar manner to the estimate (37),

$$\begin{aligned} H_3(r) &\leq \sup_{n \in \mathbb{N}} P_0 \left[ \sup_{u \in U_n(r)} \frac{|R_n^3(u)|}{(1 + |u|^2)^{1-\epsilon/2}} \geq C r^{\epsilon - \rho_1} \right] \\ &\lesssim \frac{1}{r^{K_3(\epsilon - \rho_1)}} \sup_{n \in \mathbb{N}} E_0 \left[ \sup_{|v| \leq 1} \left\{ \frac{|R_n^3(c'_n v)|}{(1 + |c'_n v|^2)^{1-\epsilon/2}} \right\}^{K_3} \right]. \end{aligned}$$

Let  $S_n(v) := |R_n^3(c'_n v)|/(1 + |c'_n v|^2)^{1-\epsilon/2}$ . If

$$\sup_{|v| \leq 1} \sup_{n \in \mathbb{N}} E_0[|S_n(v)|^{K_3}] < \infty \quad (40)$$

and if there exists a  $p' > 2$  (here “2” indicates the dimension of  $\theta$ ) such that for every  $v, v'$  satisfying  $0 \leq |v'| \leq |v| \leq 1$

$$\sup_{n \in \mathbb{N}} E_0[|S_n(v) - S_n(v')|^{K_3}] \leq C|v - v'|^{p'}, \quad (41)$$

then  $\sup_{n \in \mathbb{N}} E_0[\sup_{|v| \leq 1} |S_n(v)|^{K_3}] < \infty$  by means of Lemma 3.7, hence  $H_3(r) \leq C_L/r^L$ .

As with the estimate (31) we have  $E_0[|R_n^3(c'_n v)|^{K_3}] \lesssim n^{-1/2} |c'_n v|^{K_3+1}$ , so that, since  $\epsilon \leq 1/2$  and  $K_3 \geq 2$ ,

$$E_0[|S_n(v)|^{K_3}] \lesssim \frac{1}{\sqrt{n}} \frac{|c'_n v|^{K_3+1}}{(1 + |c'_n v|^2)^{K_3(1-\epsilon/2)}} \lesssim \frac{1}{\sqrt{n}} (1 + |c'_n v|^2)^{(1-K_3+\epsilon K_3)/2} \lesssim \frac{1}{\sqrt{n}}.$$

This leads to (40). To prove (41), we first observe that  $|S_n(v) - S_n(v')| \leq \bar{S}_n^1(v, v') + \bar{S}_n^2(v, v')$ , where

$$\begin{aligned} \bar{S}_n^1(v, v') &:= \frac{|R_n^3(c'_n v) - R_n^3(c'_n v')|}{(1 + |c'_n v|^2)^{1-\epsilon/2}}, \\ \bar{S}_n^2(v, v') &:= |R_n^3(c'_n v')| \left| \frac{1}{(1 + |c'_n v|^2)^{1-\epsilon/2}} - \frac{1}{(1 + |c'_n v'|^2)^{1-\epsilon/2}} \right|. \end{aligned}$$

Pick any  $\epsilon' > 0$ . Again as with the estimate (31), it is not difficult to show that

$$\begin{aligned} E_0[|\bar{S}_n^1(v, v')|^{K_3}] &\lesssim \frac{1}{\sqrt{n}} \frac{(c'_n)^{K_3+1} |v| |v - v'|^{K_3}}{(1 + |c'_n v'|^2)^{K_3(1-\epsilon/2)}} \\ &\lesssim |v - v'|^{2+\epsilon'} \frac{(c'_n)^{2+\epsilon'}}{\sqrt{n}} \frac{|c'_n v|^{K_3-1-\epsilon'}}{(1 + |c'_n v'|^2)^{K_3(1-\epsilon/2)}} \\ &\lesssim |v - v'|^{2+\epsilon'} \frac{(c'_n)^{2+\epsilon'}}{\sqrt{n}}. \end{aligned} \quad (42)$$

As for  $\bar{S}_n^2(v, v')$ , since  $1 - \epsilon/2 \in (0, 1)$  and  $|v'| \leq |v| \leq 1$ , we have

$$\begin{aligned} E_0[|\bar{S}_n^2(v, v')|^{K_3}] &\lesssim \frac{|c'_n v'|^{K_3+1}}{\sqrt{n}} \left\{ \frac{c'_n |c'_n v| |v - v'|}{(1 + |c'_n v|^2)(1 + |c'_n v'|^2)} \right\}^{K_3(1-\epsilon/2)} \\ &\lesssim |v - v'|^{K_3(1-\epsilon/2)} \frac{|c'_n v|^{K_3(1-\epsilon/2)}}{\sqrt{n}} \frac{|c'_n v'|^{K_3+1}}{(1 + |c'_n v'|^2)^{K_3(1-\epsilon/2)}} \frac{|c'_n v|^{K_3(1-\epsilon/2)}}{(1 + |c'_n v|^2)^{K_3(1-\epsilon/2)}} \\ &\lesssim |v - v'|^{2+\epsilon'} \frac{(c'_n)^{2+\epsilon'}}{\sqrt{n}} \frac{|c'_n v|^{2K_3(1-\epsilon/2)-(2+\epsilon')}}{(1 + |c'_n v|^2)^{K_3(1-\epsilon/2)}} \lesssim |v - v'|^{2+\epsilon'} \frac{(c'_n)^{2+\epsilon'}}{\sqrt{n}}. \end{aligned} \quad (43)$$

It follows from (42) and (43) that

$$\sup_{n \in \mathbb{N}} E_0[|S_n(v) - S_n(v')|^{K_3}] \leq |v - v'|^{2+\epsilon'} \sup_{n \in \mathbb{N}} \frac{(c'_n)^{2+\epsilon'}}{\sqrt{n}}, \quad (44)$$

for any  $\epsilon' > 0$ . Now, given any  $\alpha \in (0, 1)$  we take  $\epsilon' \in (0, 2\alpha/(1-\alpha))$ , so that  $(1-\alpha)(2+\epsilon') < 2$ . Then  $(c'_n)^{2+\epsilon'}/\sqrt{n} \lesssim b_n/\sqrt{n} = \{nh^{4(1-1/\beta)}\}^{1/2} \lesssim 1$  under Assumption 4, hence (41) follows.

If either  $\lambda_0$  or  $\gamma_0$  is known, so that the dimension of  $\theta$  is one, then, all without distinction, we can deduce (44) with the exponents “ $2 + \epsilon'$ ” in the right-hand side replaced by “ $1 + \epsilon'$ ”.

In this case  $(c'_n)^{1+\epsilon'}/\sqrt{n} \lesssim \sqrt{b_n/n} = h^{1-1/\beta} \lesssim 1$  as soon as  $\beta \geq 1$ , hence we do not need the condition  $nh^{4(1-1/\beta)} \lesssim 1$ .

*Estimate of  $H_4(r)$ .* According to Lemma 3.3,  $|p_h(0) - p_0(0)| \lesssim h^d$  for some constant  $d > 0$ . Strengthen (39) to ( $\epsilon > 0$  is the constant taken in the step of estimating  $H_3(r)$ )

$$0 < \rho_1 < \frac{\alpha \wedge d}{1 - \alpha} \wedge \epsilon. \quad (45)$$

Then we derive for any  $\kappa_4 \in (\rho_1, \frac{\alpha \wedge d}{1 - \alpha} \wedge \epsilon]$

$$\begin{aligned} H_4(r) &\leq \sup_{n \in \mathbb{N}} P_0 \left[ b_n^{\kappa_4(1-\alpha)/2} h^d \mathcal{I}_n(2) \geq C r^{\kappa_4 - \rho_1} \right] \\ &\lesssim \frac{1}{r^{K_4(\kappa_4 - \rho_1)}} \sup_{n \in \mathbb{N}} \left( \sqrt{n} h^{2-1/\beta} \cdot h^{\frac{d}{\kappa_4(1-\alpha)} - 1} \right)^{\kappa_4(1-\alpha)} \lesssim \frac{1}{r^{K_4(\kappa_4 - \rho_1)}}, \end{aligned}$$

hence  $H_4(r) \leq C_L / r^L$ .

*Estimate of  $H_5(r)$ .* We have  $|R_n^5(u)|/(1 + |u|^2) \leq |R_n^{5'}| + |R_n^{5''}|$ , where

$$\begin{aligned} R_n^{5'} &:= \frac{1}{n} \sum_{i=1}^n (w_{i-1} x^{\otimes 2} - E_0[w_{i-1} x^{\otimes 2}]), \\ R_n^{5''} &:= \frac{1}{n} \sum_{i=1}^n (E_0[w_{i-1} x^{\otimes 2}] - E_0^{\pi_0}[w_{i-1} x^{\otimes 2}]), \end{aligned}$$

where  $E_0^{\pi_0}$  denotes the expectation operator corresponding to  $P_0$  with  $X_0$  having the distribution  $\pi_0$ . Since  $X$  is exponentially absolute-regular under  $P_0$ , we can deduce

$$E_0 \left[ \max_{i \leq n} |\sqrt{n} h R_n^{5'}|^{K_5} \right] \lesssim 1$$

by invoking Yoshida [25, Lemma 4]. Strengthen (45) to

$$0 < \rho_1 < \frac{\alpha \wedge d \wedge \epsilon_0}{1 - \alpha} \wedge \epsilon,$$

and then pick a  $\kappa_5 \in (\rho_1, \epsilon_0/(1 - \alpha)]$ , where  $\epsilon_0$  is the constant appearing in Assumption 4.1. Then we get

$$E_0[|(c'_n)^{\kappa_5} R_n^{5'}|^{K_5}] \lesssim \{(c'_n)^{\kappa_5}/\sqrt{n}h\}^{K_5} \lesssim n^{\{\kappa_5(1-\alpha) - \epsilon_0\}K_5/2} \lesssim 1, \quad (46)$$

as  $h^{1-1/\beta} \lesssim 1$ . Finally we estimate  $R_n^{5''}$ . Write  $\|\lambda\|_{\bar{g}} = \sup_{|g| \leq \bar{g}} |\lambda(g)|$  for a signed measure  $\lambda$  and a nonnegative function  $\bar{g}$ , and  $P_t(x, dy)$  for the transition semigroups under  $P_0$  associated with  $X$ . Then, denoting by  $g(x)$  either  $x^2 w(x)$ ,  $-xw(x)$ , or  $w(x)$ , we have

$$\begin{aligned} |R_n^{5''}| &\lesssim \frac{1}{n} \sum_{i=1}^n \int \|P_{t_{i-1}}(x, \cdot) - \pi_0\|_{1+|g|} \eta(dx) \\ &\lesssim \frac{1}{n} \sum_{i=1}^n e^{-a_0 t_{i-1}} \int (1 + |g|)(x) \eta(dx) \lesssim \frac{1}{n} \sum_{i=1}^n e^{-a_0 t_{i-1}} \lesssim \frac{1}{nh}, \end{aligned} \quad (47)$$

where  $a_0 > 0$  is some constant; see Masuda [13, 15] for details. Now, building on (46) and (47), we arrive at the following estimate for every  $r > 0$  large enough:

$$H_5(r) \lesssim \sup_{n \in \mathbb{N}} P_0 \left[ (c'_n)^{\kappa_5} |R_n^{5'}| \geq C r^{\kappa_5 - \rho_1} \right] \lesssim \frac{1}{r^{K_5(\kappa_5 - \rho_1)}}.$$

The proof of Lemma 3.1 is thus complete.

### 3.2.2 Proof of Lemma 3.2

Fix any  $L > 0$ , and write  $\tilde{\mathbb{Y}}_n(\theta) = b_n^{1/2-\beta_2} \{\mathbb{Y}_n(\theta) - \mathbb{Y}(\theta)\}$ . Trivially we have  $\tilde{\mathbb{Y}}_n(\theta_0) \equiv 0$ , hence, by means of Lemma 3.7, it suffices to prove that

$$\exists p' > 2 \forall \theta, \theta' \in \Theta \quad \sup_{n \in \mathbb{N}} E_0 \left[ |\tilde{\mathbb{Y}}_n(\theta) - \tilde{\mathbb{Y}}_n(\theta')|^L \right] \leq C |\theta - \theta'|^{p'}.$$

From (33) and (35) we have

$$\tilde{\mathbb{Y}}_n(\theta) = b_n^{-\beta_2+1/2} \left\{ a_n(\theta - \theta_0)^\top \Delta_n + \sum_{j=1}^5 b_n^{-1} R_n^j(a_n^{-1}(\theta - \theta_0)) \right\}.$$

Hence  $|\tilde{\mathbb{Y}}_n(\theta) - \tilde{\mathbb{Y}}_n(\theta')| \leq |\theta - \theta'| b_n^{-\beta_2} |\Delta_n| + \sum_{j=1}^5 \bar{\mathbb{Y}}_n^j(\theta, \theta')$  for each  $\theta, \theta' \in \Theta$ , where

$$\bar{\mathbb{Y}}_n^j(\theta, \theta') := b_n^{-(\beta_2+1/2)} |R_n^j(a_n^{-1}(\theta - \theta_0)) - R_n^j(a_n^{-1}(\theta' - \theta_0))|$$

In view of (34), it remains to estimate  $\sup_{n \in \mathbb{N}} E_0[|\bar{\mathbb{Y}}_n^j(\theta, \theta')|^L]$  one by one.

*Case of  $j = 1$ .* From (27), we readily get  $E_0[|\bar{\mathbb{Y}}_n^1(\theta, \theta')|^L] \lesssim |\theta - \theta'|^L$ .

*Case of  $j = 2$ .* Note the elementary inequality:  $|\int_x^y g(z) dz| \lesssim |x - y| \{|g(x)| + |g(y)|\}$  for any  $x, y \in \mathbb{R}$  and  $g : \mathbb{R} \rightarrow \mathbb{R}$ . In a similar way to deriving (29), we can deduce that for every  $u, u'$

$$\begin{aligned} |R_n^2(u) - R_n^2(u')| &\lesssim |u - u'| \mathcal{I}_n(4) \\ &\quad \cdot \left\{ (|u| + |u'|) h^{2(2-1/\beta)} + (|u|^2 + |u'|^2) \frac{h^{2-1/\beta}}{\sqrt{n}} + (|u|^3 + |u'|^3) \frac{1}{n} \right\}. \end{aligned}$$

This estimate combined with the boundedness of  $\Theta$  leads to

$$\begin{aligned} \bar{\mathbb{Y}}_n^2(\theta, \theta') &\lesssim |\theta - \theta'| \mathcal{I}_n(4) \left\{ a_n^{2\beta_2-1} \left( h^{2(2-1/\beta)} + \frac{a_n^{-2}}{n} + \frac{a_n^{-1} h^{2-1/\beta}}{\sqrt{n}} \right) \right\} \\ &\lesssim |\theta - \theta'| \mathcal{I}_n(4) \cdot a_n^{2\beta_2-1} h^{2(1-1/\beta)} \\ &= |\theta - \theta'| \mathcal{I}_n(4) \cdot \left\{ \sqrt{n} h^{(1-1/\beta)(3-2\beta_2)/(1-2\beta_2)} \right\}^{1-2\beta_2}. \end{aligned}$$

Now take  $\beta_2 \in [0, 1/2)$  so large that

$$\frac{(1-1/\beta)(3-2\beta_2)}{1-2\beta_2} \geq 2 - \frac{1}{\beta} \quad \left( \iff \beta_2 \geq \frac{2-\beta}{2\beta} \right), \quad (48)$$

then we deduce  $E_0[|\bar{\mathbb{Y}}_n^2(\theta, \theta')|^L] \lesssim |\theta - \theta'|^L$ ; note that the condition  $\beta > 1$  was implicitly used.

*Case of  $j = 3$ .* Every bit as the estimate (31), we can deduce that for every  $u, u'$

$$\begin{aligned} E_0[|R_n^3(u) - R_n^3(u')|^L] &\lesssim |u - u'|^L (|u| + |u'|) \frac{1}{\sqrt{n}} \sup_{t \in \mathbb{R}_+} E_0[w(X_t)^L (1 + |X_t|)^{L+1}] \\ &\lesssim |u - u'|^L (|u| + |u'|) \frac{1}{\sqrt{n}}. \end{aligned}$$

Thus  $E_0[|\bar{\mathbb{Y}}_n^3(\theta, \theta')|^L] \lesssim |\theta - \theta'|^L b_n^{-L(1/2+\beta_2)} a_n^{-1-L} / \sqrt{n} = |\theta - \theta'|^L b_n^{-\beta_2 L} h^{1-1/\beta} \lesssim |\theta - \theta'|^L$ .

*Case of  $j = 4$ .* Reminding the proof of Theorem 2.1 (the step of estimating  $H_4(r)$ ), we have

$$\begin{aligned} |\bar{\mathbb{Y}}_n^4(\theta, \theta')| &\lesssim |\theta - \theta'| \mathcal{I}_n(2) \cdot a_n^{2\beta_2-1} |p_h(0) - p_0(0)| \\ &\lesssim |\theta - \theta'| \mathcal{I}_n(2) \left( \sqrt{n} h^{1-1/\beta+d/(1-2\beta_2)} \right)^{1-2\beta_2}. \end{aligned}$$

The desired estimate  $E_0[|\bar{\mathbb{Y}}_n^4(\theta, \theta')|^L] \lesssim |\theta - \theta'|^L$  follows on letting  $\beta_2$  fulfil that, in addition to (48),

$$1 - \frac{1}{\beta} + \frac{d}{1 - 2\beta_2} \geq 2 - \frac{1}{\beta} \quad \left( \iff \beta_2 \geq \frac{1-d}{2} \right). \quad (49)$$

*Case of  $j = 5$ .* It remains to look at  $\bar{\mathbb{Y}}_n^5(\theta, \theta')$ . In view of the step of estimating  $H_5(r)$  in the proof of Theorem 2.1, we get

$$E_0[|\bar{\mathbb{Y}}_n^5(\theta, \theta')|^L] \lesssim |\theta - \theta'|^L \left\{ E_0 \left[ (b_n^{(1-2\beta_2)/2} R_n^{5'})^L \right] + \left( \frac{b_n^{(1-2\beta_2)/2}}{nh} \right)^L \right\}.$$

Under Assumption 4 we have  $b_n^{(1-2\beta_2)/2}/(nh) \lesssim n^{(1-2\beta_2)/2-\epsilon_0} \lesssim 1$  as soon as

$$\beta_2 \geq \frac{1 - 2\epsilon_0}{2}. \quad (50)$$

Now we get  $E_0[|\bar{\mathbb{Y}}_n^5(\theta, \theta')|^L] \lesssim |\theta - \theta'|^L$  by taking  $\beta_2$  sufficiently close to  $1/2$  so that  $E_0[(b_n^{(1-2\beta_2)/2} R_n^{5'})^L] \lesssim 1$  in addition to (48), (49), and (50).

### 3.3 Proof of Theorem 2.3

First we prove  $\hat{\phi}_\beta(0)_n \rightarrow^p \phi_\beta(0)$ , based on a modification of Ling [11, Section A.2]. Since  $n^{-1} \sum_{i=1}^n w_{i-1} \rightarrow^p \int w(x) \pi_0(dx) > 0$ , it suffices to prove the two convergences:

$$\left| \frac{1}{n} \sum_{i=1}^n w_{i-1} G_{ni}(\hat{u}_n) - \phi_\beta(0) \left( \frac{1}{n} \sum_{i=1}^n w_{i-1} \right) \right| \rightarrow^p 0, \quad (51)$$

$$\left| \frac{1}{n} \sum_{i=1}^n w_{i-1} \left\{ \frac{1}{B_n} K \left( \frac{\Delta_i X - h(\hat{\gamma}_n - \hat{\lambda}_n X_{t_{i-1}})}{h^{1/\beta} B_n} \right) - G_{ni}(\hat{u}_n) \right\} \right| \rightarrow^p 0, \quad (52)$$

where (recall (18) and that  $\hat{u}_n := \sqrt{n} h^{1-1/\beta} (\hat{\theta}_n - \theta_0)$ ), for each  $u$ ,  $n$ , and  $i \leq n$ ,

$$\begin{aligned} G_{ni}(u) &:= E_0^{i-1} \left[ \frac{1}{B_n} K \left( \frac{\epsilon_{ni} + \epsilon'_{n,i-1} - u^\top x_{i-1}/\sqrt{n}}{B_n} \right) \right] \\ &= \int K(z) p_h \left( B_n z - \epsilon'_{n,i-1} + \frac{u^\top x_{i-1}}{\sqrt{n}} \right) dz \end{aligned} \quad (53)$$

the last equality following from the  $\mathcal{F}_{t_{i-1}}$ -measurability of  $\epsilon'_{n,i-1}$  and  $x_{i-1}$ . In view of Lemma 3.3, we have  $\sup_{n,i \leq n} \|G_{ni}\|_\infty < \infty$ .

We begin with (51). Put  $\delta_{ni} = -\epsilon'_{n,i-1} + \hat{u}_n^\top x_{i-1}/\sqrt{n}$ . Using the expression (53) we see that  $n^{-1} \sum_{i=1}^n w_{i-1} |G_{ni}(\hat{u}_n) - \phi_\beta(0)| \leq J_{1n} + J_{2n}$ : first,

$$\begin{aligned} J_{1n} &:= \frac{1}{n} \sum_{i=1}^n w_{i-1} \left| \int K(z) \{p_h(B_n z + \delta_{ni}) - \phi_\beta(B_n z + \delta_{ni})\} dz \right| \\ &\leq h^d \frac{1}{n} \sum_{i=1}^n w_{i-1} = O_p(h^d) \end{aligned}$$

for some  $d > 0$  by virtue of Lemma 3.3; second, again by using Lemma 3.3,

$$\begin{aligned}
J_{2n} &:= \frac{1}{n} \sum_{i=1}^n w_{i-1} \left| \int K(z) \{ \phi_\beta(B_n z + \delta_{ni}) - \phi_\beta(0) \} dz \right| \\
&\lesssim \frac{1}{n} \sum_{i=1}^n w_{i-1} \int K(z) |B_n z + \delta_{ni}| dz \\
&\lesssim O_p(B_n) + \left( h^{2-1/\beta} + \frac{|\hat{u}_n|}{\sqrt{n}} \right) \frac{1}{n} \sum_{i=1}^n w_{i-1} (1 + |X_{t_{i-1}}|) \\
&\lesssim O_p(B_n \vee h^{2-1/\beta} \vee n^{-1/2}) = o_p(1),
\end{aligned}$$

where we used (24) in proceeding from the second line to the third. Thus we get (51).

Next we turn to (52). Set

$$\chi_{ni}(u) = \frac{w_{i-1}}{n} \left\{ \frac{1}{B_n} K \left( \frac{\epsilon_{ni} + \epsilon'_{n,i-1} - u^\top x_{i-1}/\sqrt{n}}{B_n} \right) - G_{ni}(u) \right\},$$

then (52) follows on proving  $|V_n(\hat{u}_n)| \rightarrow^p 0$ , where  $V_n(u) := \sum_{i=1}^n \chi_{ni}(u)$ . For each  $u$ , the sequence  $\{\chi_{ni}(u)\}_{i \leq n}$  forms an  $(\mathcal{F}_{t_i})$ -martingale difference array with the associated quadratic characteristics tending to 0 in probability: through the change of variable as in (53), we see that

$$\sum_{i=1}^n E_0^{i-1} [\chi_{ni}(u)^2] \lesssim O_p \left( \frac{1}{n B_n} \right) + O_p \left( \frac{1}{n} \right) = o_p(1).$$

This readily implies that  $V_n(u) \rightarrow^p 0$  for each  $u$ . Now fix any  $\epsilon > 0$  and  $\epsilon'$ . Then we can find an  $A > 0$ , for which  $\sup_{n \in \mathbb{N}} P_0[|\hat{u}_n| > A] < \epsilon$ . Since

$$P_0[|V_n(\hat{u}_n)| > \epsilon'] \leq P_0 \left[ \sup_{|u| \leq A} |V_n(u)| > \epsilon' \right] + \epsilon,$$

it remains to show that  $\sup_{|u| \leq A} |V_n(u)| \rightarrow^p 0$ . To this end we apply Lemma 3.7. By means of Burkholder's inequality,

$$\begin{aligned}
E_0[|V_n(u) - V_n(u')|^p] &\lesssim \frac{1}{n^{p/2}} \frac{1}{n} \sum_{i=1}^n E_0[|n\chi_{ni}(u) - n\chi_{ni}(u')|^p] \\
&\lesssim \frac{1}{n^{p/2}} \frac{1}{n} \sum_{i=1}^n \{W'_{ni}(u, u') + W''_{ni}(u, u')\}
\end{aligned} \tag{54}$$

for each  $u, u' \in \{u : |u| \leq A\}$ , where, writing  $\delta'_{ni}(u) = \epsilon_{ni} + \epsilon'_{n,i-1} - u^\top x_{i-1}/\sqrt{n}$ ,

$$\begin{aligned}
W'_{ni}(u, u') &:= E_0 \left[ \frac{w_{i-1}^p}{B_n^p} \left| K \left( \frac{\delta'_{ni}(u)}{B_n} \right) - K \left( \frac{\delta'_{ni}(u')}{B_n} \right) \right|^p \right], \\
W''_{ni}(u, u') &:= E_0[w_{i-1}^p |G_{ni}(u) - G_{ni}(u')|^p].
\end{aligned}$$

It follows from the Lipschitz continuity of  $K$  that  $W'_{ni}(u, u') \lesssim n^{-p/2} B_n^{-2p} |u - u'|^p$ . Also, Lemma 3.3 yields that  $W''_{ni}(u, u') \lesssim n^{-p/2} |u - u'|^p$ . Substituting these estimates into (54) yields that

$$E_0[|V_n(u) - V_n(u')|^p] \lesssim \{n^{-p} + (n B_n^2)^{-p}\} |u - u'|^p \lesssim |u - u'|^p.$$

In a similar way, we can deduce

$$\sup_{n \in \mathbb{N}} \sup_{|u| \leq A} E_0[|V_n(u)|^p] \lesssim \sup_{n \in \mathbb{N}} \left\{ n^{-p/2} + \frac{B_n}{(n B_n^2)^{p/2}} \right\} < \infty.$$



Lemma 3.7 now yields that  $\sup_{|u| \leq A} |V_n(u)| \rightarrow^p 0$ . We thus get the desired convergence  $\hat{\phi}_\beta(0)_n \rightarrow^p \phi_\beta(0)$ .

From Theorem 2.1 and Slutsky's lemma,

$$2\hat{\phi}_\beta(0)_n \hat{U}_n^{-1/2} \sqrt{nh}^{1-1/\beta} (\hat{\theta}_n - \theta_0) \rightarrow^d \mathcal{N}_2(0, I_2).$$

The choice  $B_n = l_n h^{-1/\beta}$  enables us to eliminate the factor " $h^{-1/\beta}$ ":

$$\frac{2\sqrt{nh} \sum_{i=1}^n w_{i-1} K\left(\frac{\Delta_i X - h(\hat{\gamma}_n - \hat{\lambda}_n X_{t_{i-1}})}{l_n}\right)}{l_n \sum_{i=1}^n w_{i-1}} \hat{U}_n^{-1/2} (\hat{\theta}_n - \theta_0) \rightarrow^d \mathcal{N}_2(0, I_2) \quad (55)$$

for any positive sequence  $(l_n)$  satisfying the following, which corresponds to (11):

$$l_n = o(h^{1/\beta}), \quad nh^{-1/\beta} l_n \rightarrow \infty, \quad \liminf_{n \rightarrow \infty} nh^{-2/\beta} l_n^2 > 0.$$

These conditions are satisfied by the specific choice  $l_n = \sqrt{nh}^2$  under Assumption 3 and the condition (12), therefore, (13) follows from (55). The proof of Theorem 2.3 is thus complete.

### 3.4 Preliminary lemmas

The following lemma provides a local limit result concerning the distribution of  $\epsilon_{ni}$  for  $h \rightarrow 0$ , with rates of convergence depending on the structure of  $Z$ . Let  $\beta' < 2$  denote the Blumenthal-Gettoor index of the pure-jump part of  $Z$ . Also, recall that  $\beta''$  is the Blumenthal-Gettoor index associated with  $\nu''$ ; see Assumption 1(b).

**Lemma 3.3.** *The following holds true under Assumption 1.*

- (a) *The distribution of  $\epsilon_{ni}$  for each  $n \in \mathbb{N}$  and  $i \leq n$  admits a common positive smooth Lebesgue-density  $p_h$  on  $\mathbb{R}$ .*
- (b) *We have the uniform estimate*

$$\sup_{z \in \mathbb{R}} |p_h(z) - \phi_\beta(z)| \lesssim h^d,$$

where  $\phi_\beta$  is the symmetric  $\beta$ -stable density defined in Section 2.1, and the positive constant  $d$  can be taken as follows.

- (i)  $d = 1$  if either:
  - $\sigma^2 > 0$  and  $\nu(\mathbb{R}) < \infty$ ; or,
  - $\sigma^2 = 0$ ,  $\tilde{g} \equiv 0$  on  $U$ , and  $\nu''(\mathbb{R}) < \infty$ .
- (ii)  $d = 1 - \beta'_+/2$  for any  $\beta'_+ > \beta'$  if  $\sigma^2 > 0$  and  $\nu(\mathbb{R}) = \infty$ .
- (iii)  $d = 1 - \beta''_+/\beta$  for any  $\beta''_+ > \beta''$  if  $\sigma^2 = 0$ ,  $\tilde{g} \equiv 0$  on  $U$ , and  $\nu''(\mathbb{R}) = \infty$ .
- (iv)  $d = \delta'/\beta$  for any  $\delta' \in (0, \beta) \cap (0, \delta]$  if  $\sigma^2 = 0$ ,  $\delta > 0$ , and  $\nu''(\mathbb{R}) < \infty$ .
- (v)  $d = (\delta'/\beta) \vee (1 - \beta''_+/\beta)$  for any  $\delta' \in (0, \beta) \cap (0, \delta]$  and  $\beta''_+ > \beta''$  if  $\sigma^2 = 0$ ,  $\delta > 0$ , and  $\nu''(\mathbb{R}) = \infty$ .

*Proof.* Throughout this proof,  $\varphi_Y$  (resp.  $\mathcal{L}(Y)$ ) denotes the characteristic function (resp. the distribution) of a random variable  $Y$ .

First we prove (a). By the well known property of Lévy integral, we have

$$\begin{aligned}\varphi_{\epsilon_{ni}}(u) &= \exp \left\{ \int_0^h \log \varphi_{Z_1}(h^{-1/\beta} e^{-\lambda_0 s} u) ds \right\} \\ &= \exp \left\{ -\frac{\sigma^2}{2} h^{-2/\beta} \left( \frac{1 - e^{-2\lambda_0 h}}{2\lambda_0} \right) u^2 + \int_0^h (\cos(uz) - 1) \int_0^h \nu(h^{1/\beta} e^{\lambda_0 s} dz) ds \right\},\end{aligned}\quad (56)$$

so that  $\mathcal{L}(\epsilon_{ni}) = \mathcal{L}(h^{-1/\beta} \int_0^h e^{-\lambda_0 s} dZ_s)$  for each  $n \in \mathbb{N}$  and  $i \leq n$ . If  $\sigma^2 > 0$  (hence  $\beta = 2$ ), the existence of a positive smooth density is clear from the independence between the Gaussian and pure-jump parts of  $Z$ . When  $\sigma^2 = 0$ , Assumption 1.2 leads to the following estimate of  $\mathcal{L}(\epsilon_{ni})$ 's Lévy measure: we can find a (small) constant  $r_0 > 0$  such that  $(-r_0, r_0) \subset U$ , that  $\sup_{z \in (-r_0, r_0)} |\tilde{g}(z)| < 1/2$ , and that for every  $r \in (0, r_0]$

$$\begin{aligned}\int_{|z| \leq r} z^2 \int_0^h \nu(h^{1/\beta} e^{\lambda_0 s} dz) ds &\geq \int_{|z| \leq r} z^2 \int_0^h \nu_1(h^{1/\beta} e^{\lambda_0 s} dz) ds \\ &\geq \frac{c}{2} \int_0^h \int_{|z| \leq r} z^2 h^{1/\beta} e^{\lambda_0 s} |h^{1/\beta} e^{\lambda_0 s} z|^{-1-\beta} dz ds \\ &\geq \frac{c}{2-\beta} \left( \frac{1 - e^{-\lambda_0 \beta h}}{\lambda_0 \beta h} \right) r^{2-\beta} \gtrsim r^{2-\beta}.\end{aligned}$$

This estimate combined with Orey's criterion (see Sato [20, Proposition 28.3]) and Sharpe [21] ensures the existence of a positive smooth density of  $\mathcal{L}(\epsilon_{ni})$ ; indeed, the last display implies that  $\sup_{n,i} |\varphi_{\epsilon_{ni}}(u)| \leq \exp(-C|u|^\beta)$  for every  $u$  large enough.

Now we turn to (b). We have to deal with the cases where  $\sigma^2 > 0$  and  $\sigma^2 = 0$  separately. For both cases, we utilize the basic estimate

$$\begin{aligned}&\int |H^{10}(u)H^{11}(u) - H^0(u)| du \\ &\leq \int |H^0(u)| \left| \int_0^1 \left( \frac{H^{10}(u)}{H^0(u)} \right)^y dy \right| \left| \log \frac{H^{10}(u)}{H^0(u)} + \log H^{11}(u) \right| du,\end{aligned}\quad (57)$$

which is valid for any  $(0, \infty)$ -valued characteristic functions  $H^0$ ,  $H^{10}$ , and  $H^{11}$ .

First we consider the case where  $\sigma^2 > 0$ . It follows from the Lévy-Itô decomposition that we may write  $Z_t = \sigma w_t + J_t$ , where  $w$  is a standard Winer process and  $J$  is a pure-jump Lévy process with the Blumenthal-Gettoor index  $\beta' \in [0, 2)$ . For convenience, we write

$$A_h(u) = \int_0^h \log \varphi_{J_1} \left( \frac{e^{-\lambda_0 s} u}{\sqrt{h}} \right) ds.$$

Then, reminding the expression (56), we apply (57) with

$$\begin{aligned}H^0(u) &= \exp \left( -\frac{\sigma^2}{2} u^2 \right), \\ H^{10}(u) &= \exp \left\{ -\frac{\sigma^2}{2} \left( \frac{1 - \exp(-2\lambda_0 h)}{2\lambda_0} \right) u^2 \right\}, \\ H^{11}(u) &= \exp \{ A_h(u) \},\end{aligned}$$

to obtain the following estimates through the Fourier-inversion formula:

$$\begin{aligned}
& \sup_{z \in \mathbb{R}} |p_h(z) - \phi_\beta(z)| \\
& \leq \frac{1}{2\pi} \int |H^{10}(u)H^{11}(u) - H^0(u)| du \\
& \lesssim \int e^{-\sigma^2 u^2/2} \exp \left\{ \frac{\sigma^2 u^2}{2} \left| 1 - \frac{1 - e^{-2\lambda_0 h}}{2\lambda_0 h} \right| \right\} \cdot \left\{ u^2 \left| 1 - \frac{1 - e^{-2\lambda_0 h}}{2\lambda_0 h} \right| + |A_h(u)| \right\} du \\
& \lesssim \int e^{-C u^2} \{h u^2 + |A_h(u)|\} du \lesssim h + \int e^{-C u^2} |A_h(u)| du. \tag{58}
\end{aligned}$$

The estimate of  $|A_h(u)|$  may change according to the structure of  $J$ .

- If  $\nu(\mathbb{R}) < \infty$ , then clearly  $\|\log \varphi_{J_1}(\cdot)\|_\infty < \infty$  since  $J$  is compound Poisson. Therefore we get  $|A_h(u)| \lesssim h$ .
- In case where  $\nu(\mathbb{R}) = \infty$ , we have

$$|A_h(u)| I(|u| \leq \sqrt{h}) \leq h \sup_{|v| \leq 1} |\log \varphi_{J_1}(v)| \lesssim h.$$

Also, for every  $h$  sufficiently small,

$$\inf_{(s,u): 0 \leq s \leq h, |u| \geq \sqrt{h}} \left| \frac{e^{-\lambda_0 s u}}{\sqrt{h}} \right| > \frac{1}{2}.$$

Hence, reminding the fact that  $\lim_{|v| \rightarrow \infty} |v|^{-\beta'_+} |\log \varphi_{J_1}(v)| = 0$  for any  $\beta'_+ > \beta'$  (e.g. Sato [20, page 362]), we obtain

$$|A_h(u)| I(|u| > \sqrt{h}) \lesssim \int_0^h |e^{-\lambda_0 s u} / \sqrt{h}|^{\beta'_+} ds \lesssim |u|^{\beta'_+} h^{1-\beta'_+/2}.$$

Thus it follows that  $|A_h(u)| \lesssim h^{1-\beta'_+/2} (1 + |u|^{\beta'_+})$ .

Piecing together the last two items and (58), we get the claim for  $\sigma^2 > 0$ .

Now we look at the case where  $\sigma^2 = 0$ . First we introduce some notation. For a  $V \subset \mathbb{R}$ , define Lévy processes  $Z'(V)$  and  $Z''$  by  $\log \varphi_{Z'(V)_1}(u) = \int_V \{\cos(uz) - 1\} \nu'(dz)$  and  $\log \varphi_{Z''_1}(u) = \int \{\cos(uz) - 1\} \nu''(dz)$ . Then  $Z = Z^\sharp + Z^\flat$ , where  $Z^\sharp := Z'(U)$  and  $Z^\flat := Z'(U^c) + Z''$ . Obviously,  $Z^\sharp$  (resp.  $Z^\flat$ ) has the Blumenthal-Gettoor index  $\beta$  (resp.  $\beta''$ ). Let

$$u \mapsto \exp \left\{ \int (\cos(zu) - 1) c |z|^{-1-\beta} dz \right\} =: \exp(-c_1 |u|^\beta).$$

denote the symmetric  $\beta$ -stable characteristic function corresponding to the density  $\phi_\beta$ ; specifically,  $c$  and  $c_1$  are related as  $c_1 = 2c\beta^{-1}\Gamma(1-\beta)\cos(\beta\pi/2)$ . Then, we are going to apply (57) with

$$\begin{aligned}
H^0(u) &= \exp(-c_1 |u|^\beta), \\
H^{10}(u) &= \exp \left\{ \int_0^h \log \varphi_{Z^\sharp_1}(e^{-\lambda_0 s} u h^{-1/\beta}) ds \right\}, \\
H^{11}(u) &= \exp \left\{ \int_0^h \log \varphi_{Z^\flat_1}(e^{-\lambda_0 s} u h^{-1/\beta}) ds \right\}.
\end{aligned}$$

We may set  $U = (-1, 1)$  without loss of generality. Then, we have for each  $u \in \mathbb{R}$

$$\begin{aligned}
& \log \frac{H^{10}(u)}{H^0(u)} \\
&= \int \{\cos(uz) - 1\} \left( -\frac{c}{|z|^{1+\beta}} + \int_0^h I(|z| \leq e^{-\lambda_0 s} h^{-1/\beta}) e^{\lambda_0 s} h^{1/\beta} g(e^{\lambda_0 s} h^{1/\beta} z) ds \right) dz \\
&= \int \{\cos(uz) - 1\} \frac{c}{|z|^{1+\beta}} \left( \frac{1}{h} \int_0^h e^{-\lambda_0 \beta s} I(|z| \leq e^{-\lambda_0 s} h^{-1/\beta}) ds - 1 \right) dz \\
&\quad + \int \{\cos(uz) - 1\} \frac{c}{|z|^{1+\beta}} \frac{1}{h} \int_0^h e^{-\lambda_0 \beta s} I(|z| \leq e^{-\lambda_0 s} h^{-1/\beta}) \tilde{g}(e^{\lambda_0 s} h^{1/\beta} z) ds dz \\
&=: A_h^0(u) + A_h^1(u), \quad \text{say.} \tag{59}
\end{aligned}$$

For  $A_h^0$ , we have

$$\begin{aligned}
|A_h^0(u)| &\leq \left| \int \{\cos(uz) - 1\} \frac{c}{|z|^{1+\beta}} \left( \frac{1}{h} \int_0^h e^{-\lambda_0 \beta s} ds - 1 \right) dz \right| \\
&\quad + \left| \int \{\cos(uz) - 1\} \frac{c}{|z|^{1+\beta}} \frac{1}{h} \int_0^h e^{-\lambda_0 \beta s} I(|z| > e^{-\lambda_0 s} h^{-1/\beta}) ds dz \right| \\
&\lesssim h \left| \int \{\cos(uz) - 1\} \frac{c}{|z|^{1+\beta}} dz \right| \\
&\quad + \left| \frac{1}{h} \int_0^h e^{-\lambda_0 \beta s} \left\{ \int_{|z| > e^{-\lambda_0 s} h^{-1/\beta}} (\cos(uz) - 1) |z|^{-(1+\beta)} dz \right\} ds \right| \\
&\lesssim h|u|^\beta + \frac{1}{h} \int_0^h e^{-\lambda_0 \beta s} \left( \int_{e^{-\lambda_0 s} h^{-1/\beta}}^\infty z^{-1-\beta} dz \right) ds \lesssim h(|u|^\beta + 1). \tag{60}
\end{aligned}$$

We now turn to  $A_h^1$ . Trivially,  $A_h^1 \equiv 0$  if  $\tilde{g} \equiv 0$  on  $U$ ; otherwise,  $|\tilde{g}(z)| \lesssim |z|^{\delta'}$  on  $U$  for every  $\delta' \in (0, \delta]$ . Fix any  $\delta' \in (0, \beta) \cap (0, \delta]$  in the sequel. We readily get

$$|A_h^1(u)| \lesssim -h^{\delta'/\beta} \int \{\cos(uz) - 1\} |z|^{-1-(\beta-\delta')} dz \lesssim h^{\delta'/\beta} |u|^{\beta-\delta'}. \tag{61}$$

Summarizing (59), (60), and (61) yields

$$\begin{aligned}
\sup_{z \in \mathbb{R}} |p_h(z) - \phi_\beta(z)| &\lesssim \int \exp \left[ -c_1 |u|^\beta + C \{ h^{\delta'/\beta} |u|^{\beta-\delta'} + h(|u|^\beta + 1) \} \right] \\
&\quad \cdot \left\{ h^{\delta'/\beta} |u|^{\beta-\delta'} + h(|u|^\beta + 1) + |\log H^{11}(u)| \right\} du, \tag{62}
\end{aligned}$$

where the terms involving  $\delta'$  vanish if  $\tilde{g} \equiv 0$  on  $U$ . Just like the foregoing estimates of  $|A_h(u)|$ , we see that:

- $|\log H^{11}(u)| \lesssim h$  if  $Z^\flat$  is compound Poisson;
- otherwise, for any  $\beta''_+ > \beta''$  we have

$$\begin{aligned}
|\log H^{11}(u)| &\leq |\log H^{11}(u)| I(|u| \leq h^{1/\beta}) + |\log H^{11}(u)| I(|u| > h^{1/\beta}) \\
&\lesssim h \sup_{|v| \leq 1} |\log H^{11}(v)| + \int_0^h |e^{-\lambda_0 s} h^{-1/\beta} u|^{\beta''_+} ds \\
&\lesssim h^{1-\beta''_+/\beta} (|u|^{\beta''_+} + 1).
\end{aligned}$$

Combining these estimates with (62), it is straightforward to complete the proof of (b).  $\square$

**Lemma 3.4.** Suppose  $\int_{|z|>1} |z|^q \nu(dz) < \infty$  for some  $q > 0$ . Then,

$$\sup_{t,s \in \mathbb{R}_+ : |t-s| \leq h} E_0[|X_t - X_s|^{2 \wedge q}] \lesssim h^{1 \wedge (q/2)}.$$

*Proof.* Using the expression (17) we get

$$\begin{aligned} & \sup_{t,s \in \mathbb{R}_+ : |t-s| \leq h} E_0[|X_t - X_s|^{2 \wedge q}] \\ & \lesssim h^{2 \wedge q} \sup_{t \in \mathbb{R}_+} E_0[|X_t|^{2 \wedge q}] + \sup_{t,s \in \mathbb{R}_+ : |t-s| \leq h} E \left[ \left| \int_0^{t-s} e^{-\lambda_0 v} dZ_v \right|^{2 \wedge q} \right] =: U_h^1 + U_h^2(Z). \end{aligned}$$

Trivially  $U_h^1 \lesssim h^{2 \wedge q}$ , so that it suffices to deduce  $U_h^2(Z) \lesssim h^{1 \wedge (q/2)}$ .

For  $q \geq 2$ , we have  $U_h^2(Z) = E[Z_1^2] \int_0^h e^{-2\lambda_0 s} ds \lesssim h = h^{1 \wedge (q/2)}$ . For  $q < 2$ , we consider the decomposition  $Z = Z^0 + Z^1$ , where  $Z^0$  (resp.  $Z^1$ ) has only jumps whose sizes in magnitude are less than or equal to 1 (resp. greater than 1). Then  $U_h^2(Z^1) \lesssim h$  since  $Z^1$  is compound Poisson and  $\int_{|z|>1} |z|^q \nu(dz) < \infty$ . As for  $U_h^2(Z^0)$ , since  $Z_1^0$  admits finite variance, Schwarz's inequality yields that

$$U_h^2(Z^0) \leq \sup_{t,s \in \mathbb{R}_+ : |t-s| \leq h} \left\{ E_0 \left[ \left| \int_0^{t-s} e^{-\lambda_0 v} dZ_v^0 \right|^2 \right] \right\}^{q/2} \lesssim h^{q/2}.$$

Hence we get  $U_h^2(Z) \lesssim \{U_h^2(Z^0) + U_h^2(Z^1)\} \lesssim h^{1 \wedge (q/2)}$  as desired.  $\square$

**Lemma 3.5.** Suppose Assumption 2 and that  $\int_{|z|>1} |z|^q \nu(dz) < \infty$  for some  $q > 0$ . Let

$$g_{k,l}(x) := x^k w(x)^l$$

for  $k \in \{0, 1, 2\}$  and  $l \in \{1, 2\}$ , and let

$$\delta_n(g_{k,l}) := \left| \frac{1}{nh} \int_0^{nh} g_{k,l}(X_s) ds - \frac{1}{n} \sum_{i=1}^n g_{k,l}(X_{t_{i-1}}) \right|.$$

Then  $\delta_n(g_{k,l}) \xrightarrow{p} 0$  for  $k \in \{0, 1, 2\}$  and  $l \in \{1, 2\}$ .

*Proof.* The claim follows on proving

$$\bar{\delta}_n(g_{k,l}) := \sup_{t,s : |t-s| \leq h} E_0[|g_{k,l}(X_t) - g_{k,l}(X_s)|] \rightarrow 0. \quad (63)$$

We divide the proof into two steps.

*Step 1.* First we prove (63) for  $l = 1$ . It suffices to show that  $\bar{\delta}_h'(k) \vee \bar{\delta}_h''(k) \rightarrow 0$ , where

$$\begin{aligned} \bar{\delta}_h'(k) &:= \sup_{t,s : |t-s| \leq h} E_0[|X_t|^k |w(X_t) - w(X_s)|], \\ \bar{\delta}_h''(k) &:= \sup_{t,s : |t-s| \leq h} E_0[w(X_s) |X_t^k - X_s^k|]. \end{aligned}$$

Note that the growth rate of  $g_{k,1}$  is connected with how large  $q$  can be taken:  $g_{1,1}$  and  $g_{2,1}$  can be unbounded if, e.g.,  $q$  is set to be large and  $w$  is a positive constant.

First we target  $k = 0$ . Fix an arbitrary  $\epsilon > 0$ , then we can find a  $\rho > 0$  such that  $|x - y| < \rho$  implies  $|w(x) - w(y)| < \epsilon/2$ . Observe that from Lemma 3.4 we have

$$\bar{\delta}_h'(0) \leq \frac{\epsilon}{2} + 2\|w\|_\infty \sup_{t,s : |t-s| \leq h} E_0 \left[ \left( \frac{|X_t - X_s|}{\rho} \right)^{2 \wedge q} \right] \leq \frac{\epsilon}{2} + Ch^{1 \wedge (q/2)},$$

which implies that  $\bar{\delta}_n(g_{k,1}) < \epsilon$  for every  $n$  large enough, hence  $\bar{\delta}_n(g_{k,1}) \rightarrow 0$ .

Next, we target  $k \in \{1, 2\}$  and  $q > 2$ . It follows from Hölder's and Schwarz's inequalities and Lemma 3.4 that for  $k \in \{1, 2\}$

$$\bar{\delta}_h''(1) \vee \bar{\delta}_h''(2) \lesssim \left( \sup_{t,s:|t-s|\leq h} E_0[|X_t - X_s|^2] \right)^{1/2} \lesssim \sqrt{h}. \quad (64)$$

Using the uniform continuity of  $w$  as before, we derive

$$\begin{aligned} \bar{\delta}_h'(k) &\leq \epsilon \sup_{t \in \mathbb{R}_+} E_0[|X_t|^k] + \frac{2\|w\|_\infty}{\rho^r} \sup_{t,s:|t-s|\leq h} E_0[|X_t|^k |X_t - X_s|^r] \\ &\lesssim \epsilon + \sup_{t,s:|t-s|\leq h} \{E_0[|X_t|^{ku}]\}^{1/u} \{E_0[|X_t - X_s|^{rv}]\}^{1/v} \end{aligned}$$

for any  $r > 0$  and  $u > 1$  together with  $u$ 's conjugate exponent  $v$ . Now take  $u = q/k$ ,  $v = q/(q-k)$ , and  $r = 2(q-k)/q$  to conclude that for  $k \in \{1, 2\}$

$$\bar{\delta}_h'(k) \lesssim \epsilon + \left( \sup_{t,s:|t-s|\leq h} E_0[|X_t - X_s|^2] \right)^{r/2} \lesssim \epsilon + h^{r/2}. \quad (65)$$

Now (64) and (65) yields that  $\bar{\delta}_n(g_{k,1}) \rightarrow 0$  for  $k \in \{1, 2\}$ .

We are left to considering the case where  $k \in \{1, 2\}$  and  $q \in (0, 2]$ . However, in this case we can see that Assumption 2.2 implies the boundedness and uniform continuity of  $g_{1,1}$  and  $g_{2,1}$ , so that we can follow the same line as in the case where  $k = 0$ . This completes the proof of (63) for  $l = 1$ .

*Step 2.* In case of  $l = 2$ , we observe that for every  $x, y \in \mathbb{R}$

$$|g_{k,2}(x) - g_{k,2}(y)| \leq \|w\|_\infty \{|g_{k,1}(x) - g_{k,1}(y)| + |y|^k |w(y) - w(x)|\},$$

from which it follows that  $\bar{\delta}_h(g_{k,2}) \lesssim \bar{\delta}_h(g_{k,1}) + \bar{\delta}_h'(k) \rightarrow 0$  in view of what we have seen in the previous step. This leads to (63) for  $l = 2$ . The proof of Lemma 3.5 is thus complete.  $\square$

**Remark 3.6.** *It is a simple matter to show  $\sup_{t,s:|t-s|\leq h} E_0[|k(X_t) - k(X_s)|^r] \lesssim \sqrt{h}$  for each  $r \geq 2$  if  $q$  can be large enough and  $k$  is a  $C^1$ -function with the derivative being at most polynomial growth. This trivially leads to the ergodic theorem for the discrete-time samples:  $n^{-1} \sum_{i=1}^n k_*(X_{t_{i-1}}) \rightarrow^P \pi_0(k_*)$  for smooth  $k_*$ . (Of course, this can be valid for much more general class of diffusions with jumps.) On the other hand, Lemma 3.5 enables us to deal with small  $q$  (i.e., heavy-tailed cases) too, without imposing the global differentiability of  $w$ . A practical candidate for  $w$  when suspecting heavy-tailed nature in data would be the standard Gaussian density, however, it is also possible to take, e.g., the standard Laplace density  $w(x) = \exp(-\sqrt{2}|x|)/\sqrt{2}$  (although  $w$  need not to be a probability density).*

The following lemma are used to deduce the uniform estimates of some martingale terms. See Kunita [10, Theorem 1.4.7] for details.

**Lemma 3.7.** *Let  $\{\Xi_n(h) : h = (h^j)_{j=1}^p \in H^-\}$  be a sequence of  $\mathbb{R}$ -valued continuous random fields defined on a probability space  $(\Omega, \mathcal{F}, P)$ , where  $H$  is a bounded domain in  $\mathbb{R}^p$ . Suppose there exist constants  $a > p$ ,  $b > 0$  and  $c > 0$  such that for every  $h_1, h_2 \in H$*

$$\begin{aligned} \sup_{n \in \mathbb{N}} P[|\Xi_n(h_1) - \Xi_n(h_2)|^b] &\leq c|h_1 - h_2|^a, \\ \sup_{n \in \mathbb{N}} P[|\Xi_n(h_1)|^b] &\leq c. \end{aligned}$$

*Then the family  $\{\Xi_n(\cdot)\}_n$  is tight with respect to the supremum norm over  $H$ , and moreover for any compact convex set  $K \subset H^-$  we have*

$$\sup_{n \in \mathbb{N}} P \left[ \sup_{h \in K} |\Xi_n(h)|^b \right] < \infty.$$

## Acknowledgment

This work was partly supported by Grant-in-Aid for Young Scientists (B), Japan, and Cooperative Research Program of the Institute of Statistical Mathematics. The author is grateful to Prof. Nakahiro Yoshida for giving me a better understanding of a part of his results [25].

## References

- [1] Andrews, B., Calder, M. and Davis, R. A. (2009), Maximum likelihood estimation for  $\alpha$ -stable autoregressive processes. *Ann. Statist.* **37**, 1946–1982.
- [2] Davis, R. A. and Dunsmuir, W. T. M. (1997), Least absolute deviation estimation for regression with ARMA errors. *J. Theoret. Probab.* **10**, 481–497.
- [3] Davis, R. A., Knight, K. and Liu, J. (1992),  $M$ -estimation for autoregressions with infinite variance. *Stochastic Process. Appl.* **40**, 145–180.
- [4] Doob, J. L. (1942), The Brownian movement and stochastic equations. *Ann. of Math. (2)* **43**, 351–369.
- [5] Garbaczewski, P. and Olkiewicz, R. (2000), Ornstein-Uhlenbeck-Cauchy process. *J. Math. Phys.* **41**, 6843–6860.
- [6] Hjrt, N. L. and Pollard, D. (1993), Asymptotics for minimisers of convex processes. Statistical Research Report, University of Oslo.
- [7] Hu, Y. and Long, H. (2009), Least squares estimator for Ornstein-Uhlenbeck processes driven by  $\alpha$ -stable motions. *Stochastic Processes Appl.* **119**, 2465–2480.
- [8] Knight, K. (1998), Limiting distributions for  $L_1$  regression estimators under general conditions. *Ann. Statist.* **26**, 755–770.
- [9] Koenker, R. (2005), *Quantile Regression*. Cambridge University Press, Cambridge.
- [10] Kunita, H. (1990), *Stochastic Flows and Stochastic Differential Equations*. Cambridge University Press, Cambridge.
- [11] Ling, S. (2005), Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67**, 381–393.
- [12] Luschgy, H. (1992), Local asymptotic mixed normality for semimartingale experiments. *Probab. Theory Related Fields* **92**, 151–176.
- [13] Masuda, H. (2004), On multi-dimensional Ornstein-Uhlenbeck processes driven by a general Lvy process. *Bernoulli* **10**, 1–24.
- [14] Masuda, H. (2005), Simple estimators for parametric Markovian trend of ergodic processes based on sampled data. *J. Japan Statist. Soc.* **35**, 147–170.
- [15] Masuda, H. (2007), Ergodicity and exponential  $\beta$ -mixing bound for multidimensional diffusions with jumps. *Stochastic Processes Appl.* **117**, 35–56. [Erratum: (2009) *Stochastic Processes Appl.* **119**, 676–678.]
- [16] Portnoy, S. and Koenker, R. (1997), The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators. *Statist. Sci.* **12**, 279–300.
- [17] Raible, S. (1999), *Lvy Processes in Finance: Theory, Numerics, and Empirical Facts*. PhD thesis, University of Freiburg, 2000.

- [18] Rosiński, J. (2007), Tempering stable processes. *Stochastic Process. Appl.* **117**, 677–707.
- [19] Sakamoto, Y. and Yoshida, N. (2004), Asymptotic expansion formulas for functionals of  $\epsilon$ -Markov processes with a mixing property. *Ann. Inst. Statist. Math.* **56**, 545–597.
- [20] Sato, K. (1999), *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge.
- [21] Sharpe, M. (1969), Zeroes of infinitely divisible densities. *Ann. Math. Statist.* **40**, 1503–1505.
- [22] Sørensen, M. (1991), Likelihood methods for diffusions with jumps. *Statistical Inference in Stochastic Processes*, 67–105, Probab. Pure Appl., 6, Dekker, New York.
- [23] Uchida, M. and Yoshida, N. (2001), Information criteria in model selection for mixing processes. *Stat. Inference Stoch. Process.* **4**, 73–98.
- [24] Uchida, M. and Yoshida, N. (2006), Asymptotic expansion and information criteria. *SUT J. Math.* **42**, 31–58.
- [25] Yoshida, N. (2005), Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. To appear in *Ann. Inst. Statist. Math.*



# List of MI Preprint Series, Kyushu University

The Global COE Program  
Math-for-Industry Education & Research Hub

MI

- MI2008-1 Takahiro ITO, Shuichi INOKUCHI & Yoshihiro MIZOGUCHI  
Abstract collision systems simulated by cellular automata
- MI2008-2 Eiji ONODERA  
The initial value problem for a third-order dispersive flow into compact almost Hermitian manifolds
- MI2008-3 Hiroaki KIDO  
On isosceles sets in the 4-dimensional Euclidean space
- MI2008-4 Hirofumi NOTSU  
Numerical computations of cavity flow problems by a pressure stabilized characteristic-curve finite element scheme
- MI2008-5 Yoshiyasu OZEKI  
Torsion points of abelian varieties with values in infinite extensions over a p-adic field
- MI2008-6 Yoshiyuki TOMIYAMA  
Lifting Galois representations over arbitrary number fields
- MI2008-7 Takehiro HIROTSU & Setsuo TANIGUCHI  
The random walk model revisited
- MI2008-8 Silvia GANDY, Masaaki KANNO, Hirokazu ANAI & Kazuhiro YOKOYAMA  
Optimizing a particular real root of a polynomial by a special cylindrical algebraic decomposition
- MI2008-9 Kazufumi KIMOTO, Sho MATSUMOTO & Masato WAKAYAMA  
Alpha-determinant cyclic modules and Jacobi polynomials

- MI2008-10 Sangyeol LEE & Hiroki MASUDA  
Jarque-Bera Normality Test for the Driving Lévy Process of a Discretely Observed Univariate SDE
- MI2008-11 Hiroyuki CHIHARA & Eiji ONODERA  
A third order dispersive flow for closed curves into almost Hermitian manifolds
- MI2008-12 Takehiko KINOSHITA, Kouji HASHIMOTO and Mitsuhiro T. NAKAO  
On the  $L^2$  a priori error estimates to the finite element solution of elliptic problems with singular adjoint operator
- MI2008-13 Jacques FARAUT and Masato WAKAYAMA  
Hermitian symmetric spaces of tube type and multivariate Meixner-Pollaczek polynomials
- MI2008-14 Takashi NAKAMURA  
Riemann zeta-values, Euler polynomials and the best constant of Sobolev inequality
- MI2008-15 Takashi NAKAMURA  
Some topics related to Hurwitz-Lerch zeta functions
- MI2009-1 Yasuhide FUKUMOTO  
Global time evolution of viscous vortex rings
- MI2009-2 Hidetoshi MATSUI & Sadanori KONISHI  
Regularized functional regression modeling for functional response and predictors
- MI2009-3 Hidetoshi MATSUI & Sadanori KONISHI  
Variable selection for functional regression model via the  $L_1$  regularization
- MI2009-4 Shuichi KAWANO & Sadanori KONISHI  
Nonlinear logistic discrimination via regularized Gaussian basis expansions
- MI2009-5 Toshiro HIRANOUCI & Yuichiro TAGUCHI  
Flat modules and Groebner bases over truncated discrete valuation rings

- MI2009-6 Kenji KAJIWARA & Yasuhiro OHTA  
Bilinearization and Casorati determinant solutions to non-autonomous 1+1 dimensional discrete soliton equations
- MI2009-7 Yoshiyuki KAGEI  
Asymptotic behavior of solutions of the compressible Navier-Stokes equation around the plane Couette flow
- MI2009-8 Shohei TATEISHI, Hidetoshi MATSUI & Sadanori KONISHI  
Nonlinear regression modeling via the lasso-type regularization
- MI2009-9 Takeshi TAKAISHI & Masato KIMURA  
Phase field model for mode III crack growth in two dimensional elasticity
- MI2009-10 Shingo SAITO  
Generalisation of Mack's formula for claims reserving with arbitrary exponents for the variance assumption
- MI2009-11 Kenji KAJIWARA, Masanobu KANEKO, Atsushi NOBE & Teruhisa TSUDA  
Ultradiscretization of a solvable two-dimensional chaotic map associated with the Hesse cubic curve
- MI2009-12 Tetsu MASUDA  
Hypergeometric  $q$ -functions of the  $q$ -Painlevé system of type  $E_8^{(1)}$
- MI2009-13 Hidenao IWANE, Hitoshi YANAMI, Hirokazu ANAI & Kazuhiro YOKOYAMA  
A Practical Implementation of a Symbolic-Numeric Cylindrical Algebraic Decomposition for Quantifier Elimination
- MI2009-14 Yasunori MAEKAWA  
On Gaussian decay estimates of solutions to some linear elliptic equations and its applications
- MI2009-15 Yuya ISHIHARA & Yoshiyuki KAGEI  
Large time behavior of the semigroup on  $L^p$  spaces associated with the linearized compressible Navier-Stokes equation in a cylindrical domain

- MI2009-16 Chikashi ARITA, Atsuo KUNIBA, Kazumitsu SAKAI & Tsuyoshi SAWABE  
Spectrum in multi-species asymmetric simple exclusion process on a ring
- MI2009-17 Masato WAKAYAMA & Keitaro YAMAMOTO  
Non-linear algebraic differential equations satisfied by certain family of elliptic functions
- MI2009-18 Me Me NAING & Yasuhide FUKUMOTO  
Local Instability of an Elliptical Flow Subjected to a Coriolis Force
- MI2009-19 Mitsunori KAYANO & Sadanori KONISHI  
Sparse functional principal component analysis via regularized basis expansions and its application
- MI2009-20 Shuichi KAWANO & Sadanori KONISHI  
Semi-supervised logistic discrimination via regularized Gaussian basis expansions
- MI2009-21 Hiroshi YOSHIDA, Yoshihiro MIWA & Masanobu KANEKO  
Elliptic curves and Fibonacci numbers arising from Lindenmayer system with symbolic computations
- MI2009-22 Eiji ONODERA  
A remark on the global existence of a third order dispersive flow into locally Hermitian symmetric spaces
- MI2009-23 Stjepan LUGOMER & Yasuhide FUKUMOTO  
Generation of ribbons, helicoids and complex scherk surface in laser-matter Interactions
- MI2009-24 Yu KAWAKAMI  
Recent progress in value distribution of the hyperbolic Gauss map
- MI2009-25 Takehiko KINOSHITA & Mitsuhiro T. NAKAO  
On very accurate enclosure of the optimal constant in the a priori error estimates for  $H_0^2$ -projection

- MI2009-26 Manabu YOSHIDA  
Ramification of local fields and Fontaine's property (Pm)
- MI2009-27 Yu KAWAKAMI  
Value distribution of the hyperbolic Gauss maps for flat fronts in hyperbolic three-space
- MI2009-28 Masahisa TABATA  
Numerical simulation of fluid movement in an hourglass by an energy-stable finite element scheme
- MI2009-29 Yoshiyuki KAGEI & Yasunori MAEKAWA  
Asymptotic behaviors of solutions to evolution equations in the presence of translation and scaling invariance
- MI2009-30 Yoshiyuki KAGEI & Yasunori MAEKAWA  
On asymptotic behaviors of solutions to parabolic systems modelling chemotaxis
- MI2009-31 Masato WAKAYAMA & Yoshinori YAMASAKI  
Hecke's zeros and higher depth determinants
- MI2009-32 Olivier PIRONNEAU & Masahisa TABATA  
Stability and convergence of a Galerkin-characteristics finite element scheme of lumped mass type
- MI2009-33 Chikashi ARITA  
Queueing process with excluded-volume effect
- MI2009-34 Kenji KAJIWARA, Nobutaka NAKAZONO & Teruhisa TSUDA  
Projective reduction of the discrete Painlevé system of type  $(A_2 + A_1)^{(1)}$
- MI2009-35 Yosuke MIZUYAMA, Takamasa SHINDE, Masahisa TABATA & Daisuke TAGAMI  
Finite element computation for scattering problems of micro-hologram using DtN map

MI2009-36 Reiichiro KAWAI & Hiroki MASUDA

Exact simulation of finite variation tempered stable Ornstein-Uhlenbeck processes

MI2009-37 Hiroki MASUDA

On statistical aspects in calibrating a geometric skewed stable asset price model

MI2010-1 Hiroki MASUDA

Approximate self-weighted LAD estimation of discretely observed ergodic Ornstein-Uhlenbeck processes