



# Optimal stable Ornstein–Uhlenbeck regression

Hiroki Masuda<sup>1</sup>

Received: 1 August 2022 / Revised: 15 January 2023 / Accepted: 15 February 2023 /  
Published online: 13 March 2023  
© The Author(s) 2023

## Abstract

We prove asymptotically efficient inference results concerning an Ornstein–Uhlenbeck regression model driven by a non-Gaussian stable Lévy process, where the output process is observed at high frequency over a fixed period. The local asymptotics of non-ergodic type for the likelihood function is presented, followed by a way to construct an asymptotically efficient estimator through a suboptimal, yet very simple preliminary estimator.

**Keywords** Asymptotic mixed normality · Optimal regression · Stable Ornstein-Uhlenbeck process

## 1 Introduction

### 1.1 Objective and background

Given an underlying filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, P)$ , we consider the following Ornstein–Uhlenbeck (OU) regression model

$$Y_t = Y_0 + \int_0^t (\mu \cdot X_s - \lambda Y_s) ds + \sigma J_t, \quad t \in [0, T], \quad (1.1)$$

where  $J$  is the symmetric  $\beta$ -stable (càdlàg) Lévy process characterized by

$$E[e^{iuJ_t}] = \exp(-t|u|^\beta), \quad t \geq 0, u \in \mathbb{R},$$

and is independent of the initial variable  $Y_0$ , and where  $X = (X_t)_{t \in [0, T]}$  is an  $\mathbb{R}^q$ -valued non-random càdlàg function such that

---

✉ Hiroki Masuda  
hmasuda@ms.u-tokyo.ac.jp

<sup>1</sup> Graduate School of Mathematical Sciences, The University of Tokyo, 3-8-1 Komaba Meguro-ku, Tokyo 153-8914, Japan

$$\lambda_{\min} \left( \int_0^T X_t^{\otimes 2} dt \right) > 0, \quad (1.2)$$

with  $\lambda_{\min}(A)$  denoting the minimum eigenvalue of a square matrix  $A$ . Throughout, the terminal sampling time  $T > 0$  is a fixed constant. Let

$$\theta := (\lambda, \mu, \beta, \sigma) \in \Theta,$$

where  $\Theta \subset \mathbb{R}^p$  ( $p := q + 3$ ) is a bounded convex domain such that its closure  $\overline{\Theta} \subset \mathbb{R} \times \mathbb{R}^q \times (0, 2) \times (0, \infty)$ . The primary objective of this paper is the asymptotically efficient estimation of  $\theta$  when available data is  $(X_t)_{t \in [0, T]}$  and  $(Y_{t_j})_{j=1}^n$ , where  $t_j = t_j^n := jh$  with  $h = h_n := T/n$ ; later on, we will consider cases where we observe  $(X_{t_j})_{j=1}^n$  instead of the full continuous-time record. We will denote the true value of  $\theta$  by  $\theta_0 = (\lambda_0, \mu_0, \beta_0, \sigma_0) \in \Theta$ .

Analysis of the (time-homogeneous) OU process driven by a stable Lévy process goes back to Doob (1942), where Doob treated the model in a genuinely analytic manner without Itô's formula, which has not been published as yet at that time. Nowadays, the OU models have been used in a wide variety of applications, such as electric consumption modeling (Perninge et al. 2011; Borovkova and Schmeck 2017; Verdejo et al. 2019), ecology (Jhwueng and Maroulas 2014), and protein dynamics modeling (Challis and Schmidler 2012), to mention a few.

The model (1.1) may be seen as a continuous-time counterpart of the simple first-order ARX (autoregressive exogenous) model. Nevertheless, any proper form of the efficient estimation result has been missing in the literature, probably due to the lack of background theory to estimate all the parameters involved under the bounded-domain infill asymptotics. Let us note that, when  $J$  is a Wiener process ( $\beta = 2$ ), the drift parameters are consistently estimable only when the terminal sampling time tends to infinity, and the associated statistical experiments are known to possess essentially different properties according to the sign of  $\lambda$ . That is to say, the model is: locally asymptotically normal for  $\lambda > 0$  (ergodic case); locally asymptotically Brownian functional for  $\lambda = 0$  (unit-root case); locally asymptotically mixed normal (LAMN) for  $\lambda < 0$  (non-ergodic (explosive) case). Turning back to the stable driven case, we should note that the least-squares type estimator would not work unless  $T_n \rightarrow \infty$ , as is expected from Hu and Long (2009) and Zhang and Zhang (2013); there, the authors proved that (when  $\beta$  is known) the rate of convergence when  $\lambda > 0$  equals  $(T_n / \log n)^{1/\beta}$  and the asymptotic distribution is given by a ratio of two independent stable distributions.

## 1.2 Contributions in brief

First, in Sect. 2, we will show that the model is locally asymptotically mixed normal (LAMN) at  $\theta_0 \in \Theta$ , and also that the likelihood equation has a root that is asymptotically efficient in the classical sense of Hajék-Le Cam-Jeganathan. The asymptotic results presented here are uniformly valid in a single manner over any compact subset of the parameter space  $\Theta$ . In particular, the sign of the autoregressive parameter  $\lambda_0$

does not matter, revealing that (i) the results can be described in a unified manner regardless of whether the model is ergodic or not, and also that (ii) the conventional unit-root problem (see Samarakoon and Knight (2009) and the references therein) is not relevant here at all; this is in sharp contrast to the case of ARX time-series models and the Gaussian OU models. Besides, in Sect. 3, we will provide a way to provide an asymptotically efficient estimator through a suboptimal, yet very simple preliminary estimator, which enables us to bypass not only computationally demanding numerical optimization of the likelihood function involving the  $\beta$ -stable density, but also possible multiple-root problem (Lehmann 1999, Section 7.3).

## 2 Local likelihood asymptotic

### 2.1 Preliminaries and result

Let  $P_\theta$  denote the image measure of  $(J, Y)$  associated with the value  $\theta \in \Theta$ . We show the non-trivial stochastic expansion of the log-likelihood ratio of  $P_{\theta+\varphi_n(\theta)v_{n,n}}^Y$  with respect to  $P_{\theta,n}^Y$  for appropriate norming matrix  $\varphi_n(\theta)$  introduced later and bounded sequence  $(v_n) \subset \mathbb{R}^p$ , where  $P_{\theta,n}^Y$  stands for the restriction of  $P_\theta$  to  $\sigma(Y_{t_j} : j \leq n)$ . The distribution  $\mathcal{L}(Y_0)$  may vary according to  $\theta$ ; we will assume that for any  $\epsilon > 0$ , there exists an  $M > 0$  such that  $\sup_{\theta \in \bar{\Theta}} P_\theta[|Y_0| \geq M] < \epsilon$ .

Let  $\phi_\beta$  denote the  $\beta$ -stable density of  $J_1: P_\theta[J_1 \in dy] = \phi_\beta(y)dy$ . It is known that  $\phi_\beta(y) > 0$  for each  $y \in \mathbb{R}$ , that  $\phi_\beta$  is smooth in  $(y, \beta) \in \mathbb{R} \times (0, 2)$ , and that for each  $k, l \in \mathbb{Z}_+$ ,

$$\limsup_{|y| \rightarrow \infty} \frac{|y|^{\beta+1+k}}{\log^l(1 + |y|)} |\partial^k \partial_\beta^l \phi_\beta(y)| < \infty. \tag{2.1}$$

See DuMouchel (1973) for details. Here, we write  $\partial^k \partial_\beta^l \phi_\beta(y) := (\partial^k / \partial y^k)(\partial^l / \partial \beta^l) \phi_\beta(y)$ ; analogous notation for the partial derivatives will be used in the sequel.

To proceed, we need to introduce further notation. Any asymptotics will be taken for  $n \rightarrow \infty$  unless otherwise mentioned. We denote by  $\rightarrow_u$  the uniform convergence of non-random quantities concerning  $\theta$  over  $\bar{\Theta}$ . We write  $C$  for positive universal constant which may vary at each appearance, and  $a_n \lesssim b_n$  when  $a_n \leq Cb_n$  for every  $n$  large enough. Given positive functions  $a_n(\theta)$  and  $b_n(\theta)$ , we write  $b_n(\theta) = o_u(a_n)$  and  $b_n(\theta) = O_u(a_n)$  if  $a_n^{-1}b_n(\theta) \rightarrow_u 0$  and  $\sup_\theta |a_n^{-1}b_n(\theta)| = O(1)$ , respectively. The symbol  $a_n(\theta) \lesssim_u b_n(\theta)$  means that  $\sup_\theta |a_n(\theta)/b_n(\theta)| \lesssim 1$ . We write  $\int_j$  instead of  $\int_{t_{j-1}}^{t_j}$ .

By integrating by parts applied to the process  $t \mapsto e^{\lambda t} Y_t$ , we obtain the explicit càdlàg solution process: under  $P_\theta$ ,

$$Y_t = e^{-\lambda(t-s)} Y_s + \mu \cdot \int_s^t e^{-\lambda(t-u)} X_u du + \sigma \int_s^t e^{-\lambda(t-u)} dJ_u, \quad t > s. \tag{2.2}$$

For  $x, \lambda \in \mathbb{R}$ , we write

$$\eta(x) = \frac{1}{x}(1 - e^{-x}), \quad \zeta_j(\lambda) = \frac{1}{h} \int_j e^{-\lambda(t_j-s)} X_s ds.$$

The basic property of the Lévy integral and the fact that  $\log E_\theta[e^{iuJ_1}] = -|u|^\beta$  give

$$\begin{aligned} \log E_\theta \left[ \exp \left( iu \sigma \int_j e^{-\lambda(t_j-s)} dJ_s \right) \right] &= \int_j \log E_\theta \left[ \exp \left( iue^{-\lambda(t_j-s)} \sigma J_1 \right) \right] ds \\ &= -|\sigma u|^\beta \int_j e^{-\lambda\beta(t_j-s)} ds \\ &= -|\sigma h^{1/\beta} \eta(\lambda\beta h)^{1/\beta} u|^\beta. \end{aligned}$$

Hence,

$$\epsilon_j(\theta) := \frac{Y_{t_j} - e^{-\lambda h} Y_{t_{j-1}} - \mu \cdot \zeta_j(\lambda)h}{\sigma h^{1/\beta} \eta(\lambda\beta h)^{1/\beta}} \stackrel{P_\theta}{\sim} \text{i.i.d. } \mathcal{L}(J_1). \tag{2.3}$$

Now, the exact log-likelihood function  $\ell_n(\theta) = \ell_n(\theta; (X_t)_{t \in [0, T]}, (Y_{t_j})_{j=0}^n)$  is given by

$$\begin{aligned} \ell_n(\theta) &= \sum_{j=1}^n \log \left( \frac{1}{\sigma h^{1/\beta} \eta(\lambda\beta h)^{1/\beta}} \phi_\beta(\epsilon_j(\theta)) \right) \\ &= \sum_{j=1}^n \left( -\log \sigma + \frac{1}{\beta} \log(1/h) - \frac{1}{\beta} \log \eta(\lambda\beta h) + \log \phi_\beta(\epsilon_j(\theta)) \right). \end{aligned} \tag{2.4}$$

We introduce the non-random  $p \times p$ -matrix

$$\varphi_n = \varphi_n(\theta) := \text{diag} \left( \frac{1}{\sqrt{nh^{1-1/\beta}}} I_{1+q}, \frac{1}{\sqrt{n}} \begin{pmatrix} \varphi_{11,n}(\theta) & \varphi_{12,n}(\theta) \\ \varphi_{21,n}(\theta) & \varphi_{22,n}(\theta) \end{pmatrix} \right), \tag{2.5}$$

where the real entries  $\varphi_{kl,n} = \varphi_{kl,n}(\theta)$  are assumed to be continuously differentiable in  $\theta \in \Theta$  and to satisfy the following conditions for some finite values  $\bar{\varphi}_{kl} = \bar{\varphi}_{kl}(\theta)$ :

$$\begin{cases} \varphi_{11,n}(\theta) \rightarrow_u \bar{\varphi}_{11}(\theta), \\ \varphi_{12,n}(\theta) \rightarrow_u \bar{\varphi}_{12}(\theta), \\ s_{21,n}(\theta) := \beta^{-2} \log(1/h_n) \varphi_{11,n}(\theta) + \sigma^{-1} \varphi_{21,n}(\theta) \rightarrow_u \bar{\varphi}_{21}(\theta), \\ s_{22,n}(\theta) := \beta^{-2} \log(1/h_n) \varphi_{12,n}(\theta) + \sigma^{-1} \varphi_{22,n}(\theta) \rightarrow_u \bar{\varphi}_{22}(\theta), \\ \inf_{\theta} |\bar{\varphi}_{11}(\theta) \bar{\varphi}_{22}(\theta) - \bar{\varphi}_{12}(\theta) \bar{\varphi}_{21}(\theta)| > 0, \\ \max_{(k,l)} |\partial_\theta \varphi_{kl,n}(\theta)| \lesssim_u \log^2(1/h). \end{cases} \tag{2.6}$$

The matrix  $\varphi_n(\theta)$  will turn out to be the right norming with which  $u \mapsto \ell_n(\theta + \varphi_n(\theta)u) - \ell_n(\theta)$  under  $P_\theta$  has an asymptotically quadratic structure in  $\mathbb{R}^p$ ;

see Brouste and Masuda (2018) and Clément and Gloter (2020) for the related previous studies. Note that  $\sqrt{nh_n}^{1-1/\beta} \rightarrow_u \infty$  and  $|\varphi_{21,n}(\theta)| \vee |\varphi_{22,n}(\theta)| \lesssim \log(1/h)$ . By the same reasoning as in Brouste and Masuda (2018, page 292), we have  $\inf_{\theta} |\varphi_{11,n}(\theta)\varphi_{22,n}(\theta) - \varphi_{12,n}(\theta)\varphi_{21,n}(\theta)| \gtrsim 1$  and  $|\varphi_n(\theta)| \rightarrow_u 0$  under (2.6).

Let

$$f_{\beta}(y) := \frac{\partial_{\beta}\phi_{\beta}}{\phi_{\beta}}(y), \quad g_{\beta}(y) := \frac{\partial\phi_{\beta}}{\phi_{\beta}}(y),$$

and define the block-diagonal random matrix

$$\mathcal{I}(\theta) = \text{diag}(\mathcal{I}_{\lambda,\mu}(\theta), \mathcal{I}_{\beta,\sigma}(\theta)), \tag{2.7}$$

where, for a random variable  $\epsilon \stackrel{P_{\theta}}{\sim} \int \phi_{\beta}(y)dy$  and by denoting by  $A^{\top}$  the transpose of matrix  $A$ ,

$$\mathcal{I}_{\lambda,\mu}(\theta) := \frac{1}{\sigma^2} E_{\theta} [g_{\beta}(\epsilon)^2] \frac{1}{T} \int_0^T \begin{pmatrix} Y_t^2 & -Y_t X_t^{\top} \\ -Y_t X_t & X_t^{\otimes 2} \end{pmatrix} dt, \tag{2.8}$$

$$\mathcal{I}_{\beta,\sigma}(\theta) := \begin{pmatrix} \bar{\varphi}_{11} & \bar{\varphi}_{12} \\ -\bar{\varphi}_{21} & -\bar{\varphi}_{22} \end{pmatrix}^{\top} \begin{pmatrix} E_{\theta} [f_{\beta}(\epsilon)^2] & E_{\theta} [\epsilon f_{\beta}(\epsilon)g_{\beta}(\epsilon)] \\ E_{\theta} [\epsilon f_{\beta}(\epsilon)g_{\beta}(\epsilon)] & E_{\theta} [(1 + \epsilon g_{\beta}(\epsilon))^2] \end{pmatrix} \begin{pmatrix} \bar{\varphi}_{11} & \bar{\varphi}_{12} \\ -\bar{\varphi}_{21} & -\bar{\varphi}_{22} \end{pmatrix}. \tag{2.9}$$

Note that  $\mathcal{I}(\theta)$  does depend on the choice of  $\bar{\varphi}(\theta) = \{\bar{\varphi}_{kl}(\theta)\}$ ; if  $\bar{\varphi}(\theta)$  is free from  $(\lambda, \mu)$ , then so is  $\mathcal{I}(\theta)$ .

Also, we note that  $\mathcal{I}(\theta) > 0$  ( $P_{\theta}$ -a.s.,  $\theta \in \Theta$ ) under (1.2). Indeed, it was verified in Brouste and Masuda (2018, Theorem 1) that  $\mathcal{I}_{\beta,\sigma}(\theta) > 0$  a.s. To deduce that  $\mathcal{I}_{\lambda,\mu}(\theta) > 0$  a.s., we note that  $\int_0^T Y_t^2 dt > 0$  a.s. and that, by Schwarz’s inequality,

$$\begin{aligned} & u^{\top} \left\{ \int_0^T X_t^{\otimes 2} dt - \left( \int_0^T Y_t X_t dt \right) \left( \int_0^T Y_t^2 dt \right)^{-1} \left( \int_0^T Y_t X_t dt \right)^{\top} \right\} u \\ &= \int_0^T (u \cdot X_t)^2 dt - \left( \int_0^T Y_t^2 dt \right)^{-1} \left( \int_0^T Y_t (u \cdot X_t) dt \right)^2 > 0 \end{aligned}$$

for every nonzero  $u \in \mathbb{R}^q$ , since for any constant real  $\xi$  we have  $Y \neq (u \cdot X)\xi$  a.s. as functions on  $[0, T]$ . Apply the identity  $\det \begin{pmatrix} A & B^{\top} \\ B & C \end{pmatrix} = \det(A) \det(C - BA^{-1}B^{\top})$  to conclude the  $P_{\theta}$ -a.s. positive definiteness of  $\mathcal{I}(\theta)$ .

The normalized score function  $\Delta_n(\theta_0)$  and the normalized observed information matrix  $\mathcal{I}_n(\theta_0)$  are given by

$$\begin{aligned} \Delta_n(\theta) &:= \varphi_n(\theta)^{\top} \partial_{\theta} \ell_n(\theta), \\ \mathcal{I}_n(\theta) &:= -\varphi_n(\theta)^{\top} \partial_{\theta}^2 \ell_n(\theta) \varphi_n(\theta), \end{aligned}$$

respectively. Let  $MN_{p,\theta}(0, \mathcal{I}(\theta)^{-1})$  denote the covariance mixture of  $p$ -dimensional normal distribution, corresponding to the characteristic function  $u \mapsto E_{\theta}[\exp(-u^{\top}$

$\mathcal{I}(\theta)^{-1}u/2]$ . Finally, we write  $M[u] = \sum_i M_i u_i$  for a linear form  $M = \{M_i\}$  and similarly  $Q[u, u] = Q[u^{\otimes 2}] = \sum_{i,j} Q_{ij} u_i u_j$  for a quadratic form  $Q = \{Q_{ij}\}$ . Now, we are ready to state the main claim of this section.

**Theorem 2.1** *The following statements hold for any  $\theta \in \Theta$ .*

(1) *For any bounded sequence  $(v_n) \subset \mathbb{R}^p$ , it holds that*

$$\ell_n(\theta + \varphi_n(\theta)v_n) - \ell_n(\theta) = \Delta_n(\theta)[v_n] - \frac{1}{2}\mathcal{I}_n(\theta)[v_n, v_n] + o_{P_\theta}(1),$$

where we have the convergence in distribution under  $P_\theta$ :  $\mathcal{L}(\Delta_n(\theta), \mathcal{I}_n(\theta)|P_\theta) \Rightarrow \mathcal{L}(\mathcal{I}(\theta)^{1/2}Z, \mathcal{I}(\theta))$ , where  $Z \sim N_p(0, I)$  is independent of  $\mathcal{I}(\theta)$ , defined on an extended probability space.

(2) *There exists a local maximum point  $\hat{\theta}_n$  of  $\ell_n(\theta)$  with  $P_\theta$ -probability tending to 1 for which*

$$\varphi_n(\theta)^{-1}(\hat{\theta}_n - \theta) = \mathcal{I}_n(\theta)^{-1}\Delta_n(\theta) + o_{P_\theta}(1) \Rightarrow MN_{p,\theta}\left(0, \mathcal{I}(\theta)^{-1}\right).$$

It is worth mentioning that the particular non-diagonal form of  $\varphi_n(\theta)$  is, as in Brouste and Masuda (2018), inevitable to deduce the asymptotically non-degenerate joint distribution of the maximum-likelihood estimator (MLE), the good local maximum point  $\hat{\theta}_n$  in Theorem 2.1(2).

**Remark 2.2** Here are some comments on the model timescale.

- (1) We fix the terminal sampling time  $T$ , so that the rate of convergence  $\sqrt{nh}^{1-1/\beta} = n^{1/\beta-1/2}T^{1-1/\beta} = O(n^{1/\beta-1/2})$  for  $(\lambda, \mu)$ . If  $\beta > 1$  (resp.  $\beta < 1$ ), then a longer period would lead to a better (resp. worse) performance of estimating  $(\lambda, \mu)$ . The Cauchy case  $\beta = 1$ , where the two rates of convergence coincide, is exceptional.
- (2) We can explicitly associate a change of the terminal sampling time  $T$  with those of the components of  $\theta$ . Specifically, changing the model timescale from  $t$  to  $tT$  in (1.1), we see that the process

$$Y^T = (Y_t^T)_{t \in [0,1]} := (Y_{tT})_{t \in [0,1]}$$

satisfies exactly the same integral equation as in (1.1), except that  $\theta = (\lambda, \mu, \beta, \sigma)$  is replaced by

$$\theta_T = (\lambda_T, \mu_T, \beta_T, \sigma_T) := (T\lambda, T\mu, \beta, T^{1/\beta}\sigma)$$

( $\beta$  is unchanged),  $X_t$  by  $X_t^T := X_{tT}$ , and  $J_t$  by  $J_t^T := T^{-1/\beta}J_{tT}$ :

$$Y_t^T = Y_0^T + \int_0^t (\mu_T \cdot X_s^T - \lambda_T Y_s^T) ds + \sigma_T J_t^T, \quad t \in [0, 1].$$

Note that  $(J_t^T)_{t \in [0,1]}$  defines the standard  $\beta$ -stable Lévy process. This indeed shows that we may set  $T \equiv 1$  in the virtual (model) world without loss of generality. This

is impossible for diffusion-type models, where we cannot consistently estimate the drift coefficient unless we let the terminal sampling time  $T$  tend to infinity.

**Remark 2.3** The present framework allows us to do unit period-wise, for example, day-by-day inference for both trend and scale structures, providing a sequence of period-wise estimates with theoretically valid approximate confidence sets. This, though informally, suggests an aspect of change-point analysis in high-frequency data: if we have high-frequency sample over  $[k - 1, k]$  for  $k = 1, \dots, [T]$ , then we can construct a sequence of estimators  $\{\hat{\theta}_n(k)\}_{k=1}^{[T]}$ ; then it would be possible in some way to reject the constancy of  $\theta$  over  $[0, [T]]$  if  $k \mapsto \hat{\theta}_n(k)$  ( $k = 1, \dots, [T]$ ) is not likely to stay unchanged.

**Remark 2.4** It is formally straightforward to extend the model (1.1) to the following form:

$$Y_t = Y_0 + \int_0^t \{a(X_s, \lambda)Y_s + b(X_s, \mu)\} ds + \int_0^t c(X_{s-}, \sigma)dJ_s, \quad t \in [0, T].$$

Under mild regularity conditions on the function  $(a, b, c)$  as well as on the non-random process  $X$ , a solution process  $Y$  is explicitly given by (see Cheridito et al. 2003, Appendix)

$$Y_t = e^{\psi(s,t;\lambda)} Y_s + \int_s^t e^{\psi(u,t;\lambda)} (b(X_u, \mu)du + c(X_{u-}, \sigma)dJ_u), \quad 0 \leq s < t,$$

where  $\psi(s, t; \lambda) := \int_s^t a(X_v, \lambda)dv$ . However, the corresponding likelihood asymptotics becomes much messier. It is worth mentioning that the optimal rate matrix can be diagonal if, for example,  $\partial_t \partial_\theta \log c(t, \sigma) \neq 0$  with  $X_t = t$ : for details, see the previous study Clément and Gloter (2020) that treated the general time-homogeneous Markovian case.

### 2.2 Proof of Theorem 2.1

In this proof, we make use of the general result Sweeting (1980) about the exact-likelihood asymptotics in a more or less analogous way to that of Brouste and Masuda (2018, Theorem 1): under the uniform nature of the exact-likelihood asymptotics, we will deduce the joint convergence in distribution of the normalized score  $\Delta_n(\theta_0)$  and the normalized observed information  $\mathcal{I}_n(\theta_0)$  from the uniform convergence in probability of  $\mathcal{I}_n(\cdot)$  in an appropriate sense. Consequently, we will not need to derive the stable convergence in law of  $\Delta_n(\theta_0)$ , which is often crucial when concerned with high-frequency sampling for a process with dependent increments.

We have  $\sup_{t \in [0, T]} |X_t| < \infty$ , since  $X : [0, T] \rightarrow \mathbb{R}^q$  is assumed to be càdlàg. Through the localization procedure, we may and do suppose that the driving stable Lévy process does not have jumps of size greater than some fixed threshold (see Masuda 2019, Section 6.1 for a concise account). In that case, the Lévy measure of  $J$  is compactly supported; hence in particular,

$$\sup_{\theta \in \Theta} E_{\theta} \left[ |J_1|^K \right] < \infty \tag{2.10}$$

for any  $K > 0$ . Further, since the Lévy measure of  $J$  is symmetric, the removal of large-size jumps does not change the parametric form of the drift coefficient. We also localize the initial variable  $Y_0$  so that  $|Y_0|$  is essentially bounded uniformly in  $\theta$ . It follows from (2.2) and (2.10) that  $\sup_{\theta \in \Theta} \sup_{0 \leq t \leq T} E_{\theta} \left[ |Y_t|^K \right] < \infty$  for  $t \in [0, T]$  as well.

To proceed, we introduce some further notation. Given continuous random functions  $\xi_0(\theta)$  and  $\xi_n(\theta)$ ,  $n \geq 1$ , we write  $\xi_n(\theta) \xrightarrow{p}_u \xi_0(\theta)$  if the joint distribution of  $\xi_n$  and  $\xi_0$  are well defined under  $P_{\theta}$  and if  $P_{\theta} [|\xi_n(\theta) - \xi_0(\theta)| > \epsilon] \rightarrow_u 0$  for every  $\epsilon > 0$  as  $n \rightarrow \infty$ . Additionally, for a sequence  $a_n > 0$ , we write  $\xi_n(\theta) = o_{u,p}(a_n)$  if  $a_n^{-1} \xi_n(\theta) \xrightarrow{p}_u 0$ , and also  $\xi_n(\theta) = O_{u,p}(a_n)$  if for every  $\epsilon > 0$  there exists a constant  $K > 0$  for which  $\sup_{\theta} P_{\theta} [|a_n^{-1} \xi_n(\theta)| > K] < \epsilon$ . Similarly, for any random functions  $\chi_{nj}(\theta)$  doubly indexed by  $n$  and  $j \leq n$ , we write  $\xi_{nj}(\theta) = O_p^*(a_n)$  if

$$\sup_n \max_{j \leq n} \sup_{\theta} E_{\theta} \left[ |a_n^{-1} \chi_{nj}(\theta)|^K \right] < \infty$$

for any  $K > 0$ . Finally, let

$$\mathfrak{N}_n(c; \theta) := \left\{ \theta' \in \Theta : |\varphi_n(\theta)^{-1}(\theta' - \theta)| \leq c \right\}.$$

We will complete the proof of Theorem 2.1 by verifying the three statements corresponding to the conditions (12), (13), and (14) in Brouste and Masuda (2018), which here read

$$\mathcal{I}_n(\theta) \xrightarrow{p}_u \mathcal{I}(\theta), \tag{2.11}$$

$$\sup_{\theta' \in \mathfrak{N}_n(c; \theta)} |\varphi_n(\theta')^{-1} \varphi_n(\theta) - I_p| \rightarrow_u 0, \tag{2.12}$$

$$\sup_{\theta^1, \dots, \theta^p \in \mathfrak{N}_n(c; \theta)} \left| \varphi_n(\theta)^{\top} \{ \partial_{\theta}^2 \ell_n(\theta^1, \dots, \theta^p) - \partial_{\theta}^2 \ell_n(\theta) \} \varphi_n(\theta) \right| \xrightarrow{p}_u 0, \tag{2.13}$$

respectively, where (2.12) and (2.13) should hold for all  $c > 0$  and where  $\partial_{\theta}^2 \ell_n(\theta^1, \dots, \theta^p)$ ,  $\theta^k \in \Theta$ , denotes the  $p \times p$  Hessian matrix of  $\ell_n(\theta)$ , whose  $(k, l)$ th element is given by  $\partial_{\theta_k} \partial_{\theta_l} \ell_n(\theta^k)$ , in which  $\theta =: (\theta_l)_{l=1}^p$ . Having obtained (2.11), (2.12) and (2.13), Sweeting (1980, Theorem 1 and 2) immediately concludes Theorem 2.1. We can verify (2.12) exactly as in Brouste and Masuda (2018), so we will look at (2.11) and (2.13).

*Proof of (2.11).* Recall the expression (2.4). To look at the entries of  $\partial_{\theta}^2 \ell_n(\theta)$ , we introduce several shorthands for notational convenience; they may look somewhat daring, but would not bring confusion. Let us omit the subscript  $\beta$  and the argument  $\epsilon_j$  of the aforementioned notation, such as  $\phi := \phi_{\beta}(\epsilon_j)$ ,  $g := g_{\beta}(\epsilon_j)$  and so on. For



brevity, we also write

$$l' = \log(1/h), \quad c = \eta(\lambda\beta h)^{-1/\beta}, \quad \epsilon = \epsilon_j(\theta),$$

so that (2.4) becomes

$$\ell_n(\theta) = \sum_{j=1}^n \left( -\log \sigma + \frac{1}{\beta} l' + \log c + \log \phi \right).$$

Further, the partial differentiation with respect to a variable will be denoted by the braced subscript such as  $\epsilon_{(a)} := \partial_a \epsilon_j(\theta)$  and  $\epsilon_{(a,b)} := \partial_a \partial_b \epsilon_j(\theta)$ . Then, direct computations give the first-order partial derivatives:

$$\begin{aligned} \partial_\lambda \ell_n(\theta) &= \sum_{j=1}^n \left( (\log c)_{(\lambda)} + \epsilon_{(\lambda)} g \right), \\ \partial_\mu \ell_n(\theta) &= \sum_{j=1}^n \epsilon_{(\mu)} g, \\ \partial_\beta \ell_n(\theta) &= \sum_{j=1}^n \left( -\beta^{-2} l' + (\log c)_{(\beta)} + \epsilon_{(\beta)} g + f \right), \\ \partial_\sigma \ell_n(\theta) &= \sum_{j=1}^n \left( -\sigma^{-1} + \epsilon_{(\sigma)} g \right), \end{aligned}$$

followed by the second-order ones:

$$\begin{aligned} \partial_\lambda^2 \ell_n(\theta) &= \sum_{j=1}^n \left\{ (\epsilon_{(\lambda)})^2 (\partial g) + \epsilon_{(\lambda,\lambda)} g + (\log c)_{(\lambda,\lambda)} \right\}, \\ \partial_\mu^2 \ell_n(\theta) &= \sum_{j=1}^n (\epsilon_{(\mu)})^2 (\partial g), \\ \partial_\beta^2 \ell_n(\theta) &= \sum_{j=1}^n \left\{ 2\beta^{-3} l' + (\log c)_{(\beta,\beta)} + \epsilon_{(\beta,\beta)} g + \epsilon_{(\beta)} g_{(\beta)} \right. \\ &\quad \left. + (\epsilon_{(\beta)})^2 (\partial g) + f_{(\beta)} + \epsilon_{(\beta)} (\partial f) \right\}, \\ \partial_\sigma^2 \ell_n(\theta) &= \sum_{j=1}^n \left\{ \sigma^{-2} + (\epsilon_{(\sigma)})^2 (\partial g) + \epsilon_{(\sigma,\sigma)} g \right\}, \\ \partial_\lambda \partial_\mu \ell_n(\theta) &= \sum_{j=1}^n \left\{ \epsilon_{(\lambda)} \epsilon_{(\mu)} (\partial g) + \epsilon_{(\mu,\lambda)} g \right\}, \end{aligned}$$

$$\begin{aligned} \partial_\lambda \partial_\beta \ell_n(\theta) &= \sum_{j=1}^n \{(\log c)_{(\lambda, \beta)} + \epsilon_{(\lambda)} g(\beta) + \epsilon_{(\beta)} \epsilon_{(\lambda)} (\partial g) + \epsilon_{(\lambda, \beta)} g\}, \\ \partial_\lambda \partial_\sigma \ell_n(\theta) &= \sum_{j=1}^n \{\epsilon_{(\sigma)} \epsilon_{(\lambda)} (\partial g) + \epsilon_{(\lambda, \sigma)} g\}, \\ \partial_\mu \partial_\beta \ell_n(\theta) &= \sum_{j=1}^n \{\epsilon_{(\mu)} g(\beta) + \epsilon_{(\beta)} \epsilon_{(\mu)} (\partial g) + \epsilon_{(\beta, \mu)} g\}, \\ \partial_\mu \partial_\sigma \ell_n(\theta) &= \sum_{j=1}^n \{\epsilon_{(\mu)} \epsilon_{(\sigma)} (\partial g) + \epsilon_{(\mu, \sigma)} g\}, \\ \partial_\beta \partial_\sigma \ell_n(\theta) &= \sum_{j=1}^n \{\epsilon_{(\sigma)} g(\beta) + \epsilon_{(\beta)} \epsilon_{(\sigma)} (\partial g) + \epsilon_{(\beta, \sigma)} g\}. \end{aligned}$$

It is straightforward to see which term is the leading one in each expression above. We do not list all the details here, but for later reference mention a few of the points:

- $\partial^k \log \eta(y) = O_u(1)$  for  $|y| \rightarrow 0$  whatever  $k \in \mathbb{Z}_+$  is;
- $(\log c)_{(\lambda, \dots, \lambda)} = O_u(h^k)$  ( $k$ -times,  $k \in \mathbb{Z}_+$ ),  $(\log c)_{(\lambda, \beta)} = O_u(h^2)$ ,  $(\log c)_{(\beta)} = O_u(h^2)$ ,  $(\log c)_{(\beta, \beta)} = O_u(h^4)$ , and so forth;
- $\max_{j \leq n} |\partial_\lambda^k \zeta_j(\lambda)| = O(h^k)$  for  $k \in \mathbb{Z}_+$ ;
- recalling the definition (2.3) and because of the consequence (2.10) of the localization, concerning the partial derivatives of  $\epsilon_j(\theta)$  we obtain the asymptotic representations:  $\epsilon_{(\mu, \sigma)} = (1 + o_u(1))\sigma^{-2}h^{1-1/\beta}$ ,  $\epsilon_{(\mu, \lambda)} = (1 + o_u(1))\sigma^{-1}h^{2-1/\beta}/2$ ,  $\epsilon_{(\sigma, \lambda)} = (1 + o_u(1))\{-\sigma^{-2}h^{1-1/\beta}Y_{t_{j-1}} + O_p^*(h \vee h^{2-1/\beta})\}$ ,  $\epsilon_{(\lambda, \lambda)} = O_p^*(h^{2-1/\beta})$ ,  $\epsilon_{(\beta, \beta)} = O_p^*(h^2(l')^2) + \epsilon O_p^*((l')^2)$ ,  $\epsilon_{(\lambda, \beta)} = O_p^*(l'h^{1-1/\beta})$ , and so on; the terms “ $o_u(1)$ ” therein are all valid uniformly in  $j \leq n$ .

Now, we write

$$\mathcal{I}_n(\theta) = \begin{pmatrix} \mathcal{I}_{11,n}(\theta) & \mathcal{I}_{12,n}(\theta) \\ \mathcal{I}_{12,n}(\theta)^\top & \mathcal{I}_{22,n}(\theta) \end{pmatrix}$$

with  $\mathcal{I}_{11,n}(\theta) \in \mathbb{R}^{1+q} \otimes \mathbb{R}^{1+q}$ ,  $\mathcal{I}_{22,n}(\theta) \in \mathbb{R}^2 \otimes \mathbb{R}^2$  and  $\mathcal{I}_{12,n}(\theta) \in \mathbb{R}^{1+q} \otimes \mathbb{R}^2$ . We can deduce  $\mathcal{I}_{22,n}(\theta) \xrightarrow{p}_u \mathcal{I}_{\beta, \sigma}(\theta)$  in exactly the same way as in the proof of Eq.(12) in Brouste and Masuda (2018). Below, we will show  $\mathcal{I}_{11,n}(\theta) \xrightarrow{p}_u \mathcal{I}_{\lambda, \mu}(\theta)$  and  $\mathcal{I}_{12,n}(\theta) \xrightarrow{p}_u 0$ .

The Burkholder inequality ensures that

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n \pi(X_{t_{j-1}}, Y_{t_{j-1}}; \theta) U(\epsilon_j(\theta)) = O_{u,p}(1) \tag{2.14}$$

for any continuous  $\pi(x, y; \theta)$  and for any  $U(\epsilon_j(\theta))$  such that  $E_\theta[U(\epsilon_j(\theta))] = 0$  ( $\theta \in \Theta$ ) and that the left-hand side of (2.14) is continuous over  $\theta \in \overline{\Theta}$ . Also, note that

the right continuity of  $t \mapsto X_t$  implies that  $(X_t^{\otimes 1} := X_t)$

$$\lim_{n \rightarrow \infty} \max_{l=1,2} \max_{j \leq n} \left| \frac{1}{h} \int_j (X_s^{\otimes l} - X_{t_{j-1}}^{\otimes l}) ds \right| = 0.$$

These basic facts will be repeatedly used below without mentioning them.

For convenience, we will write

$$r_n = r_n(\beta) = \sqrt{n}h^{1-1/\beta} \tag{2.15}$$

and denote by  $\mathbf{1}_{u,p}$  any random array  $\xi_{nj}(\theta)$  such that  $\max_{j \leq n} |\xi_{nj}(\theta) - 1| \xrightarrow{P} 0$ . Direct computations give the following expressions for the components of  $\mathcal{I}_{11,n}(\theta) = -r_n^{-2} \partial_{(\lambda,\mu)}^2 \ell_n(\theta)$ :

$$\begin{aligned} -\frac{1}{r_n^2} \partial_\mu^2 \ell_n(\theta) &= -\frac{1}{n} \sum_{j=1}^n \sigma^{-2} (\partial g) \zeta_j(\lambda)^{\otimes 2} + o_{u,p}(1) \\ &= \frac{1}{n} \sum_{j=1}^n \sigma^{-2} g^2 \zeta_j(\lambda)^{\otimes 2} + o_{u,p}(1), \\ -\frac{1}{r_n^2} \partial_\lambda^2 \ell_n(\theta) &= -\frac{1}{n} \sum_{j=1}^n \sigma^{-2} (\partial g) Y_{t_{j-1}}^2 \mathbf{1}_{u,p} + o_{u,p}(1) + O_{u,p}(h^{1/\beta}) \\ &= \frac{1}{n} \sum_{j=1}^n \sigma^{-2} g^2 Y_{t_{j-1}}^2 + o_{u,p}(1), \\ -\frac{1}{r_n^2} \partial_\lambda \partial_\mu \ell_n(\theta) &= -\frac{1}{n} \sum_{j=1}^n \left\{ (\partial g) \left( \mathbf{1}_{u,p} \sigma^{-1} Y_{t_{j-1}} \zeta_j(\lambda) + O_p^*(h^{1/\beta}) \right) \right. \\ &\quad \left. (-\sigma^{-1} \mathbf{1}_{u,p}) + O_p^*(h^{1/\beta}) \right\} \\ &= \mathbf{1}_{u,p} \left( -\frac{1}{n} \sum_{j=1}^n \sigma^{-2} (\partial g) Y_{t_{j-1}} \zeta_j(\lambda) + o_{u,p}(1) \right) + o_{u,p}(1) \\ &= \frac{1}{n} \sum_{j=1}^n \sigma^{-2} g^2 Y_{t_{j-1}} \zeta_j(\lambda) + o_{u,p}(1). \end{aligned}$$

We can deduce that  $\mathcal{I}_{11,n}(\theta) \xrightarrow{P} \mathcal{I}_{\lambda,\mu}(\theta)$  as follows.

- First, noting that  $\epsilon_j = \epsilon_j(\theta) \stackrel{P_\theta}{\text{i.i.d.}} \mathcal{L}(J_1)$ , we make the compensation  $g^2 = E_\theta[g^2] + (g^2 - E_\theta[g^2])$  in the summands in the rightmost sides of the last three displays and then pick up the leading part involving  $E_\theta[g^2]$ ; the other one becomes negligible by the Burkholder inequality.

- Then, the a.s. Riemann integrability of  $t \mapsto (X_t(\omega), Y_t(\omega))$  allows us to conclude that, for  $k, l \in \{0, 1, 2\}$  and under  $P_\theta$  for each  $\theta$ ,

$$\begin{aligned}
 D_n(k, l) &:= \left| \frac{1}{n} \sum_{j=1}^n Y_{t_{j-1}}^k X_{t_{j-1}}^{\otimes l} - \frac{1}{T} \int_0^T Y_t^k X_t^{\otimes l} dt \right| \\
 &\lesssim \frac{1}{n} \sum_{j=1}^n \frac{1}{h} \int_j \left( |Y_t - Y_{t_{j-1}}| (1 + |Y_t| + |Y_{t_{j-1}}|)^C \right. \\
 &\quad \left. + |Y_t|^k \left| \left( \frac{1}{h} \int_j X_t dt + O(h) \right)^{\otimes l} - X_t^{\otimes l} \right| \right) dt \\
 &\lesssim \frac{1}{n} \sum_{j=1}^n \frac{1}{h} \int_j \left( |Y_t - Y_{t_{j-1}}| (1 + |Y_t| + |Y_{t_{j-1}}|)^C + |Y_t|^k o(1) \right) dt \xrightarrow{P} 0,
 \end{aligned}$$

where the order symbols in the estimates are valid uniformly in  $j \leq n$ . By (2.2), under the localization, we have  $\max_{j \leq n} \sup_\theta E_\theta[|Y_t|^M] = O(1)$  and  $\max_{j \leq n} \sup_\theta E_\theta[|Y_t - Y_{t_{j-1}}|^M] = o_u(1)$  for any  $M > 0$ , from which it follows that  $D_n(k, l) = o_{u,p}(1)$ .

Specifically, for the case of  $-r_n^{-2} \partial_\mu^2 \ell_n(\theta)$ , we have

$$\begin{aligned}
 -\frac{1}{r_n^2} \partial_\mu^2 \ell_n(\theta) &= \frac{1}{n} \sum_{j=1}^n \sigma^{-2} E_\theta[g^2] \zeta_j(\lambda)^{\otimes 2} + o_{u,p}(1) \\
 &= \sigma^{-2} E_\theta[g_\beta(\epsilon_1(\theta))^2] \frac{1}{T} \int_0^T X_t^{\otimes 2} dt + o_{u,p}(1) \xrightarrow{P} \mathcal{I}_{\lambda,\mu;22}(\theta)
 \end{aligned}$$

with  $\mathcal{I}_{\lambda,\mu;22}(\theta)$  denoting the lower left  $q \times q$  component of  $\mathcal{I}_{\lambda,\mu}(\theta)$ . The others can be handled analogously.

Next, we turn to looking at  $\mathcal{I}_{12,n}(\theta) = \{\mathcal{I}_{12,n}^{kl}(\theta)\}_{k,l}$ :

$$\begin{aligned}
 \mathcal{I}_{12,n}^{11}(\theta) &= \varphi_{11,n}(\theta) \partial_\lambda \partial_\beta \ell_n(\theta) + \varphi_{21,n}(\theta) \partial_\mu \partial_\beta \ell_n(\theta), \\
 \mathcal{I}_{12,n}^{12}(\theta) &= \varphi_{11,n}(\theta) \partial_\lambda \partial_\sigma \ell_n(\theta) + \varphi_{21,n}(\theta) \partial_\mu \partial_\sigma \ell_n(\theta), \\
 \mathcal{I}_{12,n}^{21}(\theta) &= \varphi_{12,n}(\theta) \partial_\lambda \partial_\beta \ell_n(\theta) + \varphi_{22,n}(\theta) \partial_\mu \partial_\beta \ell_n(\theta), \\
 \mathcal{I}_{12,n}^{22}(\theta) &= \varphi_{12,n}(\theta) \partial_\lambda \partial_\sigma \ell_n(\theta) + \varphi_{22,n}(\theta) \partial_\mu \partial_\sigma \ell_n(\theta).
 \end{aligned}$$

We can deduce that  $\mathcal{I}_{12,n}(\theta) \xrightarrow{P} 0$  just by inspecting the four components separately in a similar way that we managed  $\mathcal{I}_{11,n}(\theta)$ . Let us only mention the lower-left  $q \times 1$  component: recalling the properties (2.1) and  $|\varphi_{22,n}| \lesssim_u l'$ , we see that

$$\mathcal{I}_{12,n}^{21}(\theta) = -\frac{(h^{1-1/\beta})^{-1}}{n} \sum_{j=1}^n \left( \varphi_{12,n}(\log c)_{(\lambda,\beta)} + \varphi_{12,n} \epsilon_{(\lambda)} g(\beta) + \varphi_{12,n} \epsilon_{(\lambda)} \epsilon_{(\beta)} (\partial g) \right)$$

$$\begin{aligned}
 & + \varphi_{12,n} \epsilon_{(\lambda,\beta)} g + \varphi_{22,n} \epsilon_{(\mu)} g(\beta) + \varphi_{22,n} \epsilon_{(\beta)} \epsilon_{(\mu)} (\partial g) + \varphi_{22,n} \epsilon_{(\beta,\mu)} g \Big) \\
 = & O(h^{1+1/\beta}) + O_{u,p}(n^{-1/2} \vee h^{1/\beta}) + O_{u,p}((n^{-1/2} \vee h^{1/\beta}) l') \\
 & + O_{u,p}(n^{-1/2}) + O_{u,p}(n^{-1/2}(l')^2) + O_{u,p}(n^{-1/2}(l')^2) \xrightarrow{P} 0.
 \end{aligned}$$

Thus, the claim (2.11) follows.

*Proof of (2.13).* Note that

$$\sup_{\theta' \in \mathfrak{N}_n(c;\theta)} |\epsilon_j(\theta')| \lesssim_u |\epsilon_j(\theta)| + \bar{s}_{nj}(\theta; c),$$

where  $|\bar{s}_{nj}(\theta; c)| \lesssim o_u(1)(1 + |Y_{t_{j-1}}|)$ . Also, for each  $k, l, m \in \mathbb{Z}_+$ , we have  $P_\theta$ -a.s. the (rough) estimate:

$$\frac{1}{n} \left| \partial_\beta^k \partial_\sigma^l \partial_{(\lambda,\mu)}^m \ell_n(\theta) \right| \lesssim_u (l')^k h^{(1-1/\beta)m} \frac{1}{n} \sum_{j=1}^n (1 + |Y_{t_{j-1}}|)^m \{1 + \log(1 + |\epsilon_j(\theta)|^2)\}^k.$$

Then, as in the proof of Eq.(14) in Brouste and Masuda (2018), for each  $c > 0$ , we can find a constant  $R = R(c) > 0$  such that (still rough, but sufficient)

$$\begin{aligned}
 & \sup_{\theta^1, \dots, \theta^p \in \mathfrak{N}_n(c;\theta)} \left| \varphi_n(\theta)^\top \{ \partial_\theta^2 \ell_n(\theta^1, \dots, \theta^p) - \partial_\theta^2 \ell_n(\theta) \} \varphi_n(\theta) \right| \\
 & \lesssim_u \sup_{\theta', \theta^1, \dots, \theta^p \in \mathfrak{N}_n(c;\theta)} \left| \varphi_n(\theta)^\top \left\{ \partial_\theta^3 \ell_n(\theta^1, \dots, \theta^p) [\theta' - \theta] \right\} \varphi_n(\theta) \right| \\
 & \lesssim_u \frac{(l')^C}{\sqrt{n}} \sup_{\beta', \beta'' \in \bar{B}(\beta; R/l')} h^{(1/\beta' - 1/\beta'')^3} \sup_{\theta' \in \mathfrak{N}_n(c;\theta)} \frac{1}{n} \sum_{j=1}^n (1 + |Y_{t_{j-1}}|)^m \{1 + \log(1 + |\epsilon_j(\theta')|)\}^3 \\
 & \lesssim_u \frac{(l')^C}{\sqrt{n}} \frac{1}{n} \sum_{j=1}^n (1 + |Y_{t_{j-1}}|)^m \{1 + \log(1 + |\epsilon_j(\theta)|)\}^3 \lesssim O_{u,p} \left( \frac{(l')^C}{\sqrt{n}} \right) \xrightarrow{P} 0,
 \end{aligned}$$

where  $\bar{B}(\beta; R/l'_n)$  denotes the closed ball with center  $\beta$  and radius  $R/l'$ . This shows (2.13). The proof of Theorem 2.1 is complete.

### 3 Asymptotically efficient estimator

From now on, we fix a true value  $\theta_0 \in \Theta$ , and the stochastic symbols and convergences will be taken under  $P := P_{\theta_0}$ ; accordingly, we write  $E := E_{\theta_0}$ . Having Theorem 2.1 in hand, we can proceed with the construction of an asymptotically efficient estimator. It is known that any asymptotically centering estimator  $\hat{\theta}_n^*$ :

$$\varphi_n(\theta_0)^{-1} (\hat{\theta}_n^* - \theta_0) = \mathcal{I}_n(\theta_0)^{-1} \Delta_n(\theta_0) + o_p(1) \tag{3.1}$$

is regular; by Theorem 2.1, the right-hand side converges in distribution to  $MN_{p,\theta_0}(0, \mathcal{I}(\theta_0)^{-1})$ . This, together with the convolution theorem in turn, gives the

asymptotic minimax theorem: for any measurable (loss) function  $\mathfrak{L} : \mathbb{R}^p \rightarrow \mathbb{R}_+$  such that  $\mathfrak{L}(u) = \tau(|u|)$  for some non-decreasing  $\tau : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  with  $\tau(0) = 0$ , we have

$$\liminf_{n \rightarrow \infty} E \left[ \mathfrak{L} \left( \varphi_n(\theta_0)^{-1} (\hat{\theta}_n^* - \theta_0) \right) \right] \geq E \left[ \mathfrak{L}(\mathcal{I}(\theta_0)^{-1/2} Z) \right]. \quad (3.2)$$

Recalling that  $\mathcal{L}(\Delta_n(\theta), \mathcal{I}_n(\theta) | P_\theta) \Rightarrow \mathcal{L}(\mathcal{I}(\theta)^{1/2} Z, \mathcal{I}(\theta))$ , where  $Z \sim N_p(0, I)$  (Theorem 2.1) and in view of the lower bound in (3.2), we may call that any estimator  $\hat{\theta}_n^*$  satisfying (3.1) asymptotically efficient. Again by Theorem 2.1, the good local maximum point  $\hat{\theta}_n$  of  $\ell_n(\theta)$  is asymptotically efficient. We refer to Jeganathan (1982, Theorems 2 and 3, and Proposition 2) and also Jeganathan (1995, Theorem 8) for more information and details of the above arguments.

Theorem 2.1 is based on the classical Cramér-type argument. The well-known shortcoming is its local character: the result just tells us the existence of an asymptotically well behaving root of the likelihood equation, but does not give information about which local maxima is the one when there are multiple local maxima, and equivalently multiple roots for the likelihood equations Lehmann (1999, Section 7.3). Indeed, the log-likelihood function  $\ell_n$  of (2.4) is highly nonlinear and non-concave. In this section, we try to get rid of the locality by a Newton–Raphson type of improvement, which in our case will not only remedy the aforementioned inconvenience of the multiple-root problem, but also enable us to bypass the numerical optimization involving the stable density  $\phi_\beta$ . In Brouste and Masuda (2018, Section 3), for the  $\beta$ -stable Lévy process (the special case of (1.1) with  $\lambda = 0$  and  $X \equiv 1$ ), we provided an initial estimator based on the sample median and the method of moments associated with logarithm and/or lower-order fractional moments. However, it was essential in Brouste and Masuda (2018) that the model is a Lévy process, for which we could apply the median-adjusted central limit theorem for an i.i.d. sequence of random variables. In the present case, we need a different sort of argument.

In Theorem 2.1, the process  $X = (X_t)_{t \in [0, T]}$  was assumed to be observed continuously in  $[0, T]$ . In this section, we will instead deal with a discrete-time sample  $(X_{t_j})_{j=0}^n$  under the additional condition:

$$\exists \kappa \in (1/2, 1], \quad \max_{j \leq n} \left| \frac{1}{h} \int_j^{j+h} (X_t - X_{t_{j-1}}) dt \right| \lesssim h^\kappa. \quad (3.3)$$

We will explicitly construct an estimator  $\hat{\theta}_n^*$  which is asymptotically equivalent to the MLE  $\hat{\theta}_n$ , by verifying the asymptotically centering property (3.1); for this much-thinned sample, we may and do keep calling such a  $\hat{\theta}_n^*$  asymptotically efficient.

### 3.1 Newton–Raphson procedure

To proceed with a discrete-time sample  $\{(X_{t_j}, Y_{t_j})\}_{j=0}^n$ , we introduce the approximate-likelihood function  $\mathbb{H}_n(\theta)$  by replacing  $\zeta_j(\lambda)$  by  $X_{t_{j-1}}$  in the definition (2.4) of the

genuine log-likelihood function  $\ell_n(\theta)$  (recall the notation  $l' := \log(1/h)$ ):

$$\mathbb{H}_n(\theta) = \sum_{j=1}^n \left( -\log \sigma + \frac{1}{\beta} l' - \frac{1}{\beta} \log \eta(\lambda\beta h) + \log \phi_\beta \left( \epsilon'_j(\theta) \right) \right), \tag{3.4}$$

where

$$\epsilon'_j(\theta) := \frac{Y_{t_j} - e^{-\lambda h} Y_{t_{j-1}} - \mu \cdot X_{t_{j-1}} h}{\sigma h^{1/\beta} \eta(\lambda\beta h)^{1/\beta}}. \tag{3.5}$$

Of course, this approximation is not for free: to manage the resulting discretization error specified later on, we additionally impose that

$$\beta_0 > \frac{2}{1 + 2\kappa}. \tag{3.6}$$

Then we have at least  $\beta_0 > 2/3$ , so that small values of  $\beta_0$  are excluded; this is the price we have to pay for dealing with a discrete-time sample from  $X$  in an efficient way. Accordingly, in the sequel, we will reset the parameter space of  $\beta$  to be a domain  $\Theta_\beta$  such that  $\Theta_\beta \subset (2/3, 2)$ .

Toward construction of an asymptotically efficient estimator  $\hat{\theta}_n^*$  satisfying (3.1), we will prove a basic result about a Newton–Raphson type of procedure. As in (2.15), we write  $r_n = r_n(\beta_0) = \sqrt{nh}^{1-1/\beta_0}$ . Write  $n^{-1/2} \tilde{\varphi}_n(\theta)$  for the lower-right  $2 \times 2$ -part of  $\varphi_n(\theta)$ , so that the definition (2.5) with  $\theta = \theta_0$  becomes  $\varphi_n(\theta_0) = \text{diag}(r_n^{-1} I_{q+1}, n^{-1/2} \tilde{\varphi}_n(\theta_0))$ . We then introduce the diagonal matrix

$$\varphi_{0,n} = \varphi_{0,n}(\beta_0) := \text{diag} \left( r_n^{-1} I_{q+1}, n^{-r/2} \begin{pmatrix} 1 & 0 \\ 0 & l' \end{pmatrix} \right) \tag{3.7}$$

for a constant

$$0 < r \leq 1. \tag{3.8}$$

The difference between  $\varphi_n$  and  $\varphi_{0,n}$  is only in the lower-right component for  $(\beta, \sigma)$ , and note that the matrix  $\varphi_n^{-1} \varphi_{0,n}$  may diverge in norm. Then, suppose that we are given an initial estimator  $\hat{\theta}_{0,n} = (\hat{\lambda}_{0,n}, \hat{\mu}_{0,n}, \hat{\beta}_{0,n}, \hat{\sigma}_{0,n})$  such that  $\varphi_{0,n}^{-1}(\hat{\theta}_{0,n} - \theta_0) = O_p(1)$ , namely,

$$\left( r_n(\hat{\lambda}_{0,n} - \lambda_0), r_n(\hat{\mu}_{0,n} - \mu_0), n^{r/2}(\hat{\beta}_{0,n} - \beta_0), \frac{n^{r/2}}{l'}(\hat{\sigma}_{0,n} - \sigma_0) \right) = O_p(1).$$

Let us write  $a = (\lambda, \mu)$  and  $b = (\beta, \sigma)$ . Based on the approximate-likelihood function (3.4) and  $\hat{\theta}_{0,n}$ , we recursively define the  $k$ -step estimator  $\hat{\theta}_{k,n}$  ( $k \geq 1$ ) by

$$\hat{\theta}_{k,n} = \hat{\theta}_{k-1,n} + \left\{ \text{diag} \left( -\partial_a^2 \mathbb{H}_n(\hat{\theta}_{k-1,n}), -\partial_b^2 \mathbb{H}_n(\hat{\theta}_{k-1,n}) \right) \right\}^{-1} \partial_\theta \mathbb{H}_n(\hat{\theta}_{k-1,n}) \tag{3.9}$$

on the event  $F_{k-1,n} := \{ |\det(\partial_a^2 \mathbb{H}_n(\hat{\theta}_{k-1,n}))| \wedge |\det(\partial_b^2 \mathbb{H}_n(\hat{\theta}_{k-1,n}))| > 0 \}$  and assign an arbitrary value to  $\hat{\theta}_{k,n}$  on the complement set  $F_{k-1,n}^c$ ; below, it will be seen (as in

the proof of Theorem 2.1) that  $P[F_{k-1,n}] \rightarrow 1$ . Hence, the arbitrary property does not matter asymptotically and we may and do suppose that  $P[F_{k-1,n}] = 1$  for  $k \geq 1$ . In our subsequent arguments, the inverse-matrix part in (3.9) must be block diagonal: see Remark 3.3 below.

In what follows,  $\hat{\theta}_n$  denotes the good local maxima of the likelihood function  $\ell_n(\theta)$ , when  $(X_t)_{t \leq T}$  is observable; by Theorem 2.1, we have  $P[\partial_\theta \ell_n(\hat{\theta}_n) = 0] \rightarrow 1$  and  $\varphi_n(\theta_0)^{-1}(\hat{\theta}_n - \theta_0) = \mathcal{I}_n(\theta_0)^{-1} \Delta_n(\theta_0) + o_p(1) \Rightarrow MN_{p,\theta_0}(0, \mathcal{I}(\theta_0)^{-1})$ . Define the number

$$K := \min\{k \in \mathbb{N} : 2^{k-1}r > 1/2\} = \min\{k \in \mathbb{N} : k > \log_2(1/r)\}. \tag{3.10}$$

We deduce the asymptotic equivalence of  $\hat{\theta}_n$  and  $\hat{\theta}_{K,n}$ :

$$\varphi_n(\theta_0)^{-1}(\hat{\theta}_{K,n} - \hat{\theta}_n) = o_p(1), \tag{3.11}$$

starting from the initial estimator  $\hat{\theta}_{0,n}$ ; (3.11) concludes (3.1) (hence, (3.2) as well) with  $\hat{\theta}_n^* = \hat{\theta}_{K,n}$ .

We assume that  $\varphi_{0,n}^{-1}(\hat{\theta}_{0,n} - \theta_0) = O_p(1)$  (hence also  $\varphi_{0,n}^{-1}(\hat{\theta}_{0,n} - \hat{\theta}_n) = O_p(1)$ ). Then, to establish (3.11), we first look at the amount of improvement through (3.9) with  $k = 1$ . Write  $\varphi_n = \varphi_n(\theta_0)$  and  $\tilde{\varphi}_n = \tilde{\varphi}_n(\theta_0)$ , and introduce

$$\begin{aligned} \hat{\mathcal{I}}_{0,n} &:= -\varphi_n^\top \text{diag} \left( \partial_a^2 \mathbb{H}_n(\hat{\theta}_{0,n}), \partial_b^2 \mathbb{H}_n(\hat{\theta}_{0,n}) \right) \varphi_n \\ &= \text{diag} \left( -r_n^{-2} \partial_a^2 \mathbb{H}_n(\hat{\theta}_{0,n}), -\frac{1}{n} \tilde{\varphi}_n^\top \partial_b^2 \mathbb{H}_n(\hat{\theta}_{0,n}) \tilde{\varphi}_n \right) =: (\hat{\mathcal{I}}_{0,a,n}, \hat{\mathcal{I}}_{0,b,n}). \end{aligned}$$

We apply Taylor’s expansion around  $\hat{\theta}_n$  to (3.9) with  $k = 1$ : for some random point  $\hat{\theta}'_{0,n}$  on the segment joining  $\hat{\theta}_{0,n}$  and  $\hat{\theta}_n$ ,

$$\begin{aligned} \hat{\theta}_{1,n} - \hat{\theta}_n &= \hat{\theta}_{0,n} - \hat{\theta}_n + \varphi_n \hat{\mathcal{I}}_{0,n}^{-1} \varphi_n^\top \partial_\theta \mathbb{H}_n(\hat{\theta}_{0,n}) \\ &= \varphi_n \hat{\mathcal{I}}_{0,n}^{-1} \left\{ \varphi_n^\top \partial_\theta \mathbb{H}_n(\hat{\theta}_n) \right. \\ &\quad \left. + \varphi_n^\top \left( \text{diag} \left( -\partial_a^2 \mathbb{H}_n(\hat{\theta}_{0,n}), -\partial_b^2 \mathbb{H}_n(\hat{\theta}_{0,n}) \right) - \left( -\partial_b^2 \mathbb{H}_n(\hat{\theta}'_{0,n}) \right) \right) [\hat{\theta}_{0,n} - \hat{\theta}_n] \right\} \\ &=: \varphi_n \hat{\mathcal{I}}_{0,n}^{-1} (R_{0,n} + R''_{0,n}). \end{aligned} \tag{3.12}$$

In what follows, we will derive the rate of convergence of  $\hat{\theta}_{1,n} - \hat{\theta}_n$  in several steps. Here again, we may and do work under the localization (See Sect. 2.2).

*Step 1.* First, we show that  $\hat{\mathcal{I}}_{0,n}^{-1} = O_p(1)$ . We have

$$\hat{\mathcal{I}}_{0,a,n} = -r_n^{-2} \partial_a^2 \mathbb{H}_n(\theta_0) - r_n^{-2} \left( \partial_a^2 \mathbb{H}_n(\hat{\theta}_{0,n}) - \partial_a^2 \mathbb{H}_n(\theta_0) \right), \tag{3.13}$$

$$\hat{\mathcal{I}}_{0,b,n} = -\frac{1}{n} \tilde{\varphi}_n^\top \partial_b^2 \mathbb{H}_n(\theta_0) \tilde{\varphi}_n - \frac{1}{n} \tilde{\varphi}_n^\top \left( \partial_b^2 \mathbb{H}_n(\hat{\theta}_{0,n}) - \partial_b^2 \mathbb{H}_n(\theta_0) \right) \tilde{\varphi}_n. \tag{3.14}$$



The first terms on the right-hand sides above tend to  $\mathcal{I}_{\lambda,\mu}(\theta_0)$  and  $\mathcal{I}_{\beta,\sigma}(\theta_0)$  in probability, respectively. The second terms equal  $o_p(1)$ , by similar considerations to the verification of (2.13) in the proof of Theorem 2.1. Hence,  $\hat{\mathcal{I}}_{0,n} \xrightarrow{P} \mathcal{I}(\theta_0)$  and in particular  $\hat{\mathcal{I}}_{0,n}^{-1} = O_p(1)$  since  $\mathcal{I}(\theta_0)$  is a.s. positive definite.

Step 2. Next, we show that  $R'_{0,n} = \varphi_n^\top \partial_\theta \mathbb{H}_n(\hat{\theta}_n)$  is  $o_p(1)$ . Observe that

$$\varphi_n^\top \partial_\theta \mathbb{H}_n(\hat{\theta}_n) = \varphi_n^\top \partial_\theta \ell_n(\hat{\theta}_n) + \varphi_n^\top \left( \partial_\theta \mathbb{H}_n(\hat{\theta}_n) - \partial_\theta \ell_n(\hat{\theta}_n) \right).$$

For the first term, we have  $\varphi_n^\top \partial_\theta \ell_n(\hat{\theta}_n) = o_p(1)$ , since  $P[|s_n \partial_\theta \ell_n(\hat{\theta}_n)| > \epsilon] \leq P[|\partial_\theta \ell_n(\hat{\theta}_n)| \neq 0] \rightarrow 0$  for every  $\epsilon > 0$  and  $s_n \uparrow \infty$ . To manage the second term, we need to estimate the gap between  $\mathbb{H}_n(\theta)$  and  $\ell_n(\theta)$  by taking the different convergence rates of their components into account. By the definitions (2.4) and (3.4),

$$\begin{aligned} \mathbb{H}_n(\theta) - \ell_n(\theta) &= \sum_{j=1}^n \left( \log \phi_\beta(\epsilon'_j(\theta)) - \log \phi_\beta(\epsilon_j(\theta)) \right) \\ &= \sum_{j=1}^n \left( \int_0^1 g_\beta \left( \epsilon_j(\theta) + s(\epsilon'_j(\theta) - \epsilon_j(\theta)) \right) ds \right) (\epsilon'_j(\theta) - \epsilon_j(\theta)). \end{aligned} \tag{3.15}$$

From the expressions (2.3) and (3.5) and since  $\kappa \leq 1$ , a series of straightforward computations shows that the partial derivatives of

$$\begin{aligned} d_{\epsilon,j}(\theta) &:= \epsilon'_j(\theta) - \epsilon_j(\theta) \\ &= \frac{1}{\sigma h^{1/\beta} \eta(\lambda\beta h)^{1/\beta}} \left( \mu \cdot \int_j (e^{-\lambda(t_j-s)} - 1) X_s ds + \mu \cdot \int_j (X_s - X_{t_{j-1}}) ds \right) \end{aligned}$$

satisfy the following bounds:  $|\partial_\mu d_{\epsilon,j}(\theta)| \lesssim h^{1+\kappa-1/\beta}$ ,  $|\partial_\lambda d_{\epsilon,j}(\theta)| \lesssim h^{2-1/\beta}$ ,  $|\partial_\beta d_{\epsilon,j}(\theta)| \lesssim h^{1+\kappa-1/\beta} l'$ , and  $|\partial_\sigma d_{\epsilon,j}(\theta)| \lesssim h^{1+\kappa-1/\beta}$ . Obviously,  $h^{1/\tilde{\beta}_n-1/\beta_0} = 1 + o_p(1)$  for any  $\tilde{\beta}_n$  such that  $n^v(\tilde{\beta}_n - \beta_0) = O_p(1)$  for some  $v > 0$ ; below, we will repeatedly make use of this fact without mention. Further, under (3.6), it holds that

$$\exists \delta_1 > 0, \quad \sqrt{n} h^{1+\kappa-1/\beta_0} = O(n^{-1/2-\kappa+1/\beta_0}) = O(n^{-\delta_1}). \tag{3.16}$$

By piecing together these observations, the basic property (2.1), and the expression (3.15), under (3.3) we can obtain

$$\begin{aligned} &\left| \varphi_n^\top \left( \partial_\theta \mathbb{H}_n(\hat{\theta}_n) - \partial_\theta \ell_n(\hat{\theta}_n) \right) \right| \\ &\lesssim \left| r_n^{-1} \partial_\mu \mathbb{H}_n(\hat{\theta}_n) \right| + \left| r_n^{-1} \partial_\lambda \mathbb{H}_n(\hat{\theta}_n) \right| + \left| \frac{l'}{\sqrt{n}} \partial_\beta \mathbb{H}_n(\hat{\theta}_n) \right| + \left| \frac{l'}{\sqrt{n}} \partial_\sigma \mathbb{H}_n(\hat{\theta}_n) \right| \\ &\lesssim O_p \left( \sqrt{n} h^{1+\kappa-1/\beta_0} \vee \sqrt{n} h^\kappa \right) + O_p \left( \sqrt{n} h^{1+\kappa-1/\beta_0} \vee \sqrt{n} h^\kappa \right) \end{aligned}$$

$$\begin{aligned}
 &+ O_p\left(\sqrt{n} h^{1+\kappa-1/\beta_0} (l')^C\right) + O_p\left(\sqrt{n} h^{1+\kappa-1/\beta_0} (l')^C\right) \\
 &\lesssim O_p\left(n^{-\delta_1} \vee n^{1/2-\kappa}\right) \xrightarrow{P} 0.
 \end{aligned}$$

This concludes that  $R'_{0,n} = o_p(1)$ .

*Step 3.* Let  $R''_{0,n} =: (R''_{0,a,n}, R''_{0,b,n}) \in \mathbb{R}^{q+1} \times \mathbb{R}^2$ . The goal of this step is to show  $R''_{0,a,n} = o_p(1)$  and  $R''_{0,b,n} = O_p(n^{1/2-r} (l')^C)$ ; at this stage, the latter component may not be stochastically bounded if  $r \leq 1/2$  (recall (3.8)). We have  $R''_{0,n} = A_{0,n} H_{0,n}$ , where

$$\begin{aligned}
 A_{0,n} &:= \varphi_n^\top \begin{pmatrix} \partial_a^2 \mathbb{H}_n(\hat{\theta}'_{0,n}) - \partial_a^2 \mathbb{H}_n(\hat{\theta}_{0,n}) & \text{sym.} \\ \partial_a \partial_b \mathbb{H}_n(\hat{\theta}'_{0,n}) & \partial_b^2 \mathbb{H}_n(\hat{\theta}'_{0,n}) - \partial_b^2 \mathbb{H}_n(\hat{\theta}_{0,n}) \end{pmatrix} \varphi_n, \\
 H_{0,n} &:= \varphi_n^{-1}(\hat{\theta}_{0,n} - \hat{\theta}_n).
 \end{aligned}$$

Under the assumption  $\varphi_{0,n}^{-1}(\hat{\theta}_{0,n} - \theta_0) = O_p(1)$ , recalling the block-diagonal forms (2.5) and (3.7), we see that

$$H_{0,n} = \varphi_n^{-1} \varphi_{0,n} \varphi_{0,n}^{-1}(\hat{\theta}_{0,n} - \theta_0) - \varphi_n^{-1}(\hat{\theta}_n - \theta_0) = \begin{pmatrix} O_p(1) \\ O_p(n^{(1-r)/2} l') \end{pmatrix}, \tag{3.17}$$

where the components  $O_p(1) \in \mathbb{R}^{q+1}$  and  $O_p(n^{(1-r)/2} l') \in \mathbb{R}^2$ ; here and in what follows, we use the stochastic-order symbols for random variables of different dimensions, which will not cause any confusion.

We will show that all the components of  $A_{0,n}$  are at most  $O_p(n^{-r/2} (l')^C)$ :

$$A_{0,n} = O_p(n^{-r/2} (l')^C). \tag{3.18}$$

For the diagonal parts of  $A_{0,n}$ , from the same arguments as in proving (3.13) and (3.14) with the assumption  $\varphi_{0,n}^{-1}(\hat{\theta}_{0,n} - \theta_0) = O_p(1)$ , it holds that

$$\begin{aligned}
 &\left| r_n^{-2} \left( \partial_a^2 \mathbb{H}_n(\hat{\theta}'_{0,n}) - \partial_a^2 \mathbb{H}_n(\hat{\theta}_{0,n}) \right) \right| + \left| \frac{1}{n} \tilde{\varphi}_n^\top \left( \partial_b^2 \mathbb{H}_n(\hat{\theta}'_{0,n}) - \partial_b^2 \mathbb{H}_n(\hat{\theta}_{0,n}) \right) \tilde{\varphi}_n \right| \\
 &= O_p(n^{-r/2} (l')^C).
 \end{aligned}$$

Write  $\theta = (\theta_l)_{l=1}^p$  and so on, and also let  $\partial_a \partial_b \mathbb{H}_n(\hat{\theta}'_{0,n}) \in \mathbb{R}^2 \times \mathbb{R}^{q+1}$  for the size of the matrix. Then, for the non-diagonal part of  $A_{0,n}$ , we expand it as follows:

$$\begin{aligned}
 &\frac{1}{r_n \sqrt{n}} \tilde{\varphi}_n^\top \partial_a \partial_b \mathbb{H}_n(\hat{\theta}'_{0,n}) =: \frac{1}{r_n \sqrt{n}} \partial_a \partial_b \mathbb{H}_n(\theta_0) \\
 &+ \sum_{l=1}^p \left( \frac{(h^{1-1/\beta_0})^{-1}}{n} \partial_{\theta_l} \partial_a \partial_b \mathbb{H}_n(\hat{\theta}'_{0,n}) \right) (\hat{\theta}'_{0,n,l} - \theta_{0,l}).
 \end{aligned}$$

As in the previous diagonal case, the second term on the right-hand side equals  $O_p(n^{-r/2}(l')^C)$ . As for the first term, we write

$$\frac{1}{r_n\sqrt{n}}\partial_a\partial_b\mathbb{H}_n(\theta_0) = \frac{1}{r_n\sqrt{n}}\partial_a\partial_b\ell_n(\theta_0) + \frac{1}{r_n\sqrt{n}}\partial_a\partial_b(\mathbb{H}_n(\theta_0) - \ell_n(\theta_0)).$$

We have seen the explicit expressions of the components of  $\partial_\theta^2\ell_n(\theta)$  in Sect. 2.2. Based on them, it can be seen that all the components of  $r_n^{-1}n^{-1/2}\partial_a\partial_b\ell_n(\theta_0)$  take the form:

$$\frac{1}{n}\sum_{j=1}^n\pi_{j-1}(\theta_0)\psi(\epsilon_j(\theta_0)) + O(h^2)$$

for some  $\mathcal{F}_{Y_{j-1}}$ -measurable random variable  $\pi_{j-1}(\theta_0)$  such that  $|\pi_{j-1}(\theta_0)| \lesssim (1 + |Y_{j-1}|)(l')^C$  and for some odd function  $\psi$  (hence,  $E[\psi(\epsilon_j(\theta_0))] = 0$ ); the last term “ $O(h^2)$ ” only appears in  $\partial_\lambda\partial_\beta\ell_n(\theta)$ . Burkholder’s inequality for the martingale difference arrays gives  $n^{-1}\sum_{j=1}^n\pi_{j-1}(\theta_0)\psi(\epsilon_j(\theta_0)) = O_p(n^{-1/2}(l')^C)$ . We conclude that  $r_n^{-1}n^{-1/2}\partial_a\partial_b\ell_n(\theta_0) = O_p(n^{-1/2}(l')^C)$ . Next, we write  $\mathbb{H}_n(\theta) - \ell_n(\theta) = \sum_{j=1}^n B_j(\theta)d_{\epsilon,j}(\theta)$  for the expression (3.15). The following estimates hold:  $|d_{\epsilon,j}(\theta)| \lesssim h^{1+\kappa-1/\beta}$ ,  $|\partial_a\partial_b d_{\epsilon,j}(\theta)| \lesssim h^{1+\kappa-1/\beta}(1+l')$ ,  $|B_j(\theta)| \lesssim 1$ ,  $|\partial_a B_j(\theta)| \lesssim (1+|Y_{j-1}|)h^{1-1/\beta}$ ,  $|\partial_b B_j(\theta)| \lesssim 1+l'$ , and  $|\partial_a\partial_b B_j(\theta)| \lesssim (1+l')(1+|Y_{j-1}|)h^{1-1/\beta}$ . Therefore, by (3.16),

$$\begin{aligned} \left| \frac{1}{r_n\sqrt{n}}\partial_a\partial_b(\mathbb{H}_n(\theta_0) - \ell_n(\theta_0)) \right| &= \left| \frac{1}{r_n\sqrt{n}}\sum_{j=1}^n\partial_a\partial_b(B_j(\theta)d_{\epsilon,j}(\theta))\Big|_{\theta=\theta_0} \right| \\ &\lesssim (1+l')h^{1+\kappa-1/\beta_0}\frac{1}{n}\sum_{j=1}^n(1+|Y_{j-1}|) \\ &= O_p\left(\frac{(l')^C}{\sqrt{n}}\right)\sqrt{n}h^{1+\kappa-1/\beta_0} = o_p\left(\frac{(l')^C}{\sqrt{n}}\right). \end{aligned}$$

Since  $r \leq 1$ , we have concluded (3.18).

The desired stochastic orders now follows from (3.17) and (3.18):

$$R''_{0,n} = A_{0,n}H_{0,n} = O_p(n^{-r/2}(l')^C) \left( O_p(1) \right)_{O_p(n^{(1-r)/2}l')} = \left( O_p(1) \right)_{O_p(n^{1/2-r}(l')^C)}. \tag{3.19}$$

*Step 4.* We are now able to derive the convergence rate of  $\hat{\theta}_{1,n} - \hat{\theta}_n$ . Recall the definition (3.10) of  $K \in \mathbb{N}$  and the initial rate of convergence (3.7).

- First, we consider  $r > 1/2$ . Then,  $R''_{0,n} = o_p(1)$  from (3.19), so that we can take  $\varphi_{1,n} = \varphi_n$ : by Steps 1 to 3 and (3.12),  $\varphi_n^{-1}(\hat{\theta}_{1,n} - \hat{\theta}_n) = o_p(1)$ . This means that a single iteration is enough if we can take  $r > 1/2$  from the beginning.

- Turning to  $r \in (0, 1/2]$ , we pick a constant  $\epsilon' \in (0, r/2)$  (hence  $r - \epsilon' > r/2$ ), which is to be taken sufficiently small later. Define

$$\varphi_{1,n} = \varphi_{1,n}(\epsilon') := \text{diag} \left( r_n^{-1} I_{q+1}, n^{-(r-\epsilon')} \begin{pmatrix} 1 & 0 \\ 0 & I' \end{pmatrix} \right).$$

Again by Steps 1–3 and (3.12),  $\varphi_{1,n}^{-1} \varphi_n \hat{\mathcal{I}}_{0,n}^{-1} = \text{diag}(O_p(1), O_p((l')^C n^{r-\epsilon'-1/2}))$  and

$$\begin{aligned} \varphi_{1,n}^{-1}(\hat{\theta}_{1,n} - \hat{\theta}_n) &= \begin{pmatrix} O_p(1) & O \\ O & O_p((l')^C n^{r-\epsilon'-1/2}) \end{pmatrix} \left\{ o_p(1) + \begin{pmatrix} o_p(1) \\ O_p(n^{1/2-r}(l')^C) \end{pmatrix} \right\} \\ &= o_p(1) + \begin{pmatrix} o_p(1) \\ O_p(n^{-\epsilon'}(l')^C) \end{pmatrix} = o_p(1). \end{aligned}$$

It follows that the rate of convergence for estimating  $(\beta, \sigma)$  gets improved from  $\text{diag}(n^{r/2}, n^{r/2}/l')$  of  $\hat{\theta}_{0,n}$  to  $\text{diag}(n^{r-\epsilon'}, n^{r-\epsilon'}/l')$  of  $\hat{\theta}_{1,n}$ ; this can be seen as a matrix-norming counterpart of the (near-)doubling phenomenon in the one-step estimation (see for example Zacks 1971, Section 5.5). To improve the rate further, we apply (3.9) to obtain  $\hat{\theta}_{2,n}$  from  $\hat{\theta}_{1,n}$ , so that the rate of convergence for estimating  $(\beta, \sigma)$  gets improved from  $\text{diag}(n^{r-\epsilon'}, n^{r-\epsilon'}/l')$  to  $\text{diag}(n^{2r-3\epsilon'}, n^{2r-3\epsilon'}/l')$ ; here again, we can control the constant  $\epsilon' > 0$  to be sufficiently small. This procedure is iterated  $K - 1$  times, resulting in the rate  $\text{diag}(n^{2^{K-2}r-\epsilon'_0}, n^{2^{K-2}r-\epsilon'_0}/l')$  with  $\epsilon'_0$  being small enough to ensure that  $2(2^{K-2}r - \epsilon'_0) > 1/2$ . Then, the last ( $K$ th-step) application of (3.9) is the same as in the case of  $r > 1/2$  mentioned above.

These observations conclude (3.11).

Thus, we have arrived at the following claim.

**Theorem 3.1** *Suppose that  $\hat{\theta}_{0,n}$  satisfies that  $\varphi_{0,n}^{-1}(\hat{\theta}_{0,n} - \theta_0) = O_p(1)$  with (3.7) and (3.8), and define  $K$  as in (3.10). Then, the  $K$ -step estimator  $\hat{\theta}_{K,n}$ , defined through (3.9) satisfies (3.11), and hence is asymptotically efficient (by Theorem 2.1):*

$$\varphi_n(\theta_0)^{-1}(\hat{\theta}_{K,n} - \theta_0) = \mathcal{I}_n(\theta_0)^{-1} \Delta_n(\theta_0) + o_p(1) \xrightarrow{\mathcal{L}} MN_{p,\theta_0} \left( 0, \mathcal{I}(\theta_0)^{-1} \right). \tag{3.20}$$

Because of the diagonality of  $\varphi_{0,n}$ , Theorem 3.1 makes it possible to construct an initial estimator  $\hat{\theta}_{0,n} = (\hat{\lambda}_{0,n}, \hat{\mu}_{0,n}, \hat{\beta}_{0,n}, \hat{\sigma}_{0,n})$  individually for each component.

Having (3.20) in hand, we can construct consistent estimators  $\hat{\mathcal{I}}_{\lambda,\mu,n} \xrightarrow{P} \mathcal{I}_{\lambda,\mu}(\theta_0)$  and  $\hat{\mathcal{I}}_{\beta,\sigma,n} \xrightarrow{P} \mathcal{I}_{\beta,\sigma}(\theta_0)$ , and then prove the Studentization:

$$\left( \hat{\mathcal{I}}_{\lambda,\mu,n}^{1/2} \sqrt{n} h^{1-1/\hat{\beta}_{K,n}} \begin{pmatrix} \hat{\lambda}_{K,n} - \lambda_0 \\ \hat{\mu}_{K,n} - \mu_0 \end{pmatrix}, \hat{\mathcal{I}}_{\beta,\sigma,n}^{1/2} \sqrt{n} \tilde{\varphi}_n(\hat{\theta}_{K,n})^{-1} \begin{pmatrix} \hat{\beta}_{K,n} - \beta_0 \\ \hat{\sigma}_{K,n} - \sigma_0 \end{pmatrix} \right) \xrightarrow{\mathcal{L}} N_p(0, I_p). \tag{3.21}$$

Indeed, this follows by noting the following facts.

- For construction of  $\hat{\mathcal{I}}_{\lambda,\mu,n}$  and  $\hat{\mathcal{I}}_{\beta,\sigma,n}$ :

- In the expressions (2.8) and (2.9), we can replace the (Riemann)  $dt$ -integrals by the corresponding sample quantities:

$$\frac{1}{n} \sum_{j=1}^n (Y_{t_{j-1}}^2, Y_{t_{j-1}} X_{t_{j-1}}) \xrightarrow{p} \frac{1}{T} \int_0^T (Y_t^2, Y_t X_t) dt.$$

- The elements of the form  $E_{\theta_0}[H(\epsilon; \beta_0)] = \int H(\epsilon; \beta_0) \phi_{\beta_0}(\epsilon) d\epsilon$  with  $H(\epsilon; \beta)$  smooth in  $\beta$  can be evaluated through a numerical integration involving the density  $\phi_{\beta}(\epsilon)$  and its partial derivatives with respect to  $(\beta, \epsilon)$ , with plugging in the estimate  $\hat{\beta}_{K,n}$  for the value of  $\beta$  (the initial estimator  $\hat{\beta}_{0,n}$  is enough).
- Again, note that  $n^v(\hat{\beta}_{K,n} - \beta_0, \hat{\sigma}_{K,n} - \sigma_0) = o_p(1)$  for any sufficiently small  $v \in (0, 1/2)$ , so that  $h^{1-1/\hat{\beta}_{K,n}}/h^{1-1/\beta_0} = (1/h)^{1/\hat{\beta}_{K,n}-1/\beta_0} \xrightarrow{p} 1$ . The values  $\bar{\varphi}_{lm}(\theta_0)$  contained in  $\mathcal{I}_{\beta, \sigma}(\theta_0)$  are estimated by plugging in  $\hat{\theta}_{K,n}$  in (2.6):

$$\begin{cases} \hat{\beta}_{K,n}^{-2} l' \varphi_{11,n}(\hat{\theta}_{K,n}) + \hat{\sigma}_{K,n}^{-1} \varphi_{21,n}(\hat{\theta}_{K,n}) \xrightarrow{p} \bar{\varphi}_{21}(\theta_0), \\ \hat{\beta}_{K,n}^{-2} l' \varphi_{12,n}(\hat{\theta}_{K,n}) + \hat{\sigma}_{K,n}^{-1} \varphi_{22,n}(\hat{\theta}_{K,n}) \xrightarrow{p} \bar{\varphi}_{22}(\theta_0), \\ \varphi_{11,n}(\hat{\theta}_{K,n}) \xrightarrow{p} \bar{\varphi}_{11}(\theta_0), \\ \varphi_{12,n}(\hat{\theta}_{K,n}) \xrightarrow{p} \bar{\varphi}_{12}(\theta_0). \end{cases}$$

We can replace  $(\hat{\beta}_{K,n}, \hat{\sigma}_{K,n})$  by  $(\hat{\beta}_{0,n}, \hat{\sigma}_{0,n})$  all through the above.

- Since  $\varphi_n^{-1}(\hat{\theta}_{K,n} - \theta_0) = O_p(1)$ , it follows that

$$\begin{aligned} \sqrt{n} \tilde{\varphi}_n(\hat{\theta}_{K,n})^{-1} \begin{pmatrix} \hat{\beta}_{K,n} - \beta_0 \\ \hat{\sigma}_{K,n} - \sigma_0 \end{pmatrix} &= \sqrt{n} \left( \tilde{\varphi}_n(\theta_0)^{-1} + O_p((l')^C n^{-1/2}) \right) \begin{pmatrix} \hat{\beta}_{K,n} - \beta_0 \\ \hat{\sigma}_{K,n} - \sigma_0 \end{pmatrix} \\ &= \sqrt{n} \tilde{\varphi}_n(\theta_0)^{-1} \begin{pmatrix} \hat{\beta}_{K,n} - \beta_0 \\ \hat{\sigma}_{K,n} - \sigma_0 \end{pmatrix} + O_p((l')^C n^{-1/2}) \\ &= \sqrt{n} \tilde{\varphi}_n(\theta_0)^{-1} \begin{pmatrix} \hat{\beta}_{K,n} - \beta_0 \\ \hat{\sigma}_{K,n} - \sigma_0 \end{pmatrix} + o_p(1). \end{aligned}$$

The property (3.21) entails

$$\begin{aligned} &\left| \hat{\mathcal{I}}_{\lambda, \mu, n}^{1/2} \sqrt{n} h^{1-1/\hat{\beta}_{K,n}} \begin{pmatrix} \hat{\lambda}_{K,n} - \lambda_0 \\ \hat{\mu}_{K,n} - \mu_0 \end{pmatrix} \right|^2 \\ &+ \left| \hat{\mathcal{I}}_{\beta, \sigma, n}^{1/2} \sqrt{n} \tilde{\varphi}_n(\hat{\theta}_{K,n})^{-1} \begin{pmatrix} \hat{\beta}_{K,n} - \beta_0 \\ \hat{\sigma}_{K,n} - \sigma_0 \end{pmatrix} \right|^2 \xrightarrow{\mathcal{L}} \chi^2(p) = \chi^2(q + 3), \end{aligned}$$

which can be used for constructing an approximate confidence ellipsoid and for goodness-of-fit testing, in particular variable selection among the components of  $X$ .

**Remark 3.2** From the proof of Theorem 3.1, we see that it is possible to weaken (3.6) as  $\beta_0 > 2/3$  if the integrated-process sequence  $(\int_j X_s ds)_{j=1}^n$  is observable. Moreover, it is possible to remove (3.6) if the model is the Markovian  $Y_t = Y_0 + \int_0^t (\mu - \lambda Y_s) ds + \sigma J_t$  with constant  $\mu \in \mathbb{R}$  with modifying the definition (3.5) as in the estimating function of Clément and Gloter (2020). However, we worked under (3.3) and (3.5) to deal with a possibly time-varying  $X$ .

**Remark 3.3** The standard form of the one-step estimator is not (3.9), but

$$\hat{\theta}_{k,n} = \hat{\theta}_{k-1,n} + \left( -\partial_{\theta}^2 \mathbb{H}_n(\hat{\theta}_{k-1,n}) \right)^{-1} \partial_{\theta} \mathbb{H}_n(\hat{\theta}_{k-1,n}).$$

By inspecting the proof of Theorem 3.1, we found that the off-block-diagonal part  $-\partial_a \partial_b \mathbb{H}_n(\hat{\theta}_{k-1,n})$  made the claim therein invalid. This has happened since the rate of convergence for estimating the component  $b = (\beta, \sigma)$  could be too slow. Still, because of the block-diagonality of the original form (2.7), it seems to be a natural and reasonable strategy to use the block-diagonal form from the beginning of defining (3.9).

**Remark 3.4** The necessity of more than one iteration ( $K \geq 2$ ) would be a technical one. If we could verify the tail-probability estimate  $\sup_n P[|r_n(\hat{\lambda}_{0,n} - \lambda_0, \hat{\mu}_{0,n} - \mu_0)| \geq s] \lesssim s^{-M}$  for a sufficiently large  $M > 0$ , then it is possible to deduce the optimality of the one-step Newton–Raphson procedure even when a strategy of construction  $(\hat{\beta}_{0,n}, \hat{\sigma}_{0,n})$  is not smooth in  $(\hat{\lambda}_{0,n}, \hat{\mu}_{0,n})$  as in the function  $\hat{M}_n(a')$  in Section 3.2.2. However, the model under consideration is heavy tailed and it seems impossible to deduce such a bound since we cannot make use of the localization for that purpose.

### 3.2 Specific preliminary estimators

In this section, we consider a specific construction of  $\hat{\theta}_{0,n} = (\hat{\lambda}_{0,n}, \hat{\mu}_{0,n}, \hat{\beta}_{0,n}, \hat{\sigma}_{0,n})$  satisfying  $\varphi_{0,n}^{-1}(\hat{\theta}_{0,n} - \theta_0) = O_p(1)$  with  $\varphi_{0,n}$  given by (3.7). We keep assuming that the available sample is  $\{(X_{t_j}, Y_{t_j})\}_{j=0}^n$  and the conditions (3.3) and (3.6) are in force. We will proceed in two steps.

- (1) First, we will estimate the trend parameter  $(\lambda, \mu)$  by the least absolute deviation (LAD) estimator, which will turn out to be rate optimal, and asymptotically mixed-normally distributed; although the identification of the asymptotic distribution is not necessary here, it would be of independent interest (see Section 3.2.3).
- (2) Next, by plugging in the LAD estimator we construct a sequence of residuals for the noise term, based on which we will consider the lower-order fractional moment matching.

Recall that we are working under the localization (2.10) by removing large jumps of  $J$ .

### 3.2.1 LAD estimator

Let us recall the autoregressive structure together (2.2) with the approximation of the (non-random) integral:

$$\begin{aligned}
 Y_{t_j} &= e^{-\lambda_0 h} Y_{t_{j-1}} + \mu_0 \cdot \zeta_j(\lambda_0) h + \sigma_0 \int_j e^{-\lambda_0(t_j-s)} dJ_s \\
 &= Y_{t_{j-1}} - \lambda_0 h Y_{t_{j-1}} + \mu_0 \cdot X_{t_{j-1}} h + \sigma_0 \int_j e^{-\lambda_0(t_j-s)} dJ_s + h^{1/\beta_0} \delta'_{j-1}, \quad (3.22)
 \end{aligned}$$

where

$$\delta'_{j-1} = \delta'_{j-1}(\theta_0) := h^{-1/\beta_0} \left( Y_{t_{j-1}} (e^{-\lambda_0 h} - 1 + \lambda_0 h) + \mu_0 \cdot (\zeta_j(\lambda_0) - X_{t_{j-1}}) h \right)$$

is an  $\mathcal{F}_{t_{j-1}}$ -measurable random variable such that

$$|\delta'_{j-1}| \lesssim (1 + |Y_{t_{j-1}}|) h^{1+\kappa-1/\beta_0}. \quad (3.23)$$

We define the LAD estimator  $(\hat{\lambda}_{0,n}, \hat{\mu}_{0,n}) \in \mathbb{R}^{q+1}$  by any element  $(\hat{\lambda}_{0,n}, \hat{\mu}_{0,n}) \in \operatorname{argmin}_{(\lambda, \mu)} M_n(\lambda, \mu)$ , leaving  $(\beta, \sigma)$  unknown, where

$$M_n(\lambda, \mu) := \sum_{j=1}^n |Y_{t_j} - Y_{t_{j-1}} - (-\lambda Y_{t_{j-1}} + \mu \cdot X_{t_{j-1}}) h|. \quad (3.24)$$

This is a slight modification of the previously studied approximate LAD estimator in Masuda (2010) concerning the ergodic locally stable OU process.

We introduce the following convex random function on  $\mathbb{R} \times \mathbb{R}^q$  (recall the notation (2.15)):

$$\Lambda_n(u, v) := \frac{1}{\sigma_0 \eta(\lambda_0 \beta_0 h)^{1/\beta_0} h^{1/\beta_0}} \left\{ M_n \left( \lambda_0 + \frac{u}{r_n}, \mu_0 + \frac{1}{r_n} v \right) - M_n(\lambda_0, \mu_0) \right\}.$$

The minimizer of  $\Lambda_n$  is  $\hat{w}_n := (\hat{u}_n, \hat{v}_n)$ , where  $\hat{u}_n := r_n(\hat{\lambda}_{0,n} - \lambda_0)$  and  $\hat{v}_n := r_n(\hat{\mu}_{0,n} - \mu_0)$ . Further, letting  $z_{j-1} := (-Y_{t_{j-1}}, X_{t_{j-1}})$ ,  $w := (u, v)$ , and

$$\epsilon'_j := \frac{1}{\eta(\lambda_0 \beta_0 h)^{1/\beta_0} h^{1/\beta_0}} \int_j e^{-\lambda_0(t_j-s)} dJ_s \stackrel{P_{\theta_0}}{\sim} \text{i.i.d. } \mathcal{L}(J_1),$$

we also introduce the quadratic random function

$$\Lambda_n^\sharp(w) := \Delta'_n[w] + \frac{1}{2} \Gamma_0[w, w],$$

where

$$\Delta'_n := - \sum_{j=1}^n \frac{1}{s_{0,n}\sqrt{n}} \operatorname{sgn} \left( \epsilon'_j + \delta'_{j-1} \right) z_{j-1},$$

$$\Gamma_0 := \frac{2\phi_{\beta_0}(0)}{\sigma_0^2} \frac{1}{T} \int_0^T \begin{pmatrix} Y_t^2 & -Y_t X_t^\top \\ -Y_t X_t & X_t^{\otimes 2} \end{pmatrix} dt,$$

where  $s_{0,n} := \sigma_0 \eta(\lambda_0 \beta_0 h)^{1/\beta_0} = (1 + o(1))\sigma_0$ . The a.s. positive definiteness of  $\Gamma_0$  (see Sect. 2) implies that  $\operatorname{argmin} \Lambda_n^\sharp$  a.s. consists of the single point  $\hat{w}_n^\sharp := -\Gamma_0^{-1} \Delta'_n$ . Then, our objective is to prove that

$$\hat{w}_n = \hat{w}_n^\sharp + o_p(1). \tag{3.25}$$

The proof is analogous to Mas10ejjs (2010, Proof of Theorem 2.1); hence, we will appropriately omit the full technical details, referring to the corresponding parts therein.

By (3.22) and (3.24), we have

$$\Lambda_n(w) = \sum_{j=1}^n \left( \left| \epsilon'_j + \frac{\delta'_{j-1}}{s_{0,n}} - \frac{1}{s_{0,n}\sqrt{n}} w \cdot z_{j-1} \right| - \left| \epsilon'_j + \frac{\delta'_{j-1}}{s_{0,n}} \right| \right).$$

As in Masuda (2010, Eq.(4.6)), we can write  $\Lambda_n(w) = \Delta'_n[w] + Q_n(w)$ , where

$$Q_n(w) := 2 \sum_{j=1}^n \int_0^{w \cdot z_{j-1} / (s_{0,n}\sqrt{n})} \left\{ I \left( \epsilon'_j + \frac{\delta'_{j-1}}{s_{0,n}} \leq s \right) - I \left( \epsilon'_j + \frac{\delta'_{j-1}}{s_{0,n}} \leq 0 \right) \right\} ds.$$

Let us suppose that

$$\Delta'_n = O_p(1), \tag{3.26}$$

$$Q_n(w) = \frac{1}{2} \Gamma_0[w, w] + o_p(1), \quad w \in \mathbb{R}^{1+q}. \tag{3.27}$$

Then, we can make use of the argument of Hjort and Pollard (2011) to conclude (3.25). To see this, we note the inequality due to Hjort and Pollard (2011, Lemma 2): for any  $\epsilon > 0$ ,

$$P \left[ |\hat{w}_n - \hat{w}_n^\sharp| \geq \epsilon \right] \leq P \left[ \sup_{w: |w - \hat{w}_n^\sharp| \leq \epsilon} |\delta_n(w)| \geq \frac{1}{2} \left( \inf_{\substack{(w,z): |z|=1, \\ w = \hat{w}_n^\sharp + \epsilon z}} \Lambda_n^\sharp(w) - \Lambda_n^\sharp(\hat{w}_n^\sharp) \right) \right],$$



where  $\delta_n(w) := \Lambda_n(w) - \Lambda_n^\sharp(w)$ . Obviously,  $\Lambda_n^\sharp(\hat{w}_n^\sharp) = -(1/2)\Delta'_n \cdot \Gamma_0^{-1}\Delta'_n$ . By straightforward computations, we obtain

$$\inf_{\substack{(w,z): |z|=1, \\ w=\hat{w}_n^\sharp+\epsilon z}} \Lambda_n^\sharp(w) - \Lambda_n^\sharp(\hat{w}_n^\sharp) \geq \epsilon^2 \lambda_{\min}(\Gamma_0).$$

Also, because of the convexity, we have the uniform convergence  $\sup_{w \in A} |\delta_n(w)| \xrightarrow{P} 0$  for each compact  $A \subset \mathbb{R}^{1+q}$  (see Hjort and Pollard (2011, Lemma 1)). Note that  $\hat{w}_n^\sharp = O_p(1)$  by (3.26) and the a.s. positive definiteness of  $\Gamma_0$ . Given any  $\epsilon, \epsilon' > 0$ , we can find sufficiently large  $K > 0$  and  $N \in \mathbb{N}$  for which the following three estimates hold simultaneously:

$$\begin{aligned} \sup_n P[|\hat{w}_n^\sharp| > K] &< \epsilon'/3, \\ \sup_{n \geq N} P \left[ \sup_{w: |w| \leq K+\epsilon} |\delta_n(w)| > \epsilon' \right] &< \epsilon'/3, \\ P \left[ \epsilon' \geq \frac{\epsilon^2}{2} \lambda_{\min}(\Gamma_0) \right] &< \epsilon'/3. \end{aligned}$$

Piecing together the above arguments concludes that, for any  $\epsilon, \epsilon' > 0$ , there exists an  $N \in \mathbb{N}$  such that  $\sup_{n \geq N} P \left[ |\hat{w}_n - \hat{w}_n^\sharp| \geq \epsilon \right] < \epsilon'$ . This establishes (3.25), and it follows that

$$\hat{w}_n = -\Gamma_0^{-1}\Delta'_n + o_p(1) = O_p(1). \tag{3.28}$$

It remains to prove (3.26) and (3.27). Below, we will write  $P^{j-1}$  and  $E^{j-1}$  for the conditional probability and expectation given  $\mathcal{F}_{t_{j-1}}$ , respectively.

*Proof of (3.26)* follows on showing  $\Delta_n(1) = O_p(1)$  and  $R_{1,n} = o_p(1)$ , where

$$\begin{aligned} \Delta_n(t) &:= \sum_{j=1}^{\lfloor nt \rfloor} \frac{1}{\sigma_0 \sqrt{n}} \left\{ \text{sgn} \left( \epsilon'_j + \frac{\delta'_{j-1}}{s_{0,n}} \right) - E_0^{j-1} \left[ \text{sgn} \left( \epsilon'_j + \frac{\delta'_{j-1}}{s_{0,n}} \right) \right] \right\} z_{j-1}, \\ t &\in [0, 1], \\ R_{1,n} &:= \sum_{j=1}^n \frac{1}{\sigma_0 \sqrt{n}} E_0^{j-1} \left[ \text{sgn} \left( \epsilon'_j + \frac{\delta'_{j-1}}{s_{0,n}} \right) z_{j-1} \right]. \end{aligned} \tag{3.29}$$

The (matrix-valued) predictable quadratic variation process of  $\{\Delta_n(\cdot)\}_{t \in [0,1]}$  is given by

$$\langle \Delta_n(\cdot) \rangle_t := \sigma_0^{-2} \frac{1}{n} \sum_{j=1}^{\lfloor nt \rfloor} \left\{ \text{sgn} \left( \epsilon'_j + \frac{\delta'_{j-1}}{s_{0,n}} \right) - E_0^{j-1} \left[ \text{sgn} \left( \epsilon'_j + \frac{\delta'_{j-1}}{s_{0,n}} \right) \right] \right\}^2 z_{j-1}^{\otimes 2}.$$

We apply the Lenglart inequality Jacod and Shiryaev (2003, I.3.31) for the submartingale  $|\Delta_n(t)|^2$ : for any  $K, L > 0$ ,

$$\begin{aligned} \sup_n P \left[ \sup_{t \in [0,1]} |\Delta_n(t)| \geq K \right] &\lesssim \frac{L}{K} + \sup_n P \left[ \frac{1}{n} \sum_{j=1}^n |z_{j-1}|^2 \geq C\sigma_0^2 L \right] \\ &\lesssim \frac{L}{K} + \sup_n P \left[ \frac{1}{n} \sum_{j=1}^n (1 + |Y_{t_{j-1}}|)^2 \geq CL \right]. \end{aligned}$$

We have  $n^{-1} \sum_{j=1}^n (1 + |Y_{t_{j-1}}|)^2 = O_p(1)$ . To conclude that  $\Delta_n := \Delta_n(1) = O_p(1)$ , let  $L$  and  $K$  sufficiently large in this order. To see  $R_{1,n} = o_p(1)$ , we proceed in exactly the same way as in Masuda (2010, pp. 544–545): by partly using (3.6) and (3.23),

$$\begin{aligned} |R_{1,n}| &= \left| \frac{1}{n} \sum_{j=1}^n 2\sqrt{n} z_{j-1} \int_0^{\delta'_{j-1}/s_{0,n}} \phi_{\beta_0}(y) dy \right| \\ &\lesssim \frac{1}{n} \sum_{j=1}^n \sqrt{n} |z_{j-1}| |\delta'_{j-1}| \lesssim \frac{1}{n} \sum_{j=1}^n (1 + |Y_{t_{j-1}}|)^2 \sqrt{nh}^{1+\kappa-1/\beta_0} \\ &= O_p \left( \sqrt{nh}^{1+\kappa-1/\beta_0} \right) = O_p \left( h^{1/2+\kappa-1/\beta_0} \right) = o_p(1). \end{aligned}$$

Thus, we have obtained (3.26), and now we can replace  $\Delta'_n$  by  $\Delta_n$  in (3.28):

$$\hat{w}_n = -\Gamma_0^{-1} \Delta_n + o_p(1) = O_p(1). \tag{3.30}$$

*Proof of (3.27).* We decompose  $Q_n(w) =: \sum_{j=1}^n \zeta_j(w)$  as  $Q_n(w) = Q_{1,n}(w) + Q_{2,n}(w)$ , where  $Q_{1,n}(w) := \sum_{j=1}^n E^{j-1}[\zeta_j(w)]$  and  $Q_{2,n}(w) := \sum_{j=1}^n (\zeta_j(w) - E^{j-1}[\zeta_j(w)])$ . Then, for each  $w \in \mathbb{R}^{1+q}$ , we can readily mimic the flow of Masuda (2010, pp.545–546) (for handling the term  $Q_n(u)$  therein). The sketches are given below.

- We have

$$Q_{1,n}(w) = \frac{1}{2} \Gamma_n[w, w] + A_n(w),$$

where

$$\Gamma_n := \frac{2\phi_{\beta_0}(0)}{s_{0,n}^2} \frac{1}{n} \sum_{j=1}^n \begin{pmatrix} Y_{t_{j-1}}^2 & -Y_{t_{j-1}} X_{t_{j-1}}^\top \\ -Y_{t_{j-1}} X_{t_{j-1}} & X_{t_{j-1}}^{\otimes 2} \end{pmatrix} = \Gamma_0 + o_p(1),$$

and where

$$|A_n(w)| \lesssim \left| \frac{1}{n} \sum_{j=1}^n (w \cdot z_{j-1})^2 \left\{ \phi_{\beta_0} \left( -\frac{\delta'_{j-1}}{s_{0,n}} \right) - \phi_{\beta_0}(0) \right\} \right|$$

$$\begin{aligned}
 & + \left| \sum_{j=1}^n \int_0^{w \cdot z_{j-1} / (s_{0,n} \sqrt{n})} s^2 \int_0^1 (1-y) \partial \phi_{\beta_0} \left( s y - \frac{\delta'_{j-1}}{s_{0,n}} \right) dy ds \right| \\
 & \lesssim \frac{1}{n} \sum_{j=1}^n (1 + |Y_{t_{j-1}}|)^4 (1 + |w|)^4 \left( h^{2(1+\kappa-1/\beta_0)} \vee \frac{1}{n} \vee \frac{h^{1+\kappa-1/\beta_0}}{\sqrt{n}} \right) = o_p(1).
 \end{aligned}$$

- We have  $Q_{2,n}(w) = o_p(1)$ : by the Burkholder–Davis–Gundy inequality,

$$\begin{aligned}
 & E \left[ \left( \sum_{j=1}^n (\zeta_j(w) - E^{j-1}[\zeta_j(w)]) \right)^2 \right] \\
 & \lesssim \sum_{j=1}^n E \left[ \left( \int_0^{|w \cdot z_{j-1} / (s_{0,n} \sqrt{n})|} I \left( \left| \epsilon'_j + \frac{\delta'_{j-1}}{s_{0,n}} \right| \leq s \right) ds \right)^2 \right] \\
 & \lesssim \sum_{j=1}^n \frac{|w|}{\sqrt{n}} E \left[ |z_{j-1}| \int_0^{|w \cdot z_{j-1} / (s_{0,n} \sqrt{n})|} P^{j-1} \left[ \left| \epsilon'_j + \frac{\delta'_{j-1}}{s_{0,n}} \right| \leq s \right] ds \right] \\
 & \lesssim \sum_{j=1}^n \frac{|w|}{\sqrt{n}} E \left[ |z_{j-1}| \int_0^{|w \cdot z_{j-1} / (s_{0,n} \sqrt{n})|} \left( s + \left| \frac{\delta'_{j-1}}{s_{0,n}} \right| \right) ds \right] \\
 & \lesssim (1 + |w|)^3 \frac{1}{n} \sum_{j=1}^n E \left[ (1 + |Y_{t_{j-1}}|)^3 \right] \left( \frac{1}{\sqrt{n}} \vee h^{1+\kappa-1/\beta_0} \right) \\
 & = O \left( \frac{1}{\sqrt{n}} \vee h^{1+\kappa-1/\beta_0} \right) = o(1).
 \end{aligned}$$

Summarizing the above yields (3.27).

The tightness (3.30) is sufficient for our purpose. As a matter of fact, the LAD estimator  $(\hat{\lambda}_{0,n}, \hat{\mu}_{0,n})$  is asymptotically mixed-normally distributed. We give the details in Sect. 3.2.3.

### 3.2.2 Rates of convergence at the moment matching for $(\beta, \sigma)$

The remaining task is to construct a specific estimator  $(\hat{\beta}_{0,n}, \hat{\sigma}_{0,n})$  such that

$$\left( n^{r/2}(\hat{\beta}_{0,n} - \beta_0), \frac{n^{r/2}}{l^r}(\hat{\sigma}_{0,n} - \sigma_0) \right) = O_p(1). \tag{3.31}$$

This can be achieved simply by fitting some appropriate moments; for this purpose, the localization does not make sense, since precise expressions of truly existing moments without the localization come into play. Here we consider, as in Brouste and Masuda (2018), the pair of the absolute moments of order  $r$  and  $2r$ .

Let  $a' \in (0, \beta_0/2)$  and define

$$\hat{M}_n(a') := \frac{1}{n} \sum_{j=1}^n \left| Y_{t_j} - Y_{t_{j-1}} + \hat{\lambda}_{0,n} Y_{t_{j-1}} h - \hat{\mu}_{0,n} \cdot X_{t_{j-1}} h \right|^{a'}.$$

Let

$$\epsilon''_j := \frac{1}{h^{1/\beta_0}} \int_j e^{-\lambda_0(t_j-s)} dJ_s = (1 + o(1))\epsilon'_j,$$

which are approximately i.i.d. with common distribution  $\mathcal{L}(J_1)$ , and also let

$$M_n(a') := \sigma_0^{a'} h^{a'/\beta_0} \frac{1}{n} \sum_{j=1}^n \left| \epsilon''_j \right|^{a'}.$$

We can apply the central limit theorem to ensure that  $\sqrt{n} \left( h^{-a'/\beta_0} \sigma_0^{-a'} M_n(a') - m(a'; \beta_0) \right) = O_p(1)$  as soon as  $a' < \beta_0/2$ , where

$$m(a'; \beta_0) := E[|J_1|^{a'}] = \frac{2^{a'}}{\sqrt{\pi}} \frac{\Gamma((a' + 1)/2) \Gamma(1 - a'/\beta_0)}{\Gamma(1 - a'/2)}.$$

Moreover, it follows from the discussions in Sect. 3.2.1 that

$$h^{-a'/\beta_0} \sigma_0^{-a'} \hat{M}_n(a') = \frac{1}{n} \sum_{j=1}^n \left| \epsilon''_j + \frac{1}{\sqrt{n}} \left( \sqrt{n} \delta'_{j-1} - \hat{w}_n \cdot z_{j-1} \right) \right|^{a'},$$

which in turn gives

$$\begin{aligned} n^{a'/2} \left| h^{-a'/\beta_0} \sigma_0^{-a'} \left( \hat{M}_n(a') - M_n(a') \right) \right| &\leq \frac{1}{n} \sum_{j=1}^n \left( \sqrt{n} |\delta'_{j-1}| + |\hat{w}_n| |z_{j-1}| \right)^{a'} \\ &\lesssim O_p \left( \sqrt{n} h^{1+\kappa-1/\beta_0} \right) + O_p(1) = O_p(1). \end{aligned}$$

It follows that

$$n^{a'/2} \left( h^{-a'/\beta_0} \sigma_0^{-a'} \hat{M}_n(a') - m(a'; \beta_0) \right) = O_p(1 \vee n^{(a'-1)/2}) = O_p(1).$$

Now we want to take  $a' = r, 2r$ , which necessitates that  $r \in (0, \beta_0/4)$  in the current argument. Then, we conclude that

$$n^{r/2} \left( h^{-r/\beta_0} \sigma_0^{-r} \hat{M}_n(r) - m(r; \beta_0), h^{-2r/\beta_0} \sigma_0^{-2r} \hat{M}_n(2r; \beta_0) - m(2r; \beta_0) \right) = O_p(1),$$

so that

$$n^{r/2} \left( \frac{\hat{M}_n(r)^2}{\hat{M}_n(2r)} - \frac{m(r; \beta_0)^2}{m(2r; \beta_0)} \right) = O_p(1).$$

There exists a bijection  $f_r$  such that  $f_r(m(r; \beta)^2/m(2r; \beta)) = \beta$ ; see Brouste and Masuda (2018, Section 3.2) and the references therein for the related details. Therefore, taking  $\hat{\beta}_{0,n} := f_r(\hat{M}_n(r)^2/\hat{M}_n(2r))$  results in  $n^{r/2}(\hat{\beta}_{0,n} - \beta_0) = O_p(1)$ , as was to be shown. The bisection method is sufficient for numerically finding  $\hat{\beta}_{0,n}$ .

Turning to  $\hat{\sigma}_{0,n}$ , we note that

$$n^{r/2} \left( \frac{h^{-r/\beta_0} \hat{M}_n(r)}{m(r; \beta_0)} - \sigma_0^r \right) = O_p(1). \tag{3.32}$$

Let  $\hat{\sigma}_{0,n} := \left( \frac{h^{-r/\hat{\beta}_{0,n}} \hat{M}_n(r)}{m(r; \hat{\beta}_{0,n})} \right)^{1/r}$ : we claim that  $\frac{n^{r/2}}{l'}(\hat{\sigma}_{0,n} - \sigma_0) = O_p(1)$ . Since  $\frac{m(r; \hat{\beta}_{0,n})}{m(r; \beta_0)} = O_p(1)$ ,

$$\left| h^{r(1/\hat{\beta}_{0,n} - 1/\beta_0)} \frac{m(r; \hat{\beta}_{0,n})}{m(r; \beta_0)} - 1 \right| \leq \left| h^{r(1/\hat{\beta}_{0,n} - 1/\beta_0)} - 1 \right| O_p(1) + \left| \frac{m(r; \hat{\beta}_{0,n})}{m(r; \beta_0)} - 1 \right|.$$

Recall that  $n^{r/2}(\hat{\beta}_{0,n} - \beta_0) = O_p(1)$ , hence the second term in the upper bound equals  $O_p(n^{-r/2})$ . As for the first term, using that  $(1/\hat{\beta}_{0,n} - 1/\beta_0)l' = O_p(l'/n^{r/2}) = o_p(1)$ , we observe

$$h^{r(1/\hat{\beta}_{0,n} - 1/\beta_0)} - 1 = \exp \left( r(1/\hat{\beta}_{0,n} - 1/\beta_0)l' \right) - 1 = O_p \left( \frac{l'}{n^{r/2}} \right) = o_p(1).$$

These estimates combined with (3.32) conclude the claim: we have (3.31) for the above constructed  $(\hat{\beta}_{0,n}, \hat{\sigma}_{0,n})$ . Given an  $r \in (0, \beta_0/4)$ , by Theorem 3.1, the  $K$ -step estimator for  $K > \log_2(1/r)$  is asymptotically efficient; if  $\beta_0 > 1$  is supposed beforehand, then we can take an  $r > 1/4$  small enough to ensure that  $K = 2$  is enough.

### 3.2.3 Asymptotic mixed normality of the LAD estimator

Recall (3.30):  $\hat{w}_n = -\Gamma_0^{-1} \Delta_n + o_p(1)$ . To deduce the asymptotic mixed normality, it suffices to identify the appropriate asymptotic distribution of  $(\Delta_n, \Gamma_0)$ , equivalently of  $(\Delta_n, \Gamma_n)$ .

First, we clarify the leading term of  $\Delta_n$  in a simpler form. We have  $E[\text{sgn}(\epsilon'_j)] = 0$  and  $E[\text{sgn}(\epsilon'_j)^2] = 1$ , Observe that  $\Delta_n = \Delta_{0,n} + R_{1,n} + R_{2,n}$ , where  $R_{1,n}$  is given in (3.29) and

$$\Delta_{0,n} := \sum_{j=1}^n \frac{1}{\sigma_0 \sqrt{n}} \text{sgn}(\epsilon'_j) z_{j-1},$$

$$R_{2,n} := \sum_{j=1}^n \frac{1}{\sigma_0 \sqrt{n}} \left( \operatorname{sgn}(\epsilon'_j + \delta'_{j-1}) - \operatorname{sgn}(\epsilon'_j) \right) z_{j-1}.$$

We have already seen that  $R_{1,n} = o_p(1)$ . We claim that  $R_{2,n} = o_p(1)$ . Write  $R_{2,n} = \sum_{j=1}^n \xi_j$ . The claim follows on showing that both  $\sum_{j=1}^n E^{j-1}[\xi_j] = o_p(1)$  and  $|\sum_{j=1}^n E^{j-1}[\xi_j^{\otimes 2}]| = o_p(1)$ , but the first one obviously follows from  $R_{1,n} = o_p(1)$ . The second one can be shown as follows: first, we have

$$\begin{aligned} \sum_{j=1}^n E^{j-1}[\xi_j^{\otimes 2}] &= \sigma_0^{-2} \frac{1}{n} \sum_{j=1}^n E^{j-1} \left[ \left( \operatorname{sgn}(\epsilon'_j + \delta'_{j-1}) - \operatorname{sgn}(\epsilon'_j) \right)^2 \right] z_{j-1}^{\otimes 2} \\ &= 2\sigma_0^{-2} \frac{1}{n} \sum_{j=1}^n \left( 1 - E^{j-1} \left[ \operatorname{sgn}(\epsilon'_j + \delta'_{j-1}) \operatorname{sgn}(\epsilon'_j) \right] \right) z_{j-1}^{\otimes 2}. \end{aligned}$$

Moreover,

$$\begin{aligned} &E^{j-1} \left[ \operatorname{sgn}(\epsilon'_j + \delta'_{j-1}) \operatorname{sgn}(\epsilon'_j) \right] \\ &= \left( \int_{0 \vee (-\delta'_{j-1})}^{\infty} + \int_{-\infty}^{0 \vee (-\delta'_{j-1})} - \int_{-\delta'_{j-1}}^0 - \int_0^{-\delta'_{j-1}} \right) \phi_{\beta_0}(y) dy = 1 + D_{j-1} \end{aligned}$$

for some  $\mathcal{F}_{t_{j-1}}$ -measurable term  $D_{j-1}$  satisfying the estimate  $|D_{j-1}| \lesssim |\delta'_{j-1}| \lesssim (1 + |Y_{t_{j-1}}|)h^{1+\kappa-1/\beta_0}$ . These observations conclude that  $|\sum_{j=1}^n E^{j-1}[\xi_j^{\otimes 2}]| = o_p(1)$ .

It remains to look at  $\Delta_{0,n}$ . The mere convergence in distribution is unsuitable since the matrix  $\Gamma_0$  is random. We will apply the weak limit theorem for stochastic integrals: we refer the reader to Jacod and Shiryaev (2003, VI.6) for a detailed account of the limit theorems as well as the standard notation used below.

We introduce the partial sum process

$$S_t^n := \sum_{j=1}^{[nt]} \frac{1}{\sqrt{n}} \operatorname{sgn}(\epsilon'_j), \quad t \in [0, 1].$$

We apply Jacod (2007, Lemma 4.3) to derive  $S^n \xrightarrow{\mathcal{L}_s} w'$  in  $\mathcal{D}(\mathbb{R})$  (the Skorokhod space of  $\mathbb{R}$ -valued functions, equipped with the Skorokhod topology), where  $w' = (w')_{t \in [0,1]}$  denotes a standard Wiener process defined on an extended probability space and independent of  $\mathcal{F}$ . Here, the symbol  $\xrightarrow{\mathcal{L}_s}$  denotes the  $(\mathcal{F})$ -stable convergence in law, which is strictly stronger than the mere weak convergence and in particular implies the joint weak convergence in  $\mathcal{D}(\mathbb{R}^{q+2})$ :

$$(S^n, H^n) \xrightarrow{\mathcal{L}} (w', H^\infty) \quad (3.33)$$

for any  $\mathbb{R}^{q+1}$ -valued  $\mathcal{F}$ -measurable càdlàg processes  $H^n$  and  $H^\infty$  such that  $H^n \xrightarrow{p} H^\infty$  in  $\mathcal{D}(\mathbb{R}^{q+1})$ .

We note the following two points.

- We have  $S^n \xrightarrow{\mathcal{L}} w'$  in  $\mathcal{D}(\mathbb{R})$ , and for each  $n \in \mathbb{N}$  the process  $(S_t^n)_{t \in [0,1]}$  is an  $(\mathcal{F}_{[nt]/n})$ -martingale such that  $\sup_{n,t} |\Delta S_t^n| \leq 1$ . These facts combined with Jacod and Shiryaev (2003, VI.6.29) imply that the sequence  $(S^n)$  is predictably uniformly tight.
- Given any continuous function  $f : \mathbb{R}^{q+1} \rightarrow \mathbb{R}^{q'}$  (for some  $q' \in \mathbb{N}$ ), we consider the function  $H^n = (H^{1,n}, H^{2,n})$  with

$$H_t^{1,n} := (-Y_{[nt]/n}, X_{[nt]/n}),$$

$$H_t^{2,n} := \frac{1}{n} \sum_{j=1}^{[nt]} f(Y_{t_{j-1}}, X_{t_{j-1}}).$$

Then, we have  $H^{1,n} \xrightarrow{p} H^{1,\infty} := (-Y, X)$  in  $\mathcal{D}(\mathbb{R}^{q+1})$  and  $H^{2,n} \xrightarrow{p} H^{2,\infty} := \int_0^\cdot f(Y_s, X_s) ds$  in  $\mathcal{D}(\mathbb{R}^{q'})$ , with which (3.33) concludes the joint weak convergence in  $\mathcal{D}(\mathbb{R}^{2+q+q'})$ :

$$(S^n, H^{1,n}, H^{2,n}) \xrightarrow{\mathcal{L}} (w', H^{1,\infty}, H^{2,\infty}).$$

With these observations, we can apply Jacod and Shiryaev (2003, VI.6.22) to derive the weak convergence of stochastic integrals:

$$(H_-^{1,n} \cdot S^n, H_-^{2,n}) \xrightarrow{\mathcal{L}} (H_-^{1,\infty} \cdot w', H_-^{2,\infty}),$$

which entails that, for any continuous function  $f$ ,

$$\left( \Delta_{0,n}, \frac{1}{n} \sum_{j=1}^n f(Y_{t_{j-1}}, X_{t_{j-1}}) \right) \xrightarrow{\mathcal{L}} \left( \frac{\sigma_0^{-1}}{T} \int_0^T (-Y_s, X_s) dw'_s, \frac{1}{T} \int_0^T f(Y_s, X_s) ds \right)$$

$$\stackrel{\mathcal{L}}{=} \left( \left\{ \frac{\sigma_0^{-2}}{T} \int_0^T \begin{pmatrix} Y_t^2 & -Y_t X_t^\top \\ -Y_t X_t & X_t^{\otimes 2} \end{pmatrix} dt \right\}^{1/2} Z, \right.$$

$$\left. \frac{1}{T} \int_0^T f(Y_s, X_s) ds \right),$$

where  $Z \sim N(0, 1)$  is independent of  $\mathcal{F}$ . Now, by taking

$$f(x, y) = \frac{2\phi_{\beta_0}(0)}{\sigma_0^2} \begin{pmatrix} y^2 & -x^\top y \\ -xy & x^{\otimes 2} \end{pmatrix},$$

we arrive at

$$(\Delta_{0,n}, \Gamma_0) \xrightarrow{\mathcal{L}} \left( \left\{ \sigma_0^{-2} \frac{1}{T} \int_0^T \begin{pmatrix} Y_t^2 & -Y_t X_t^\top \\ -Y_t X_t & X_t^{\otimes 2} \end{pmatrix} dt \right\}^{1/2} Z, \right.$$

$$\frac{2\phi_{\beta_0}(0)}{\sigma_0^2} \frac{1}{T} \int_0^T \begin{pmatrix} Y_t^2 & -Y_t X_t^\top \\ -Y_t X_t & X_t^{\otimes 2} \end{pmatrix} dt \Bigg).$$

In sum, applying Slutsky's theorem concludes that

$$\hat{w}_n = \Gamma_0^{-1} \Delta_{0,n} + o_p(1) \xrightarrow{\mathcal{L}} MN_{q+1, \theta_0} \left( 0, \frac{\sigma_0^2}{4\phi_{\beta_0}(0)^2} \left\{ \frac{1}{T} \int_0^T \begin{pmatrix} Y_t^2 & -Y_t X_t^\top \\ -Y_t X_t & X_t^{\otimes 2} \end{pmatrix} dt \right\}^{-1} \right).$$

**Acknowledgements** The author should like to thank the anonymous reviewers for their detailed comments, based on which he could fix some essential mistakes and drastically improve the quality of the paper. This work was partly supported by JSPS KAKENHI Grant Number 22H01139 and JST CREST Grant Number JPMJCR2115, Japan.

**Funding** Open access funding provided by The University of Tokyo.

## Declarations

**Conflicts of interest** The author declares that there is no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Borovkova, S., & Schmeck, M. D. (2017). Electricity price modeling with stochastic time change. *Energy Economics*, 63, 51–65.
- Brouste, A., & Masuda, H. (2018). Efficient estimation of stable Lévy process with symmetric jumps. *Statistical Inference for Stochastic Processes*, 21(2), 289–307.
- Challis, C. J., & Schmidler, S. C. (2012). A stochastic evolutionary model for protein structure alignment and phylogeny. *Molecular Biology and Evolution*, 29(11), 3575–3587.
- Cheridito, P., Kawaguchi, H., & Maejima, M. (2003). Fractional Ornstein-Uhlenbeck processes. *Electronic Journal of Probability*, 8(3), 14. electronic.
- Clément, E., & Gloter, A. (2020). Joint estimation for SDE driven by locally stable Lévy processes. *Electronic Journal of Statistics*, 14(2), 2922–2956.
- Doob, J. L. (1942). The Brownian movement and stochastic equations. *Annals of Mathematics*, 2(43), 351–369.
- DuMouchel, W. H. (1973). On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution. *Annals of Statistics*, 1, 948–957.
- Hjort, N. L., Pollard, D. (2011). Asymptotics for minimisers of convex processes. Statistical Research Report, University of Oslo, 1993. Available at arxiv preprint [arXiv:1107.3806](https://arxiv.org/abs/1107.3806).
- Hu, Y., & Long, H. (2009). Least squares estimator for Ornstein-Uhlenbeck processes driven by  $\alpha$ -stable motions. *Stochastic Processes and Their Applications*, 119(8), 2465–2480.
- Jacod, J. (2007). Asymptotic properties of power variations of Lévy processes. *ESAIM: Probability and Statistics*, 11, 173–196.



- Jacod, J., Shiryaev, A. N. (2003). Limit theorems for stochastic processes, volume 288 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. 2nd edn. Springer-Verlag
- Jeganathan, P. (1982). On the asymptotic theory of estimation when the limit of the log-likelihood ratios is mixed normal. *Sankhyā Series A*, 44(2), 173–212.
- Jeganathan, P. (1995). Some aspects of asymptotic theory with applications to time series models. *Econometric Theory*, 11(5), 818–887. Trending multiple time series (New Haven, CT, 1993).
- Jhwueng, D.-C., & Maroulas, V. (2014). Phylogenetic Ornstein-Uhlenbeck regression curves. *Statistics and Probability Letters*, 89, 110–117.
- Lehmann, E. L. (1999). *Elements of large-sample theory*. Springer Texts in Statistics. Springer-Verlag.
- Masuda, H. (2010). Approximate self-weighted LAD estimation of discretely observed ergodic Ornstein-Uhlenbeck processes. *Electronic Journal of Statistics*, 4, 525–565.
- Masuda, H. (2019). Non-Gaussian quasi-likelihood estimation of SDE driven by locally stable Lévy process. *Stochastic Processes and Their Applications*, 129(3), 1013–1059.
- Perninge, M., Knazkins, V., Amelin, M., & Söder, L. (2011). Modeling the electric power consumption in a multi-area system. *European Transactions on Electrical Power*, 21(1), 413–423.
- Samarakoon, D. M. M., & Knight, K. (2009). A note on unit root tests with infinite variance noise. *Econometric Reviews*, 28(4), 314–334.
- Sweeting, T. J. (1980). Uniform asymptotic normality of the maximum likelihood estimator. *Annals of Statistics*, 8(6), 1375–1381. Corrections: (1982) *Annals of Statistics* 10, 320.
- Verdejo, H., Awerkin, A., Kliemann, W., & Becker, C. (2019). Modelling uncertainties in electrical power systems with stochastic differential equations. *International Journal of Electrical Power & Energy Systems*, 113, 322–332.
- Zacks, S. (1971). *The theory of statistical inference*. John Wiley & Sons Inc. Wiley Series in Probability and Mathematical Statistics.
- Zhang, S., & Zhang, X. (2013). A least squares estimator for discretely observed Ornstein-Uhlenbeck processes driven by symmetric  $\alpha$ -stable motions. *Annals of the Institute of Statistical Mathematics*, 65(1), 89–103.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.