



# Quasi-likelihood analysis for Student-Lévy regression

Hiroki Masuda<sup>1</sup> · Lorenzo Mercuri<sup>2</sup> · Yuma Uehara<sup>3</sup>

Received: 1 July 2023 / Accepted: 10 September 2024  
© The Author(s) 2024

## Abstract

We consider the quasi-likelihood analysis for a linear regression model driven by a Student- $t$  Lévy process with constant scale and arbitrary degrees of freedom. The model is observed at high frequency over an extending period, under which we can quantify how the sampling frequency affects estimation accuracy. In that setting, joint estimation of trend, scale, and degrees of freedom is a non-trivial problem. The bottleneck is that the Student- $t$  distribution is not closed under convolution, making it difficult to estimate all the parameters fully based on the high-frequency time scale. To efficiently deal with the intricate nature from both theoretical and computational points of view, we propose a two-step quasi-likelihood analysis: first, we make use of the Cauchy quasi-likelihood for estimating the regression-coefficient vector and the scale parameter; then, we construct the sequence of the unit-period cumulative residuals to estimate the remaining degrees of freedom. In particular, using full data in the first step causes a problem stemming from the small-time Cauchy approximation, showing the need for data thinning.

**Keywords** Cauchy quasi-likelihood · High-frequency sampling · Likelihood analysis · Student Lévy process

## 1 Introduction

Suppose that we have a discrete-time sample  $\{(X_{t_j}, Y_{t_j})\}_{j=0}^{\lfloor nT_n \rfloor}$  with  $t_j = j/n$  from the continuous-time (location) regression model

$$Y_t = X_t \cdot \mu + \sigma J_t, \quad t \in [0, T_n], \quad (1.1)$$

Lorenzo Mercuri and Yuma Uehara have contributed equally to this work.

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

Lorenzo Mercuri  
lorenzo.mercuri@unimi.it

Yuma Uehara  
y-uehara@kansai-u.ac.jp

<sup>1</sup> Graduate School of Mathematical Sciences, University of Tokyo, Komaba, Japan

<sup>2</sup> Department of Economics, Management and Quantitative Methods, University of Milan, Milan, Italy

<sup>3</sup> Department of Mathematics, Faculty of Engineering Science, Kansai University, Suita, Japan

where  $X = (X_t)$  is a càdlàg stochastic covariate process in  $\mathbb{R}^q$  satisfying some regularity conditions, the dot denotes the inner product in  $\mathbb{R}^q$ , and  $J = (J_t)$  is a Lévy process such that the unit-time distribution

$$\mathcal{L}(J_1) = t_\nu := t_\nu(0, 1),$$

where  $t_\nu(\mu, \sigma)$  denotes the scaled Student- $t$  distribution with the density

$$f(x; \mu, \sigma, \nu) := \frac{\Gamma(\frac{\nu+1}{2})}{\sigma\sqrt{\pi}\Gamma(\frac{\nu}{2})} \left\{ 1 + \left(\frac{x-\mu}{\sigma}\right)^2 \right\}^{-(\nu+1)/2}. \tag{1.2}$$

Our objective is to estimate the true value  $\theta_0 = (\mu_0, \sigma_0, \nu_0) \in \Theta$  from  $\{(X_{t_j}, Y_{t_j})\}_{j=0}^{[nT_n]}$  when  $T_n \rightarrow \infty$  for  $n \rightarrow \infty$ . We will write  $h = h_n = 1/n$  for the sampling step size. All the processes are defined on an underlying filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$ . The corresponding statistical model is indexed by the unknown parameter  $\theta := (\mu, \sigma, \nu) \in \Theta_\mu \times \Theta_\sigma \times \Theta_\nu = \bar{\Theta}$ . Throughout, we assume that  $\Theta$  is a bounded convex domain in  $\mathbb{R}^{q+2}$  such that its closure  $\bar{\Theta} \subset \mathbb{R}^q \times (0, \infty) \times (0, \infty)$ .

Since the Student- $t$  distribution is not closed under convolution,  $\mathcal{L}(J_h)$  for  $h \neq 1$  is no longer Student- $t$ -distributed. The exact likelihood function is given only through the Fourier inversion, resulting in the rather intractable expression: the characteristic function of  $\mathcal{L}(J_1) = t_\nu$  is given by

$$\varphi_{J_1, \nu}(u) := \frac{2^{1-\nu/2}}{\Gamma(\nu/2)} |u|^{\nu/2} K_{\nu/2}(|u|), \quad u \in \mathbb{R}, \tag{1.3}$$

hence  $\mathcal{L}(J_h)$  admits the Lebesgue density

$$\begin{aligned} x \mapsto & \frac{1}{\pi} \int_0^\infty \cos(ux) \{\varphi_{J_1, \nu}(u)\}^h du \\ &= \frac{1}{\pi} \int_0^\infty \cos(ux) \left( \frac{2^{1-\nu/2}}{\Gamma(\nu/2)} u^{\nu/2} K_{\nu/2}(u) \right)^h du \\ &= \left( \frac{2^{1-\nu/2}}{\Gamma(\nu/2)} \right)^h \frac{1}{\pi} \int_0^\infty \cos(ux) u^{h\nu/2} (K_{\nu/2}(u))^h du, \end{aligned} \tag{1.4}$$

where  $K_\nu(t)$  denotes the modified Bessel function of the second kind ( $\nu \in \mathbb{R}, t > 0$ ):

$$K_\nu(t) = \frac{1}{2} \int_0^\infty s^{\nu-1} \exp\left\{-\frac{t}{2}\left(s + \frac{1}{s}\right)\right\} ds.$$

Numerical integration in (1.4) can be unstable for very small  $h$  and also for large  $|x|$ .

Previously, the thesis (Massing 2019) studied in detail the local asymptotic behavior of the log-likelihood function with numerics for a sample-path generation about Student- $t$  Lévy process observed at high frequency. However, the degrees-of-freedom parameter  $\nu$  was supposed to be known and to be greater than 1 throughout.

In this paper, we propose the explicit two-stage estimation procedure:

1. First, in Sect. 2, we construct an estimator  $(\hat{\mu}_n, \hat{\sigma}_n)$  of  $(\mu, \sigma)$  based on the *Cauchy quasi-likelihood*;
2. Then, in Sect. 3, we make use of the *Student- $t$  quasi-likelihood* for construction of an estimator  $\hat{\nu}_n$  of  $\nu$ , through the “unit-time” residual sequence

$$\hat{\varepsilon}_i := \hat{\sigma}_n^{-1} (Y_i - Y_{i-1} - \hat{\mu}_n \cdot (X_i - X_{i-1})),$$

for  $i = 1, \dots, [T_n]$ , which is expected to be approximately i.i.d.  $t_\nu$ -distributed.

For both estimators, we will prove the asymptotic normality and convergence of their moments (Theorems 2.2, 3.1, and 3.2). Also discussed is how we can weaken the regularity conditions when only interested in deriving the asymptotic normality (Theorems 3.5 and 3.7). The important point is that we can asymptotically efficiently estimate  $(\mu, \sigma)$  with effectively leaving  $\nu$  unknown; under additional conditions, the least-squares estimator is asymptotically normal, but the associated efficiency loss can be significant (see Remark 2.5 below). We refer to Ivanenko et al. (2015) for the details of the corresponding LAN property. In our study, the rate of convergence of  $(\hat{\mu}_n, \hat{\sigma}_n)$  in the first stage depends on how much data we use through the sequence  $(B_n)$  introduced later in (2.2). Under suitable conditions, the foregoing two-stage estimation procedure enables us to estimate  $\nu_0$  as if we directly observe the latent unit-time noise sequence  $(J_i - J_{i-1})_{i=1}^{[T_n]}$ . Since the characteristic function and the Lévy measure of the Student- $t$  distribution are both intractable, we believe that using the Student- $t$  likelihood function based on the unit-time residuals is natural from the computational viewpoint. We will present some numerical experiments in Sect. 4. The proofs are gathered in Sect. 5.

Note that  $X$  may be random and that in some situations  $X$  and  $J$  may be stochastically dependent on each other. Our stepwise procedure with different time scales makes the optimization much easier than the joint estimation. The proposed estimators  $\hat{\sigma}_n$  and  $\hat{\nu}_n$  are asymptotically independent (Theorem 3.2) while the maximum-likelihood estimators in the conventional i.i.d. and time-series settings are asymptotically correlated; see (Harvey 2013, Sect. 2.2) for details.

Below, we briefly list the necessary restrictions in our approach without details.

- First, we need  $T_n \rightarrow \infty$  in the second step for estimating  $\nu_0$ . An asymptotically efficient estimation of  $\nu$  based on the full high-frequency sample  $\{(X_{t_j}, Y_{t_j})\}_{j=0}^{[nT_n]}$  is a non-trivial problem, and we do not know even whether or not the derived rate  $\sqrt{T_n}$  is optimal for estimating  $\nu_0$ . We do not address it in this paper.
- Second, the terminal sampling time  $T_n$  should not grow so quickly if we use whole data in  $[0, T_n]$ ; otherwise, we need thinning data through the sequence  $(B_n)$  satisfying the sampling-balance conditions (2.2) and (3.1) below. This implies that the accumulated Cauchy-approximation error (see Lemma 5.4) for rapidly growing  $T_n$  may crush the clean-cut asymptotic behavior of the estimator  $\hat{\nu}_n$  described in Theorem 3.1 (hence also those in Theorems 3.2, 3.5, and 3.7). We also imposed the additional sampling-design condition (3.6) to establish the asymptotic orthogonality between the proposed estimators of  $(\mu_0, \sigma_0)$  and  $\nu_0$ .

We will denote by  $P_\theta$  the underlying probability measure for  $(J, X, Y)$  associated with  $\theta$ , and by  $E_\theta$  the expectation with respect to  $P_\theta$ . Unless otherwise mentioned, any asymptotics will be taken under  $P := P_{\theta_0}$  for  $n \rightarrow \infty$ . We will denote by  $C$  a universal positive constant which is independent of  $n$  and may vary whenever it appears. The partial differentiation with respect to the variable  $x$  will be denoted by  $\partial_x$ ; we simply use  $\partial$  without specifying a variable if not confusing. The notation  $a_n \lesssim b_n$  for positive sequences  $(a_n)$  and  $(b_n)$  means that  $\limsup_n (a_n/b_n) < \infty$ . We denote by  $I_A = I(A)$  the indicator function of a set  $A$ .

## 2 Cauchy quasi-likelihood analysis

### 2.1 Construction and asymptotic results

By the expression (1.3) and the asymptotic property  $K_\nu(z) \sim \sqrt{\pi/(2z)}e^{-z}$  for  $z \rightarrow \infty$ , the characteristic function of  $h^{-1}J_h$  equals for  $h \rightarrow 0$ ,

$$\begin{aligned} (E_\theta [\exp(iuh^{-1}J_1)])^h &= (1 + o(1))|uv^{-1/2}h^{-1}|^{vh/2} K_{v/2}(|uh^{-1}|)^h \\ &= (1 + o(1)) \left( \sqrt{\frac{\pi h}{2|u|}} \exp(-h^{-1}|u|) (1 + o(1)) \right)^h \\ &\rightarrow \exp(-|u|). \end{aligned} \tag{2.1}$$

This implies the weak convergence  $h^{-1}J_h \xrightarrow{\mathcal{L}} t_1$  (the standard Cauchy distribution) as  $h \rightarrow 0$ , whatever the degrees of freedom  $\nu_0 > 0$  is. It is therefore natural to construct the Cauchy quasi-(log-)likelihood which is defined as if the conditional distribution  $Y_{t_j} | \{Y_{t_{j-1}} = y, X_{t_{j-1}}, X_{t_j}\}$  under  $P_\theta$  equals the Cauchy distribution with location  $y + \mu \cdot (X_{t_j} - X_{t_{j-1}})$  and scale  $h\sigma$ ; note that the conditional likelihood is misspecified unless  $\nu_0 = 1$ . There are both advantages and disadvantages: on the one hand, we can make an inference for  $(\mu, \sigma)$  without knowing the value  $\nu_0$ , but on the other hand, the information of  $\nu_0$  disappeared in the small-time limit.

Using the Cauchy quasi-likelihood is not free: we have to manage the Cauchy-approximation error for  $\mathcal{L}(h^{-1}J_h)$  in the mode of  $L^1(dy)$ -local-limit theorem. Since the approximation error accumulates for  $T_n \rightarrow \infty$ , using the whole sample  $\{(X_{t_j}, Y_{t_j})\}_{j=0}^{[nT_n]}$  may introduce too much error disrupting the Cauchy approximation; this will be seen from the proof of Theorem 2.2 below. Depending on the situation, we will need either to thin the sample or to control the speed of  $T_n \rightarrow \infty$ .

We consider the (possibly) partial observations over the part  $[0, B_n]$  of the entire period  $[0, T_n]$ , where  $(B_n)$  is a positive sequence such that

$$B_n \leq T_n, \quad n^{\epsilon''} \lesssim B_n \lesssim n^{1-\epsilon'} \tag{2.2}$$

for some  $\epsilon', \epsilon'' \in (0, 1)$ . We write the corresponding number of time points used as

$$N_n = [nB_n].$$

Lemma 5.4 below implies that we cannot always take the whole sample (namely  $B_n = T_n$ ) to manage the error of the local-Cauchy approximation. In other words, when one uses the whole sample for estimating  $(\mu, \sigma)$ , the condition (2.2) restricts the speed of  $T_n \rightarrow \infty$  as  $T_n = O(n^{1-\epsilon'})$ .

Write  $a = (\mu, \sigma)$ ,  $a_0 = (\mu_0, \sigma_0)$ , and  $\Delta_j \xi = \xi_{t_j} - \xi_{t_{j-1}}$  for any process  $\xi$ . Let

$$\epsilon_j(a) := \frac{\Delta_j Y - \mu \cdot \Delta_j X}{h\sigma}.$$

Let  $\phi_1(y) := \pi^{-1} (1 + y^2)^{-2}$ , the standard Cauchy density. Then, we introduce the Cauchy quasi-(log-)likelihood conditional on  $X$ :

$$\begin{aligned} \mathbb{H}_{1,n}(a) &:= \sum_{j=1}^{N_n} \log \left\{ \frac{1}{h\sigma} \phi_1 \left( \frac{\Delta_j Y - \mu \cdot \Delta_j X}{h\sigma} \right) \right\} \\ &= C_n - \sum_{j=1}^{N_n} \{ \log \sigma + \log (1 + \epsilon_j(a)^2) \}, \end{aligned}$$

where the term  $C_n$  does not depend on  $a$ . We define the Cauchy quasi-maximum likelihood estimator (CQMLE for short) by any element

$$\hat{a}_n := (\hat{\mu}_n, \hat{\sigma}_n) \in \operatorname{argmax}_{a \in \Theta_\mu \times \Theta_\sigma} \mathbb{H}_{1,n}(a).$$

We introduce the following regularity conditions on  $X$ .

**Assumption 2.1** For  $X$ , we can associate an  $(\mathcal{F}_t)$ -adapted process  $t \mapsto X'_t$  in  $\mathbb{R}^q$  having càdlàg sample paths, for which the following conditions hold for every  $K > 0$ :

1.  $\sup_{t \geq 0} E[|X'_t|^K] < \infty$ .
2. There exists a constant  $c_X > 0$  such that both

$$\sup_n \max_{j=1, \dots, [nT_n]} E \left[ \left| \frac{1}{h^{c_X}} \left\{ \frac{1}{h} (\Delta_j X - hX'_{t_{j-1}}) \right\} \right|^K \right] < \infty$$

and

$$\sup_n \max_{j=1, \dots, [nT_n]} \sup_{t_{j-1} \leq t \leq t_j} E \left[ \left| \frac{1}{h^{c_X}} (X'_t - X'_{t_{j-1}}) \right|^K \right] < \infty$$

hold. Moreover,

$$\max_{j=1, \dots, [nT_n]} E \left[ \left| \sqrt{N_n} E \left[ \left\{ \frac{1}{h} (\Delta_j X - hX'_{t_{j-1}}) \right\} \frac{\partial \phi_1}{\phi_1} \left( \frac{\Delta_j J}{h} \right) \middle| \mathcal{F}_{t_{j-1}} \right] \right|^K \right] = o(1).$$

3. There exist a constant  $c'_X > 0$  and a probability measure  $\pi_0(dx)$  on some  $q'$ -dimensional Borel space  $(\mathbb{R}^{q'}, \mathcal{B}^{q'})$  such that, for every real-valued  $\mathcal{C}^1(\mathbb{R}^q)$ -function  $f(x)$  satisfying  $\max_{k \in \{0, 1\}} |\partial_x^k f(x)| \lesssim 1 + |x|^C$ , we can associate a measurable function  $\psi_f(z)$  such that

$$\sup_n E \left[ \left| N_n^{c'_X} \left( \frac{1}{N_n} \sum_{j=1}^{N_n} f(X'_{t_{j-1}}) - \int \psi_f(z) \pi_0(dz) \right) \right|^K \right] < \infty \tag{2.3}$$

for every  $K > 0$ .

4. The  $q \times q$  matrix  $S_0 := \int \psi_f(z) \pi_0(dz)$  for  $f(x') = x'^{\otimes 2}$  is symmetric, positive-definite, and finite.

Assumption 2.1 is designed not only for convergence in distribution of our estimators but also for ensuring the convergence of their moments: in the proofs, items 1 and 2 are used for a series of moment estimates, item 3 (together with the Sobolev inequality) is required to handle the uniform-in-parameter moment estimates and to identify the limit of the law of large numbers under the ergodicity, and item 4 will ensure the positive definiteness of the asymptotic covariance matrix of the estimator. It should be noted that the moment conditions

in Assumption 2.1 can be too much to ask if one is only interested in deriving the asymptotic normality; see Theorems 3.5 and 3.7.

The last convergence in Assumption 2.1.2 is trivial if  $X'$  and  $J$  are independent so that the conditional expectation involved therein vanishes a.s. Although we did not explicitly mention, one may think of  $X_t = \int_0^t X'_s ds$  and where the integrand  $X'$  admits a unique invariant distribution  $\pi_0$  for which  $X'_t \xrightarrow{\mathcal{L}} \pi_0$  as  $t \rightarrow \infty$  and  $S_0 = \int f(x', a)\pi_0(dx')$  (in this case,  $q = q'$  and  $\psi_f = f$ ); then, since  $\int f(x')\pi_0(dx') \leq \sup_{t \geq 0} E[f(X'_t)]$  for any continuous nonnegative  $f$ , we have  $\int |x'|^K \pi_0(dx') < \infty$  for each  $K > 0$  hence in particular  $S_0$  is finite. See Sect. 2.2 for more related discussions; as we will see there, the dimension  $q'$  in Assumption 2.1.3 may differ from  $q$ .

Let

$$\begin{aligned} \Delta_{a,n} &:= \frac{1}{\sqrt{N_n}} \partial_a \mathbb{H}_{1,n}(a_0), \\ \Gamma_{a,0} &:= \text{diag} \left( \frac{1}{2\sigma_0^2} S_0, \frac{1}{2\sigma_0^2} \right). \end{aligned} \tag{2.4}$$

Let

$$\zeta(y, x'; a) := \log \left\{ \frac{\sigma_0}{\sigma} \phi_1 \left( \frac{\sigma_0}{\sigma} \left( y - \frac{1}{\sigma_0} (\mu - \mu_0) \cdot x' \right) \right) \right\} - \log \phi_1(y), \tag{2.5}$$

and

$$f'(x'; a) := \int \zeta(y, x'; a) \phi_1(y) dy,$$

which is smooth in  $a \in \overline{\Theta_\mu \times \Theta_\sigma}$ . Then, under Assumption 2.1.3, for each  $a$  we can associate measurable functions  $\psi_{f'}(z; a)$  and  $\psi_{\partial_a f'}(z; a)$  and a probability measure  $\pi_0(dz)$ , for which

$$\frac{1}{N_n} \sum_{j=1}^{N_n} \partial_a^k f'(X'_{t_{j-1}}; a) \xrightarrow{p} \int \psi_{\partial_a^k f'(\cdot; a)}(z; a) \pi_0(dz), \quad k = 0, 1.$$

We write

$$\mathbb{Y}_1(a) = \int \psi_{f'(\cdot; a)}(z; a) \pi_0(dz). \tag{2.6}$$

The main result of this section is the following.

**Theorem 2.2** *Suppose that (2.2) and Assumption 2.1 hold, that  $a \mapsto \mathbb{Y}_1(a)$  is  $\mathcal{C}^1$ -class with  $\partial_a \mathbb{Y}_1(a) = \int \psi_{\partial_a f'(\cdot; a)}(z; a) \pi_0(dz)$ , and that there exists a constant  $c_0 > 0$  for which*

$$\mathbb{Y}_1(a) \leq -c_0 |a - a_0|^2, \quad a \in \Theta_\mu \times \Theta_\sigma. \tag{2.7}$$

Then, we have the asymptotic normality

$$\hat{u}_{a,n} := \sqrt{N_n} (\hat{a}_n - a_0) = \Gamma_{a,0}^{-1} \Delta_{a,n} + o_p(1) \xrightarrow{\mathcal{L}} N \left( 0, \Gamma_{a,0}^{-1} \right) \tag{2.8}$$

and the polynomial-type tail-probability estimate:

$$\forall L > 0, \quad \sup_{r>0} \sup_n P [|\hat{u}_{a,n}| > r] r^L < \infty. \tag{2.9}$$

If  $q' = q$ ,  $\pi_0$  is the invariant distribution of  $X'$ , and if Assumption 2.1.3 and 2.1.4 hold with  $\psi_f = f$ , then

$$\mathbb{Y}_1(a) = \int \int \zeta(y, x'; a)\phi_1(y)dy \pi_0(dx').$$

In this case, the identifiability condition (2.7) is automatic. Indeed, the well-known property of the Kullback-Leibler divergence, we have  $\mathbb{Y}_1(a) \leq 0$  with the equality holding if and only if  $a = a_0$ . Moreover, the above  $\mathbb{Y}_1(a)$  is smooth,  $\partial_a \mathbb{Y}_1(a_0) = 0$ , and  $-\partial_a^2 \mathbb{Y}_1(a_0)$  is positive definite under the present assumptions (including the positive definiteness of  $S_0 = \int x'^{\otimes 2} \pi_0(dx')$ ). These observations conclude (2.7).

It is easy to construct a consistent estimator of  $S_0$ : see (3.7) in Sect. 3.2. We have the variance-stabilizing transform for  $\sigma$ :  $\sqrt{N_n/2} (\log \hat{\sigma}_n - \log \sigma_0) \xrightarrow{\mathcal{L}} N_1(0, 1)$ .

**Remark 2.3** Making use of sample in over  $[0, B_n] \subset [0, T_n]$  is not essential. It is possible to use, for example, only sample over  $\bigcup_{k \geq 1} [2k - 2, 2k - 1)$ . Also, we could consider the case where  $B_n \equiv B$  for some fixed  $B > 0$ , meaning that for estimating  $(\mu, \sigma)$  we only use a sample over the fixed period  $[0, B]$ . In that case, if  $X$  is truly random, under suitable conditions on the underlying filtration  $(\mathcal{F}_t)$ , the asymptotic distribution of  $\sqrt{n} (\hat{a}_n - a_0)$  is mixed-normal with the random asymptotic covariance depending on a sample path  $(X_t)_{t \in [0, B]}$ . See also Remark 3.3, Clément and Gloter (2019, 2020), and Masuda (2019) for related results.

**Remark 2.4** Since high-frequency data over each fixed period is enough to consistently estimate  $(\mu, \sigma)$ , one may think of a parameter-varying (randomized parameter) setting: if  $\mu$  and  $\sigma$  may vary along  $i$ , say  $i \mapsto (\mu_i, \sigma_i)$  for the  $i$ th period  $[i - 1, i)$  and if they are all to be estimated from  $\{(X_{t_j}, Y_{t_j})\}_{j \geq 1: jh \in [i-1, i)}$ , then the model setup invokes the classical Neyman-Scott problem (Neyman and Scott 1948); if  $\{(\mu_i, \sigma_i)\}_{i \geq 1}$  is an unobserved i.i.d. sequence of random vectors, then the model becomes a (partially) random-effect one. The latter would be of interest in the context of the population approach.

**Remark 2.5** (Least-squares estimator) Let  $\nu_0 > 2$ . We can rewrite (1.1) as

$$Y_t = X_t \cdot \mu + \sigma' Z_t$$

with  $\sigma' := \sigma/\sqrt{\nu - 2}$  and the Lévy process  $Z_t := \sqrt{\nu - 2} J_t$  such that  $E[Z_1] = 0$  and  $\text{var}[Z_1] = 1$ . In this case, under suitable conditions, we can apply the least-squares method for estimating  $(\mu, \sigma')$  and deduce its asymptotic normality at rate  $\sqrt{T_n}$ .

**Remark 2.6** The claims in Theorem 2.2 hold for any local-Cauchy  $J$  satisfying the local limit theorem Lemma 5.4. In particular, the model (1.1) with  $J$  replaced by the generalized hyperbolic Lévy process (except for the variance gamma one) could be handled analogously; see (Masuda 2019, Section 2.1.2) for related information.

## 2.2 Covariate process

We set Assumption 2.1 without imposing a concrete structure of  $X$ . To provide a set of more handy conditions on  $X$  and  $X'$ , it is necessary to impose more specific structures on them. Below, we will present such an example in detail.

### 2.2.1 Setup

Let us briefly discuss how to verify Assumption 2.1 for  $X_t = (X_{1,t}, X_{2,t}) \in \mathbb{R}^{q_1} \times \mathbb{R}^{q_2}$  ( $q = q_1 + q_2$ ) of the form

$$t \mapsto X_{k,t} = \int_0^t X'_{k,s} ds, \quad k = 1, 2,$$

where  $t \mapsto (X'_{1,t}, X'_{2,t})$  is Riemann-integrable process independent of the Student- $t$  Lévy process  $J$ . Further, we assume the following conditions.

- $X'_1$  is a non-random periodic  $C^1$ -function with possibly unknown period  $\tau > 0$ , and satisfies that

$$\sup_{t \geq 0} (|X_{1,t}| + |\partial_t X_{1,t}|) < \infty.$$

For example,  $X'_1$  can be the derivative of

$$X_{1,t} = (\cos(\tau_1 t), \sin(\tau_1 t), \dots, \cos(\tau_M t), \sin(\tau_M t))$$

with  $\tau_1, \dots, \tau_M$  being unknown different rational numbers ( $q_1 = 2M$ ); we may get rid of some of the components from the beginning, such as  $X_{1,t} = (\cos(5t), \sin(t))$ .

- $X'_2$  is a  $q_2$ -dimensional diffusion process:

$$dX'_{2,t} = A(X'_{2,t}) dw_t + B(X'_{2,t}) dt, \quad (2.10)$$

where  $w$  is a standard Wiener process (independent of  $J$ ). The unknown coefficients  $A$  and  $B$  are globally Lipschitz so that (2.10) admits a unique strong solution. Moreover,  $X'_2$  is exponentially ergodic in the sense that there exist some  $\kappa > 0$  and an invariant measure  $\pi_0$  on  $(\mathbb{R}^{q_2}, \mathcal{B}^{q_2})$  for which

$$\begin{aligned} & \|P_t(x, \cdot) - \pi_0(dz)\|_g \\ & := \sup_{t \geq 0} \sup_{f: |f| \leq g} \left| \int f(z) P_t(x, dz) - \int f(z) \pi_0(dz) \right| \lesssim g(x) e^{-\kappa t} \end{aligned} \quad (2.11)$$

for any  $g$  of at most polynomial growth.

Our model may be used for analyzing time series data observed at high frequency exhibiting a seasonality (Proietti and Pedregal 2022). Concerned with the exponential ergodicity (2.11) of  $X'_2$ , we refer to Kulik (2018) and the references therein for several easy-to-verify sufficient conditions.

With the above setup, we will verify Assumption 2.1 along with imposing additional conditions.

### 2.2.2 Verification

First, for Assumption 2.1.1 and 2.1.2, we can look at  $X_1$  and  $X_2$  separately. The conditions obviously hold for  $X_1$ . As for  $X_2$  of (2.10), Assumption 2.1.1 can be verified through the easy-to-apply Lyapunov-function criteria; we do not list them here, but just refer to Masuda (2007, Theorem 2.2), Kulik (2009, Lemma 3.3), and also Masuda (2013, Section 5). Let



us suppose Assumption 2.1.1 in what follows. Then, we can verify all the conditions in Assumption 2.1.2 with  $c_X = 1/2$  as follows. Since the sequence  $(j = 1, \dots, n)$

$$\frac{1}{\sqrt{h}} \left( X'_{2,t} - X'_{2,t_{j-1}} \right) = \frac{1}{\sqrt{h}} \int_{t_{j-1}}^t A(X'_{2,s}) dw_s + \sqrt{h} \frac{1}{h} \int_{t_{j-1}}^t B(X'_{2,s}) ds$$

is  $L^K(P)$ -bounded for any  $K > 0$ , so is

$$\frac{1}{h\sqrt{h}} \left( \Delta_j X_2 - hX'_{2,t_{j-1}} \right) = \frac{1}{h} \int_{t_{j-1}}^{t_j} \frac{1}{\sqrt{h}} \left( X'_{2,s} - X'_{2,t_{j-1}} \right) ds,$$

followed by the first two conditions in Assumption 2.1.2 with  $c_X = 1/2$ . As for the last one, we note the decomposition

$$\begin{aligned} & \frac{1}{h} \left( \Delta_j X_2 - hX'_{2,t_{j-1}} \right) \\ &= A(X'_{2,t_{j-1}}) \frac{1}{h} \int_{t_{j-1}}^{t_j} (w_s - w_{t_{j-1}}) ds + \left\{ \frac{1}{h} \int_{t_{j-1}}^{t_j} \int_{t_{j-1}}^s B(X'_{2,u}) du ds \right. \\ & \quad \left. + \frac{1}{h} \int_{t_{j-1}}^{t_j} \int_{t_{j-1}}^s (A(X'_{2,u}) - A(X'_{2,t_{j-1}})) dw_u ds \right\} \\ &=: T_{1,h} + T_{2,h}. \end{aligned}$$

Since  $\sup_y |((\partial\phi_1)/\phi_1)(y)| < \infty$  and  $E [((\partial\phi_1)/\phi_1)(h^{-1}\Delta_j J) | \mathcal{F}_{t_{j-1}}] = 0$  a.s., we have

$$\begin{aligned} & \left| \sqrt{N_n} E \left[ \left\{ \frac{1}{h} \left( \Delta_j X - hX'_{t_{j-1}} \right) \right\} \frac{\partial\phi_1}{\phi_1} \left( \frac{\Delta_j J}{h} \right) \middle| \mathcal{F}_{t_{j-1}} \right] \right| \\ & \lesssim h\sqrt{N_n} E \left[ \frac{1}{h} |T_{2,h}| \middle| \mathcal{F}_{t_{j-1}} \right] \\ & \quad + \left| \sqrt{N_n} A(X'_{2,t_{j-1}}) E \left[ \frac{1}{h} \int_{t_{j-1}}^{t_j} (w_s - w_{t_{j-1}}) ds \frac{\partial\phi_1}{\phi_1} \left( \frac{\Delta_j J}{h} \right) \middle| \mathcal{F}_{t_{j-1}} \right] \right|. \end{aligned}$$

The second term in the upper bound a.s. vanishes since  $w$  and  $J$  are mutually independent. Under the linear-growth property of the coefficients, by the standard estimates as before the sequence “ $E [h^{-1}|T_{2,h}| | \mathcal{F}_{t_{j-1}}]$ ” is  $L^K(P)$ -bounded for any  $K > 0$ . Since  $h\sqrt{N_n} \lesssim n^{-\epsilon'/2} \rightarrow 0$  by (2.2), the last convergence in Assumption 2.1.2 follows. Thus, we have verified Assumption 2.1.1 and 2.1.2.

Turning to Assumption 2.1.3, we take  $c'_X > 0$  so small that  $N_n^{c'_X} \lesssim h^{-c_X} = h^{-1/2}$ . To handle the periodic nature of  $X'_1$ , we note that (2.3) is derived by showing that

$$E \left[ \left| N_n^{c'_X} \left( \frac{1}{hN_n} \int_0^{hN_n} f(X'_t) dt - \int \psi_f(z) \pi_0(dz) \right) \right|^K \right] \lesssim 1$$

for suitable  $\psi_f(z)$ , since Jensen’s and Cauchy-Schwarz inequalities give

$$E \left[ \left| N_n^{c'_X} \left( \frac{1}{hN_n} \int_0^{hN_n} f(X'_t) dt - \frac{1}{N_n} \sum_{j=1}^{N_n} f(X'_{t_{j-1}}) \right) \right|^K \right]$$

$$\begin{aligned}
 &\leq E \left[ \left| N_n^{c'_X} \left( \frac{1}{N_n} \sum_{j=1}^{N_n} \frac{1}{h} \int_{(j-1)/n}^{j/n} (f(X'_t) - f(X'_{t_{j-1}})) dt \right) \right|^K \right] \\
 &\lesssim \frac{1}{N_n} \sum_{j=1}^{N_n} \frac{1}{h} \int_{(j-1)/n}^{j/n} E \left[ (1 + |X'_t|^C + |X'_{t_{j-1}}|^C) (N_n^{c'_X} |X'_t - X'_{t_{j-1}}|)^K \right] dt \\
 &\lesssim \frac{1}{N_n} \sum_{j=1}^{N_n} \frac{1}{h} \int_{t_{j-1}}^{t_j} \sup_t E [1 + |X'_t|^C]^{1/2} E \left[ |N_n^{c'_X} (X'_t - X'_{t_{j-1}})|^{2K} \right]^{1/2} dt \\
 &\lesssim \max_{j=1, \dots, \lfloor nT_n \rfloor} \sup_{t_{j-1} \leq t \leq t_j} E \left[ |h^{-1/2} (X'_t - X'_{t_{j-1}})|^{2K} \right]^{1/2} \lesssim 1.
 \end{aligned}$$

The last estimate is due to the already verified Assumption 2.1.2. Let  $m_n = \max\{i \geq 1 : i\tau \leq hN_n\} = \lfloor hN_n/\tau \rfloor$ . Then,  $m_n \rightarrow \infty$  as  $m_n \gtrsim n^{\epsilon''}$  by (2.2), and we have

$$\frac{1}{hN_n} \int_0^{hN_n} f(X'_t) dt = \frac{1}{hN_n} \int_0^{\tau m_n} f(X'_t) dt + \frac{1}{hN_n} \int_{\tau m_n}^{hN_n} f(X'_t) dt.$$

The second term on the right-hand side can be bounded by a constant multiple of  $(hN_n)^{-1} (1 + |X'_t|^C) \lesssim n^{-\epsilon''} (1 + |X'_t|^C)$ . Further letting  $c'_X > 0$  sufficiently small so that  $n^{-\epsilon''} N_n^{c'_X} \lesssim 1$  if necessary, we only need to look at the first term in the last display, say  $V_n$ . By the periodicity of  $X'_1$ , we have

$$\begin{aligned}
 V_n &= \frac{1}{hN_n} \sum_{m=1}^{m_n} \int_{(m-1)\tau}^{m\tau} f(X'_t) dt \\
 &= \frac{\tau m_n}{hN_n} \frac{1}{m_n} \sum_{m=1}^{m_n} \frac{1}{\tau} \int_0^\tau f(X'_{1, (m-1)\tau+t}, X'_{2, (m-1)\tau+t}) dt \\
 &= \frac{\tau m_n}{hN_n} \frac{1}{m_n} \sum_{m=1}^{m_n} \frac{1}{\tau} \int_0^\tau f(X'_{1,t}, X'_{2, (m-1)\tau+t}) dt \\
 &=: \frac{\tau m_n}{hN_n} \frac{1}{m_n} \sum_{m=1}^{m_n} G_{f,m}, \quad \text{say.}
 \end{aligned}$$

Now we set

$$\psi_f(z) = \frac{1}{\tau} \int_0^\tau f(X'_{1,t}, z) dt.$$

Then,

$$\begin{aligned}
 &V_n - \int \psi_f(z) \pi_0(dz) \\
 &= \left( \frac{\tau m_n}{hN_n} - 1 \right) \frac{1}{m_n} \sum_{m=1}^{m_n} G_{f,m} + \frac{1}{m_n} \sum_{m=1}^{m_n} (G_{f,m} - E[G_{f,m}]) \\
 &\quad + \frac{1}{m_n} \sum_{m=1}^{m_n} \left( E[G_{f,m}] - \int \psi_f(z) \pi_0(dz) \right) \\
 &=: \bar{V}_{1,n} + \bar{V}_{2,n} + \bar{V}_{3,n}, \quad \text{say.}
 \end{aligned}$$

Fix any  $K \geq 2$ . Since  $\sup_{m \geq 1} E [|G_{f,m}|^K] < \infty$ ,  $|\frac{\tau m}{hN_n} - 1| = O((hN_n)^{-1}) = O(n^{-\epsilon''})$  and  $n^{-\epsilon''} N_n^{c'_X} \lesssim 1$ , we get  $\sup_n E [|N_n^{c'_X} \bar{V}_{1,n}|^K] < \infty$ . We can handle the remaining two terms by the mixing property: the condition (2.11) implies that  $X'_2$  is exponentially  $\beta$ -mixing (see (Masuda 2007, Lemma 3.9)), hence also exponentially  $\alpha$ -mixing. Since  $G_{f,m}$  is  $\sigma(X'_{2,s} : s \in [(m-1)\tau, m\tau])$ -measurable, there exists a constant  $c_l = c_l(\tau) > 0$  ( $l = 1, 2$ ) for which

$$\alpha(k) := \sup_{i \geq 1} \sup_{\substack{A \in \sigma(G_{f,l}: l \leq i), \\ B \in \sigma(G_{f,l}: l \geq i+k)}} |P[A \cap B] - P[A]P[B]| \leq c_2 \exp(-c_1 k), \quad k \in \mathbb{N}.$$

Then, a direct application of Yoshida (2011, Lemma 4) yields  $\sup_n E [|N_n^{c'_X} \bar{V}_{2,n}|^K] < \infty$ .

The remaining  $\bar{V}_{3,n}$  can be treated similarly to the proof of Masuda (2013, Lemma 4.3) as follows (see how to estimate  $\Lambda''_n(f; \theta)$  therein): letting  $g(z) := \sup_{|x'_1| \leq \|X'_{1,\cdot}\|_\infty} |f(x'_1, z)|$  (hence  $g(z) \lesssim 1 + |z|^C$ ), we obtain from (2.11) that

$$\begin{aligned} & \left| E [G_{f,m}] - \int \psi_f(z) \pi_0(dz) \right| \\ &= \left| \frac{1}{\tau} \int_0^\tau \iint f(X'_{1,t}, z) \{P_{(m-1)\tau+t}(x, dz) - \pi_0(dz)\} P^{X'_{2,0}}(dx) dt \right| \\ &\leq \frac{1}{\tau} \int_0^\tau \int \|P_{(m-1)\tau+t}(x, \cdot) - \pi_0(\cdot)\|_g P^{X'_{2,0}}(dx) dt \\ &\lesssim \int g(x) P^{X'_{2,0}}(dx) \int_0^\tau \exp(-\kappa((m-1)\tau + t)) dt \lesssim \exp(-\kappa \tau m), \end{aligned}$$

where  $P^{X'_{2,0}}(\cdot)$  denotes the initial distribution of  $X'_2$ . It follows that  $|N_n^{c'_X} \bar{V}_{3,n}| \lesssim N_n^{c'_X} M_n^{-1} \lesssim n^{-\epsilon''} N_n^{c'_X} \lesssim 1$ , concluding Assumption 2.1.3 (with  $q' = q_2$ ).

Finally, it does not seem easy to give a general sufficient condition for the positive definiteness of  $S_0$  in Assumption 2.1.4, which is related to the positive definiteness at the true value  $\theta_0$  of the Fisher-information matrix.

### 3 Student quasi-likelihood analysis

Taking over the setting in Sect. 2, we now turn to the estimation of the degrees of freedom  $\nu$  from  $\{(X_{t_j}, Y_{t_j})\}_{j=0}^{[nT_n]}$ . Suppose that Assumption 2.1 and (2.2) hold, so that we have (2.9) by Theorem 2.2. Throughout this section, we assume that  $N_n$  is sufficiently large relative to  $T_n$ :

$$\frac{T_n}{N_n} = \frac{T_n}{[nB_n]} \rightarrow 0. \tag{3.1}$$

#### 3.1 Construction and asymptotics

Define the *unit-time* residual sequence

$$\hat{\epsilon}_i := \hat{\sigma}_n^{-1} (Y_i - Y_{i-1} - \hat{\mu}_n \cdot (X_i - X_{i-1})),$$

for  $i = 1, \dots, [T_n]$ . Let

$$\varepsilon_i := J_i - J_{i-1}$$

so that  $\varepsilon_1, \varepsilon_2, \dots \sim$  i.i.d.  $t_\nu$ . We will estimate  $\nu$  based on the maximum-likelihood function as if  $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_{[T_n]}$  are observed  $t_\nu$ -i.i.d. samples: we consider the explicit Student quasi-likelihood function

$$\mathbb{H}_{2,n}(\nu) = \sum_{i=1}^{[T_n]} \rho(\hat{\varepsilon}_i; \nu),$$

where, with the notation (1.2),

$$\rho(\varepsilon; \nu) := \log f(\varepsilon; 0, 1, \nu) = \log \left( \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi}\Gamma(\frac{\nu}{2})} (1 + \varepsilon^2)^{-(\nu+1)/2} \right).$$

Then, we define the Student- $t$  quasi-likelihood estimator ( $t$ -QMLE for short) of  $\nu$  by any element

$$\hat{\nu}_n \in \operatorname{argmax}_{\nu \in \bar{\Theta}_\nu} \mathbb{H}_{2,n}(\nu).$$

We have

$$\mathbb{H}_{2,n}(\nu) = \sum_{i=1}^{[T_n]} \left( -\frac{1}{2} \log \pi + \log \Gamma\left(\frac{\nu+1}{2}\right) - \log \Gamma\left(\frac{\nu}{2}\right) - \frac{\nu+1}{2} \log(1 + \hat{\varepsilon}_i^2) \right).$$

Let  $\psi(x) := \partial \log \Gamma(x)$ , the digamma function, and then  $\psi_1 := \partial \psi$ , the trigamma function. By the integral representation  $\partial^m \psi(x) = (-1)^{m+1} \int_0^\infty s^m e^{-xs} (1 - e^{-s})^{-1} ds$  for  $m \in \mathbb{N}$  (see (Abramowitz and Stegun 1992, 6.4.1)), hence  $\partial \psi_1(x) < 0$  for  $x > 0$ . From this fact and the last expression for  $\mathbb{H}_{2,n}(\nu)$ , we obtain

$$-\partial_\nu^2 \mathbb{H}_{2,n}(\nu) = \frac{[T_n]}{4} \left( \psi_1\left(\frac{\nu}{2}\right) - \psi_1\left(\frac{\nu+1}{2}\right) \right) > 0 \tag{3.2}$$

for any  $\nu > 0$ , hence  $-\mathbb{H}_{2,n}$  is a.s. convex on  $(0, \infty)$ .

Let

$$\begin{aligned} \Delta_{\nu,n} &:= \frac{1}{\sqrt{[T_n]}} \sum_{i=1}^{[T_n]} \partial_\nu \rho(\varepsilon_i; \nu_0), \\ \Gamma_{\nu,0} &:= \frac{1}{4} \left( \psi_1\left(\frac{\nu_0}{2}\right) - \psi_1\left(\frac{\nu_0+1}{2}\right) \right). \end{aligned} \tag{3.3}$$

We have  $\Gamma_{\nu,0} > 0$  by (3.2). The next result shows that we can estimate  $\nu_0$  directly as if we observe  $(\varepsilon_i)$ .

**Theorem 3.1** *Under the assumptions in Theorem 2.2 and (3.1), we have the polynomial-type tail-probability estimate:*

$$\forall L > 0, \quad \sup_{r>0} \sup_n P[|\hat{u}_{\nu,n}| > r] r^L < \infty, \tag{3.4}$$

and the asymptotic normality:

$$\hat{u}_{\nu,n} := \sqrt{[T_n]}(\hat{\nu}_n - \nu_0) = \Gamma_{\nu,0}^{-1} \Delta_{\nu,n} + o_p(1) \xrightarrow{\mathcal{L}} N\left(0, \Gamma_{\nu,0}^{-1}\right). \tag{3.5}$$

The maximum-likelihood estimation of  $\nu_0$  can become more unstable for a larger value of  $\nu_0$ . The Fisher information  $\Gamma_{\nu,0}$  quickly decreases to 0 as  $\nu_0$  increases: the asymptotic variance  $\Gamma_{\nu,0}^{-1}$  equals about 1.22, 5.63, 13.8, 25.9, and 41.9 for  $\nu_0 = 1, 2, 3, 4,$  and  $5,$  respectively. The damping speed becomes even faster in the case of conventional parametrization. See, for example, Harvey (2013, Section 2.2) and the references therein.

### 3.2 Joint asymptotic normality

Having Theorems 2.2 and 3.1 in hand, we consider the question of the joint asymptotic normality of  $\hat{a}_n$  and  $\hat{\nu}_n$ .

Recall that we are given the underlying filtration  $(\mathcal{F}_t)_{t \geq 0}$ . From the proofs of Theorems 2.2 and 3.1, we can write  $(\hat{u}_{a,n}, \hat{u}_{\nu,n}) = M_n + o_p(1)$  with

$$M_n = \frac{1}{\sqrt{T_n}} \sum_{i=1}^{[T_n]} G_i = \frac{1}{\sqrt{T_n}} \sum_{i=1}^{[T_n]} (G_{a,i}, G_{\nu,i})$$

for a martingale difference  $G_{\nu,i} = \partial_{\nu} \rho(\varepsilon_i; \nu_0) \in \mathbb{R}$  with respect to  $(\mathcal{F}_i)_{i \in \mathbb{Z}_+}$  and

$$G_{a,i} := \sqrt{\frac{T_n}{N_n}} I(i \leq [B_n]) \sum_{j \in A_i} \psi_{a,ij},$$

where  $A_i := \{k \in \mathbb{N} : (i - 1)n < k \leq in\}$  and  $\psi_{a,ij} = \psi_{a,ij}(a_0) := -\partial_a \{\log \sigma + \log(1 + \epsilon_j(a)^2)\} |_{a=a_0}$ . It will be seen in the proof of Lemma 5.3 that  $T_n^{-1/2} \sum_{i=1}^{[T_n]} E[G_{a,i} | \mathcal{F}_{t_{j-1}}] = o_p(1)$ . Hence,  $M_n$  is a partial sum of the approximate martingale difference array with respect to  $(\mathcal{F}_i)_{i \in \mathbb{Z}_+}$ .

The Lyapunov condition holds for  $M_n$ . To deduce the first-order asymptotic behavior of  $M_n$ , we need to look at the convergence in probability of the matrix-valued quadratic characteristic:

$$[M]_n := \frac{1}{T_n} \sum_{i=1}^{[T_n]} \begin{pmatrix} G_{a,i}^{\otimes 2} & G_{a,i} G_{\nu,i} \\ \text{sym.} & G_{\nu,i}^2 \end{pmatrix},$$

which explicitly depends on  $\theta_0 = (a_0, \nu_0)$ ; among others, we refer to Shiryaev (1980, Chapter VII.8) for the related basic facts. By Theorems 2.2 and 3.1, it remains to manage the cross-covariation part of  $[M]_n$ .

To that end, we further assume

$$\frac{B_n}{T_n} \rightarrow 0. \tag{3.6}$$

By Lemmas 5.2 and 5.6 in Sect. 5.1, for any  $K > 0,$

$$\max_{1 \leq i \leq [T_n]} \max_{1 \leq j \leq [nT_n]} E \left[ |\psi_{a,ij}|^K + |G_{\nu,i}|^K \right] + \max_{1 \leq i \leq [T_n]} E \left[ \left| \frac{1}{\sqrt{n}} \sum_{j \in A_i} \psi_{a,ij} \right|^K \right] = O(1).$$

This estimate leads to

$$\frac{1}{T_n} \sum_{i=1}^{[T_n]} G_{\nu,i} G_{a,i} = \sqrt{\frac{B_n}{T_n}} \frac{1}{B_n} \sum_{i=1}^{[B_n]} G_{\nu,i} \sqrt{\frac{1}{n}} \sum_{j \in A_i} \psi_{a,ij} = O_p \left( \sqrt{\frac{B_n}{T_n}} \right) = o_p(1).$$

Thus, the additional condition (3.6) approximately quantifies how much data we should discard to make the estimators  $\hat{a}_n$  and  $\hat{v}_n$  independent.

Under Assumption 2.1,

$$\begin{aligned}\hat{S}_n &:= \frac{1}{N_n} \sum_{j=1}^{N_n} \left( \frac{1}{h} \Delta_j X \right)^{\otimes 2} \\ &= \frac{1}{N_n} \sum_{j=1}^{N_n} \left\{ \left( \frac{1}{h} \Delta_j X - X'_{t_{j-1}} \right) + X'_{t_{j-1}} \right\}^{\otimes 2} \\ &= \frac{1}{N_n} \sum_{j=1}^{N_n} \left( X'_{t_{j-1}} \right)^{\otimes 2} + o_p(1) \xrightarrow{P} S_0, \quad n \rightarrow \infty,\end{aligned}\quad (3.7)$$

Recalling the notation (2.4) and (3.3), we introduce

$$\begin{aligned}\hat{\Gamma}_{a,n} &:= \text{diag} \left( \frac{1}{2\hat{\sigma}_n^2} \hat{S}_n, \frac{1}{2\hat{\sigma}_n^2} \right), \\ \hat{\Gamma}_{v,n} &:= \frac{1}{4} \left( \psi_1 \left( \frac{\hat{v}_n}{2} \right) - \psi_1 \left( \frac{\hat{v}_n + 1}{2} \right) \right).\end{aligned}$$

Obviously we have  $(\hat{\Gamma}_{a,n}, \hat{\Gamma}_{v,n}) \xrightarrow{P} (\Gamma_{a,0}, \Gamma_{v,0})$ , which combined with (2.8) and (3.5) leads to the joint asymptotic normality:

**Theorem 3.2** *Under the assumptions in Theorem 2.2, (3.1), and (3.6), we have*

$$\left( \hat{\Gamma}_{a,n}^{1/2} \hat{u}_{a,n}, \hat{\Gamma}_{v,n}^{1/2} \hat{u}_{v,n} \right) = \text{diag} \left( \hat{\Gamma}_{a,n}^{1/2}, \hat{\Gamma}_{v,n}^{1/2} \right) M_n + o_p(1) \xrightarrow{\mathcal{L}} N_{q+2} \left( 0, I_{q+2} \right). \quad (3.8)$$

**Remark 3.3** (Asymptotic mixed normality at the first stage) We have focused on a diverging  $B_n$  to estimate  $a = (\mu, \sigma)$  at the first stage (recall (2.2)). Nevertheless, we could consider a constant  $B_n$ , say  $B_n \equiv B \in \mathbb{N}$  (then (3.6) is trivial), with which the CQMLE considered in Sect. 2 is asymptotically mixed normal (MN):

$$\left( \hat{u}_{a,n}, \hat{u}_{v,n} \right) \xrightarrow{\mathcal{L}} MN_{q+1} \left( 0, \Gamma_{a,0}^{-1} \right) \otimes N_1 \left( 0, \Gamma_{v,0}^{-1} \right),$$

now  $\Gamma_{a,0} = (2\sigma_0^2)^{-1} \text{diag}(S_0, 1)$  with  $S_0 = B^{-1} \int_0^B X_t'^{\otimes 2} dt$  being random. Then, (3.8) remains valid by the proof of (3.5) in Theorem 3.1. Without going into details, we give some brief remarks. The asymptotic mixed normality is deduced as in Masuda (2019) with a slight modification of the proof of the stable convergence in law of  $\Delta_{a,n}$  (to handle the filtration-structure issue caused by a Wiener process  $w$  driving  $X'_2$ ; see (Masuda 2019, Section 6.4.2)). Because of the stability of the convergence in law, the conclusion of Theorem 3.2 remains valid as it is with all the “ $N_n$ ” therein replaced by “ $nB$ ”.

Our primary theoretical interest was to deduce the mighty convergence of  $(\hat{u}_{a,n}, \hat{u}_{v,n})$ : not only the asymptotic normality but also the tail-probability estimate. If we are only interested in deriving the asymptotic normality, there are several possible ways to significantly weaken Assumption 2.1. Here is a version with the essential boundedness of  $X'$ .

**Assumption 3.4** For  $X$ , we can associate an  $(\mathcal{F}_t)$ -adapted process  $t \mapsto X'_t$  having càdlàg sample paths, for which the following conditions hold:

1.  $\sup_{t \geq 0} |X'_t| < \infty$  a.s.
2. The same condition as in Assumption 2.1.3 holds except that (2.3) is weakened to

$$\frac{1}{N_n} \sum_{j=1}^{N_n} f(X'_{t_{j-1}}) \xrightarrow{P} \int \psi_f(z) \pi_0(dz).$$

3.  $\frac{1}{N_n} \sum_{j=1}^{N_n} E \left[ \left| \frac{1}{h} (\Delta_j X - h X'_{t_{j-1}}) \right|^2 \middle| \mathcal{F}_{t_{j-1}} \right] = o_p(1).$
4. Assumption 2.1.4 holds.
5.  $\frac{1}{\sqrt{N_n}} \sum_{j=1}^{N_n} E \left[ \frac{1}{h} (\Delta_j X - h X'_{t_{j-1}}) \frac{\partial \phi_1}{\phi_1} \left( \frac{\Delta_j J}{h} \right) \middle| \mathcal{F}_{t_{j-1}} \right] = o_p(1)$

**Theorem 3.5** *Let  $v_0 > 1$ . Suppose that (2.2), (3.1), (3.6), and Assumption 3.4 hold. Then, we have the joint asymptotic normality (3.8).*

### 3.3 Student-Lévy driven Markov process and Euler approximation

We can handle a class of Lévy driven stochastic differential equation for  $X_2$  in the context of Theorem 3.5. Consider a sample  $(Y_{t_j})_{j=0}^{[nT_n]}$  from a solution to the Markov process ( $Y_0 = 0$ ) described by the model

$$Y_t = \mu \cdot \int_0^t b(Y_s) ds + \sigma J_t,$$

where  $v_0 > 1$  and  $b : \mathbb{R} \rightarrow \mathbb{R}^q$  is a known measurable function. We associate  $X_t = \int_0^t b(Y_s) ds$  with  $X'_t := b(Y_t)$ . So far it is assumed that  $\Delta_j X = \int_{t_{j-1}}^{t_j} b(Y_s) ds$  is observable, which may seem unnatural in the present context. In this section, we will establish a variant of the asymptotic normality (3.8) when only  $(Y_{t_j})_{j=0}^{[nT_n]}$  is observable. To this end, we assume the following conditions, which will guarantee the ergodicity of  $Y$ :

**Assumption 3.6** The function  $b$  is continuously differentiable and satisfies that

$$\sup_{y \in \mathbb{R}} (|b(y)| \vee |\partial_y b(y)|) < \infty \quad \text{and} \quad \limsup_{|y| \rightarrow \infty} y \mu_0 \cdot b(y) = -\infty.$$

Let  $b_{j-1} := b(Y_{t_{j-1}})$  and introduce the following variant of  $\mathbb{H}_{1,n}(a)$ :

$$\tilde{\mathbb{H}}_{1,n}(a) := \sum_{j=1}^{N_n} \log \left\{ \frac{1}{h\sigma} \phi_1 \left( \frac{\Delta_j Y - h\mu \cdot b_{j-1}}{h\sigma} \right) \right\} =: \sum_{j=1}^{N_n} \log \left\{ \frac{1}{h\sigma} \phi_1 (\epsilon'_j(a)) \right\}.$$

Define  $\tilde{a}_n = (\tilde{\mu}_n, \tilde{\sigma}_n)$  by any  $\tilde{a}_n \in \operatorname{argmax} \tilde{\mathbb{H}}_{1,n}$ . We write

$$\tilde{\epsilon}_i = \tilde{\sigma}^{-1} \left( Y_i - Y_{i-1} - h \tilde{\mu}_n \cdot \sum_{j \in A_i} b_{j-1} \right)$$

for  $i = 1, \dots, [T_n]$  (recall the notation  $A_i = \{k \in \mathbb{N} : (i-1)n < k \leq in\}$ ). Next, we introduce

$$\tilde{\mathbb{H}}_{2,n}(v) := \sum_{i=1}^{[T_n]} \rho(\tilde{\epsilon}_i; v)$$

and  $\tilde{v}_n \in \operatorname{argmax} \tilde{\mathbb{H}}_{2,n}$ ; recall the definition  $\rho(\varepsilon; n) = \log f(\varepsilon; 0, 1, \nu)$  (Sect. 3.1). Let  $\tilde{u}_{a,n} := \sqrt{N_n}(\tilde{a}_n - a_0)$ ,  $\tilde{u}_{v,n} := \sqrt{T_n}(\tilde{v}_n - \nu_0)$ , and moreover

$$\tilde{\Gamma}_{a,n} := \operatorname{diag} \left( \frac{1}{2\tilde{\sigma}_n^2} \tilde{S}_n, \frac{1}{2\tilde{\sigma}_n^2} \right), \quad \tilde{\Gamma}_{v,n} := \frac{1}{4} \left( \psi_1 \left( \frac{\tilde{v}_n}{2} \right) - \psi_1 \left( \frac{\tilde{v}_n + 1}{2} \right) \right),$$

where  $\tilde{S}_n := N_n^{-1} \sum_{j=1}^{N_n} b_{j-1}^{\otimes 2}$ .

**Theorem 3.7** *Let  $\nu_0 > 2$  and suppose that (3.1), (3.6), and Assumption 3.6 hold. Then,  $Y$  is ergodic and admits a unique invariant distribution  $\pi_{0,Y}(dy)$ . Moreover, we have*

$$\left( \tilde{\Gamma}_{a,n}^{1/2} \tilde{u}_{a,n}, \tilde{\Gamma}_{v,n}^{1/2} \tilde{u}_{v,n} \right) = \operatorname{diag} \left( \tilde{\Gamma}_{a,n}^{1/2}, \tilde{\Gamma}_{v,n}^{1/2} \right) M_n + o_p(1) \xrightarrow{\mathcal{L}} N_{q+2} \left( 0, I_{q+2} \right)$$

provided that  $S_0 := \int b(y)^{\otimes 2} \pi_{0,Y}(dy)$  is positive definite.

We see from the proof of Theorem 3.7 that rephrasing observed variables from  $(X_{t_j})$  to  $(X'_{t_j})$  when  $X_t = \int_0^t X'_s ds$  is possible in a more general framework.

### 4 Numerical experiment

In this section, we consider two deterministic and periodic regressors and the dynamics of the model have the following form:

$$Y_t = \mu_1 \cos(5t) + \mu_2 \sin(t) + \sigma J_t. \tag{4.1}$$

Since the Student- $t$  distribution is not closed under convolution as was stated in the introduction, we adopt an approximation method based on the inversion of the characteristic function for generating  $J_h$  with general  $h \neq 1$ . More specifically, our approximation is as follows: First, on some grid  $x_0, \dots, x_N$ , we calculate the value of the density function of  $J_h$  based on the following general inversion formula:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-iux} \varphi(u) du,$$

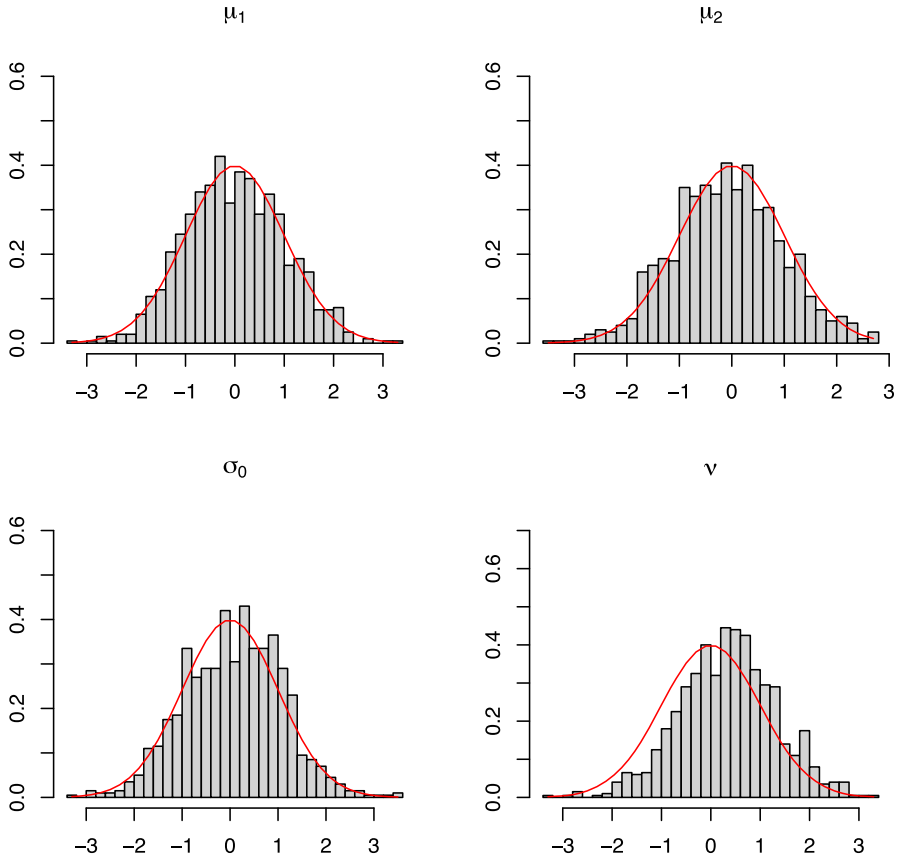
by using a numerical integration method, for example, Fast Fourier Transformation, Fourier Cosine expansion, and so on. Next, we approximate the cumulative distribution function using the Left-Riemann summation computed on the grid  $x_0, \dots, x_N$ , therefore the cumulative distribution function  $F(\cdot)$  for each  $x_j$  on the grid is determined as follows:

$$\hat{F}(x_j) = \sum_{x_k < x_j} \hat{f}_N(x_k) \Delta x,$$

where  $\hat{f}_N(x_k)$  denotes the approximated value of  $f(x_k)$ . Finally, we evaluate the cumulative distribution function at any  $x \in (x_{j-1}, x_j)$ , by interpolating linearly its value using the couples  $(x_{j-1}, \hat{F}(x_{j-1}))$  and  $(x_j, \hat{F}(x_j))$ . The random numbers can be obtained using the inversion sampling method. Our simulation method and estimation procedure are available in YUIMA package on R; see the companion paper (Masuda et al. 2024) for details.

For the first simulation exercise, we set  $T_n = 400$ ,  $B_n = 20$  and  $h = 1/200$ . We simulate 1000 trajectories of the model with deterministic regressors in (4.1) with  $\mu_1 = 5$ ,  $\mu_2 = -1$ ,



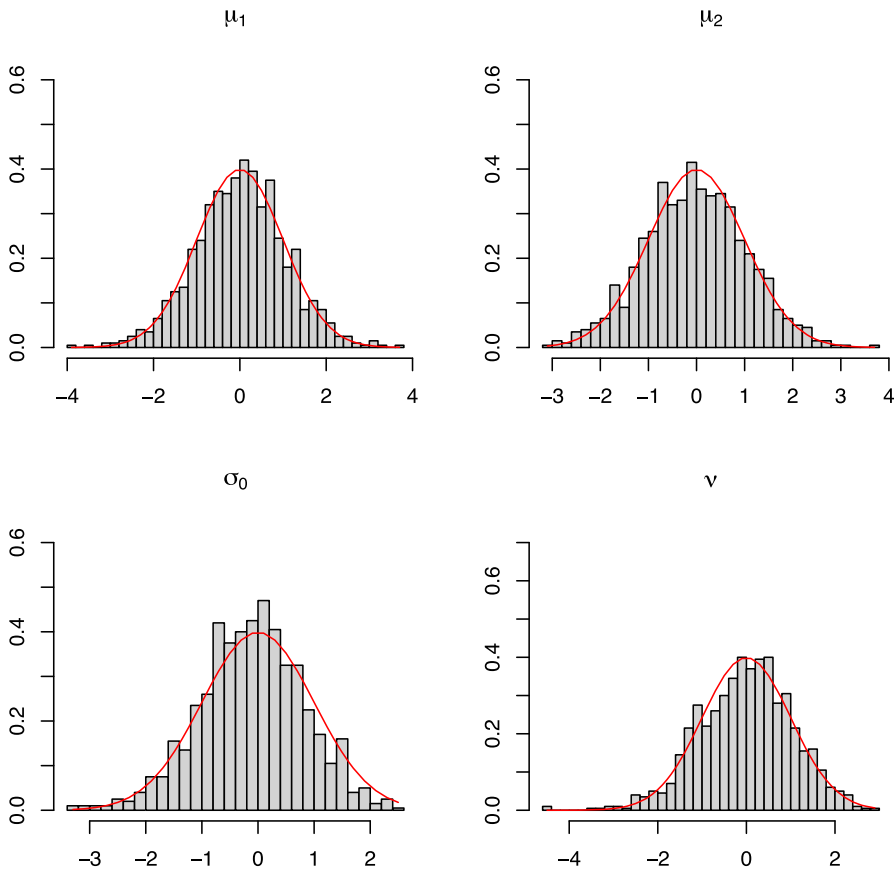


**Fig. 1** Simulated distribution of studentized parameters. In this case, we have  $\mu_1 = 5, \mu_2 = -1, \sigma_0 = 3$  and  $\nu = 1$

$\sigma_0 = 3$  and  $\nu = 1$ . Estimating the parameters for each trajectory, we obtain the empirical distribution of the studentized estimates. Figure 1 shows the simulated distribution of studentized estimates. Notably, all histograms demonstrate that the standard normal approximation adequately captures the behavior of model parameters. Favorable results are obtained for the regressors ( $\mu_1, \mu_2$ ) and the scale coefficient  $\sigma_0$ . However, a small upward bias appears for the degree of freedom  $\nu$ . This bias can be controlled by increasing the numerical precision and/or considering a larger value for  $T_n$ .

In the second simulation exercise, we focus on the case of  $\nu = 2$  and all other inputs remain unchanged. Figure 2 shows the histograms of the studentized estimators. The comparison demonstrates a favorable agreement between the simulated density functions and the standard normal distribution for all model parameters. This comparison suggests that a larger value of the degree of freedom  $\nu$  requires a smaller step-size value  $h$  to ensure accurate approximations.

In conclusion, the analysis of estimator behavior through these simulation exercises provides valuable insights into their performance within the Student Lévy Regression model. The results indicate satisfactory performance for most estimators, with minor biases observed for the degree of freedom parameter. Adjustments in numerical precision and step-time intervals



**Fig. 2** Simulated distribution of studentized parameters. In this case, we have  $\mu_1 = 5, \mu_2 = -1, \sigma_0 = 3$  and  $\nu = 2$

can effectively control these biases. We refer to Masuda et al. (2024) for further numerical experiments and real data analysis.

## 5 Proofs

### 5.1 Proof of Theorem 2.2

Let

$$\mathbb{Y}_{1,n}(a) := \frac{1}{N_n} (\mathbb{H}_{1,n}(a) - \mathbb{H}_{1,n}(a_0)),$$

$$\Gamma_{a,n} := -\frac{1}{N_n} \partial_a^2 \mathbb{H}_{1,n}(a_0).$$

We will prove the following three lemmas (recall (2.6) for the definition of  $\mathbb{Y}_1(a)$ ).

**Lemma 5.1** Under (2.2) and Assumption 2.1, there exist constants  $c_0 > 0$  and  $c_2 \in (0, 1/2]$  such that for all  $K > 0$ ,

$$\sup_n E \left[ \sup_a (N_n^{c_2} |\mathbb{Y}_{1,n}(a) - \mathbb{Y}_1(a)|)^K \right] < \infty.$$

**Lemma 5.2** Under (2.2) and Assumption 2.1, there exist a constant  $c_1 \in (0, 1/2)$  such that for every  $K > 0$ ,

$$\begin{aligned} & \sup_n E \left[ |\Delta_{a,n}|^K \right] + \sup_n E \left[ (N_n^{c_1} |\Gamma_{a,n} - \Gamma_{a,0}|)^K \right] \\ & + \sup_n E \left[ \sup_a \left| \frac{1}{N_n} \partial_a^3 \mathbb{H}_{1,n}(a) \right|^K \right] < \infty. \end{aligned} \tag{5.1}$$

**Lemma 5.3** Under (2.2) and Assumption 2.1, we have  $(\Delta_{a,n}, \Gamma_{a,n}) \xrightarrow{\mathcal{L}} (\Gamma_{a,0}^{1/2} \eta, \Gamma_{a,0})$ , where  $\eta \sim N_{q+1}(0, I_{q+1})$ .

Theorem 2.2 follows from the above lemmas and (2.7): with Lemmas 5.1 and 5.2, we can establish the tail-probability estimate (2.9) through (Yoshida 2011, Theorem 3(c)); moreover, the asymptotic normality (2.8) follows from the standard likelihood-asymptotics argument together with Lemmas 5.3. We omit the details.

By (1.4), the probability density of  $\mathcal{L}(h^{-1}J_h)$  is given by

$$\begin{aligned} f_h(y; \nu) &= \frac{1}{2\pi} \int \cos(uy) \{ \varphi_{J_1, \nu}(h^{-1}u) \}^h du \\ &= \left( \frac{2^{1-\nu/2} h^{-\nu/2}}{\Gamma(\nu/2)} \right)^h \frac{h}{\pi} \int_0^\infty \cos(uhx) u^{\nu h/2} (K_{\nu/2}(u))^h du, \end{aligned} \tag{5.2}$$

the existence of which is ensured by Bertoin and Doney (1997, Proposition 1) together with the locally Cauchy property  $h^{-1}J_h \xrightarrow{\mathcal{L}} t_1$ . The next lemma quantifies the speed of the local Cauchy approximation in  $L^1(dy)$  and serves as a basic tool for deriving limit theorems.

**Lemma 5.4** For any  $r > 0$  and any measurable function  $\zeta : \mathbb{R} \rightarrow \mathbb{R}$  such that  $|\zeta(y)| \lesssim 1 + \{\log(1 + |y|)\}^K$  for some  $K > 0$ , we have

$$\int |\zeta(y)| |f_h(y) - \phi_1(y)| dy \lesssim h^{1-r}. \tag{5.3}$$

In particular, under (2.2) we have

$$\sqrt{N_n} \int |\zeta(y)| |f_h(y) - \phi_1(y)| dy \rightarrow 0. \tag{5.4}$$

Related remarks on the estimate under the total variation distance can be found in Clément and Gloter (2020, Remark 2.1).

### 5.1.1 Proof of Lemma 5.1

Let  $\epsilon_j = \epsilon_{n,j} := \epsilon_j(a_0)$ , so that  $\epsilon_1, \epsilon_2 \dots \sim$  i.i.d.  $\mathcal{L}(h^{-1}J_h)$ , which admits the density  $f_h(y)$  defined by (5.2); this may be an abuse of notation (recall  $\epsilon_i = J_i - J_{i-1}$  introduced in Sect. 3.1), but no confusion is likely to arise. We have

$$\epsilon_j(a) = \frac{\sigma_0}{\sigma} \epsilon_j - \frac{1}{\sigma} (\mu - \mu_0) \cdot X'_{t_{j-1}} - \frac{1}{\sigma} (\mu - \mu_0) \cdot R_{n,j},$$

where  $R_{n,j} := h^{-1} (\Delta_j X - h X'_{t_{j-1}})$ . Let  $g_1(\epsilon) := \partial_\epsilon \phi_1(\epsilon) / \phi_1(\epsilon)$  and recall the definition (2.5) of  $\zeta(y, x'; a)$ . We have the decomposition  $\mathbb{Y}_{1,n}(a) = \mathbb{Y}_{1,n}^*(a) + \delta_{11,n}(a)$  where

$$\mathbb{Y}_{1,n}^*(a) = \frac{1}{N_n} \sum_{j=1}^{N_n} \zeta(\epsilon_j, X'_{t_{j-1}}; a), \tag{5.5}$$

$$\begin{aligned} \delta_{11,n}(a) &:= -\frac{1}{N_n} \sum_{j=1}^{N_n} \int_0^1 g_1 \left( \frac{\sigma_0}{\sigma} \epsilon_j - \frac{1}{\sigma} (\mu - \mu_0) \cdot X'_{t_{j-1}} - \frac{s}{\sigma} (\mu - \mu_0) \cdot R_{n,j} \right) ds \\ &\quad \times \frac{1}{\sigma} (\mu - \mu_0) \cdot R_{n,j}. \end{aligned} \tag{5.6}$$

Fix any  $K > 2$ . By Assumption 2.1 and (2.2), we can pick a sufficiently small  $c_2 \in (0, \frac{c_X}{2-\epsilon'} \wedge \frac{1}{2})$  to obtain

$$\max_{j \leq N_n} E \left[ |N_n^{c_2} R_{n,j}|^K \right] \lesssim \max_{j \leq N_n} E \left[ |n^{c_2(2-\epsilon')} R_{n,j}|^K \right] \leq \max_{j \leq [nT_n]} E \left[ |n^{c_2(2-\epsilon')} R_{n,j}|^K \right] \rightarrow 0.$$

Since  $g_1$  is essentially bounded, we have

$$E \left[ \sup_a |N_n^{c_2} \delta_{11,n}(a)|^K \right] \lesssim \max_{j \leq N_n} E \left[ |N_n^{c_2} R_{n,j}|^K \right] \rightarrow 0.$$

Write  $E^{j-1}[\cdot] = E[\cdot | \mathcal{F}_{t_{j-1}}]$ . Then,

$$\mathbb{Y}_{1,n}^*(a) = \frac{1}{\sqrt{N_n}} \sum_{j=1}^{N_n} \frac{1}{\sqrt{N_n}} \tilde{\zeta}_j(a) + \frac{1}{N_n} \sum_{j=1}^{N_n} E^{j-1} \left[ \zeta(\epsilon_j, X'_{t_{j-1}}; a) \right], \tag{5.7}$$

where  $\tilde{\zeta}_j(a) := \zeta(\epsilon_j, X'_{t_{j-1}}; a) - E^{j-1}[\zeta(\epsilon_j, X'_{t_{j-1}}; a)]$ . Obviously,

$$\sup_n \sup_a \left( E \left[ \left| \sum_{j=1}^{N_n} \frac{1}{\sqrt{N_n}} \tilde{\zeta}_j(a) \right|^K \right] + E \left[ \left| \sum_{j=1}^{N_n} \frac{1}{\sqrt{N_n}} \partial_a \tilde{\zeta}_j(a) \right|^K \right] \right) < \infty.$$

Since  $\Theta$  is assumed to be a bounded convex domain, the Sobolev inequality (Adams and Fournier 2003) holds:

$$\sup_a |F(a)|^r \lesssim \int_{\Theta_\mu \times \Theta_\sigma} (|F(a)|^r + |\partial_a F(a)|^r) da$$

for any  $\mathcal{C}^1$ -function  $F$  and  $r > \dim(a) = q + 1$  (see Adams and Fournier (2003) for details). Applying this and Fubini's theorem, we conclude that

$$\sup_n E \left[ \sup_a \left| \sum_{j=1}^{N_n} \frac{1}{\sqrt{N_n}} \tilde{\zeta}_j(a) \right|^K \right] < \infty,$$

followed by

$$\sup_n E \left[ \sup_a \left( N_n^{c_2} \left| \mathbb{Y}_{1,n}(a) - \frac{1}{N_n} \sum_{j=1}^{N_n} E^{j-1}[\zeta(\epsilon_j, X'_{t_{j-1}}; a)] \right| \right)^K \right] < \infty.$$

Observe that

$$\begin{aligned} & \frac{1}{N_n} \sum_{j=1}^{N_n} E^{j-1} [\zeta (\epsilon_j, X'_{t_{j-1}}; a)] \\ &= \frac{1}{N_n} \sum_{j=1}^{N_n} f' (X'_{t_{j-1}}; a) + \frac{1}{N_n} \sum_{j=1}^{N_n} \int \zeta (y, X'_{t_{j-1}}; a) (f_h(y) - \phi_1(y)) dy \\ &=: \mathbb{Y}_{1,n}^{**}(a) + \delta_{12,n}(a). \end{aligned} \tag{5.8}$$

Note that  $\sup_a |\zeta (y, X'_{t_{j-1}}; a)| \lesssim 1 + \log (1 + |X'_{t_{j-1}}|) + \log(1 + |y|) \lesssim \{1 + \log (1 + |X'_{t_{j-1}}|)\} \{1 + \log(1 + |y|)\}$ . Also,  $N_n^{c_2} h^{1-c} \lesssim n^{c_2(2-\epsilon')} n^{c-1} \rightarrow 0$  with  $c_2, c > 0$  small enough. It follows from Lemma 5.4 and Assumption 2.1 that

$$\begin{aligned} E \left[ \sup_a |N_n^{c_2} \delta_{12,n}(a)|^K \right] &\lesssim \max_{j \leq N_n} E \left[ \left| (1 + |X'_{t_{j-1}}|)^C N_n^{c_2} h^{1-c} \right|^K \right] \\ &\lesssim (N_n^{c_2} h^{1-c})^K \sup_{0 \leq t \leq h N_n} E \left[ \left(1 + |X'_{t_{j-1}}|\right)^C \right] \rightarrow 0. \end{aligned}$$

Moreover, since  $\max_{k \in \{0,1\}} \sup_a |\partial_a^k f'(x'; a)| \lesssim 1 + |x|^C$ , Assumption 2.1.3 ensures that

$$\max_{k \in \{0,1\}} \sup_n \sup_a E \left[ \left| N_n^{c_X} \left( \partial_a^k \mathbb{Y}_{1,n}^{**}(a) - \int \psi_{\partial_a^k f'(\cdot; a)}(z) \pi_0(dz) \right) \right|^K \right] < \infty$$

for every  $K > 0$ .

By piecing together what we have seen with the Sobolev inequality as before, we obtain the following estimate with a sufficiently small  $c_2 > 0$ :

$$\begin{aligned} & \sup_n E \left[ \sup_a (N_n^{c_2} |\mathbb{Y}_{1,n}(a) - \mathbb{Y}_1(a)|)^K \right] \\ & \lesssim 1 + \sup_n E \left[ \sup_a (N_n^{c_2} |\mathbb{Y}_{1,n}^{**}(a) - \mathbb{Y}_1(a)|)^K \right] \lesssim 1. \end{aligned}$$

### 5.1.2 Proof of Lemma 5.2

We will look at the three terms in (5.1) separately.

#### Quasi-score function

Recall that  $\epsilon_j = \epsilon_j(a_0) \sim \text{i.i.d. } \mathcal{L}(h^{-1} J_h)$ . Direct calculations give

$$\begin{aligned} \Delta_{a,n} &= -\sigma_0^{-1} \frac{1}{\sqrt{N_n}} \sum_{j=1}^{N_n} \left( \frac{(h^{-1} \Delta_j X) g_1(\epsilon_j)}{1 + \epsilon_j g_1(\epsilon_j)} \right) \\ &= -\sigma_0^{-1} \frac{1}{\sqrt{N_n}} \sum_{j=1}^{N_n} \left( \frac{X'_{t_{j-1}} g_1(\epsilon_j)}{1 + \epsilon_j g_1(\epsilon_j)} \right) - \sigma_0^{-1} \frac{1}{N_n} \sum_{j=1}^{N_n} \left( \frac{\sqrt{N_n} R_{n,j} g_1(\epsilon_j)}{0} \right). \end{aligned} \tag{5.9}$$

The first term in (5.9) equals

$$\begin{aligned} & -\sigma_0^{-1} \frac{1}{\sqrt{N_n}} \sum_{j=1}^{N_n} \left( X'_{t_{j-1}} (g_1(\epsilon_j) - E[g_1(\epsilon_j)]) \right) \\ & -\sigma_0^{-1} \frac{1}{N_n} \sum_{j=1}^{N_n} \sqrt{N_n} \left( X'_{t_{j-1}} (E[g_1(\epsilon_j)] - \int g_1(y)\phi_1(y)dy) \right. \\ & \left. - \int (1 + y\partial g_1(y))\phi_1(y)dy \right). \\ & =: r'_{1,n} + r''_{1,n}. \end{aligned}$$

The Burkholder inequality gives the  $L^K(P)$ -boundedness of  $(r'_{1,n})_n$  for any  $K > 0$ . By Lemma 5.4, the summand in  $r''_{1,n}$  can be bounded in absolute value by a constant multiple of  $(1 + |X'_{t_{j-1}}|) \sqrt{N_n} h^{1-c} \lesssim (1 + |X'_{t_{j-1}}|) n^{c-\epsilon'/2}$ . The sequence  $(r''_{1,n})_n$  is also  $L^K(P)$ -bounded since  $c > 0$  can be taken arbitrarily small so that  $c \leq \epsilon'/2$ .

For the second term in (5.9), by the compensation and the Burkholder inequality it suffices to show that for any  $K \geq 2$ ,

$$\frac{1}{N_n} \sum_{j=1}^{N_n} E \left[ \left| \sqrt{N_n} E^{j-1} [R_{n,j} g_1(\epsilon_j)] \right|^K \right] = o(1).$$

This is ensured by the last convergence in Assumption 2.1.2.

**Quasi-observed information**

The components of  $\Gamma_{a,n}$  consists of

$$\begin{aligned} -\frac{1}{N_n} \partial_\mu^2 \mathbb{H}_{1,n}(a_0) &= -\sigma_0^{-2} \frac{1}{N_n} \sum_{j=1}^{N_n} (h^{-1} \Delta_j X)^{\otimes 2} \partial g_1(\epsilon_j), \\ -\frac{1}{N_n} \partial_\sigma^2 \mathbb{H}_{1,n}(a_0) &= -\sigma_0^{-2} \frac{1}{N_n} \sum_{j=1}^{N_n} (1 + 2\epsilon_j g_1(\epsilon_j) + \epsilon_j^2 \partial g_1(\epsilon_j)), \\ -\frac{1}{N_n} \partial_\mu \partial_\sigma \mathbb{H}_{1,n}(a_0) &= -\sigma_0^{-2} \frac{1}{N_n} \sum_{j=1}^{N_n} (h^{-1} \Delta_j X) (g_1(\epsilon_j) - \epsilon_j \partial g_1(\epsilon_j)). \end{aligned}$$

To conclude that  $\sup_n E \left[ (N_n^{c_1} |\Gamma_{a,n} - \Gamma_{a,0}|)^K \right] < \infty$ , we note the following three preliminary steps (we can take  $c_1 > 0$  as small as we want):

- First, we replace  $h^{-1} \Delta_j X$  in the summands by  $R_{n,j} + X'_{t_{j-1}}$ ; this together with Assumption 2.1 enables us to replace  $h^{-1} \Delta_j X$  in the summands by  $X'_{t_{j-1}}$ ;
- Second, we extract the martingale terms by replacing the parts of the form  $\eta(\epsilon_j)$  by  $\eta(\epsilon_j) - E[\eta(\epsilon_j)]$  and then apply the Burkholder inequality to the latter parts;
- Third, we apply Lemma 5.4 to the expressions  $E[\eta(\epsilon_j)]$  with the facts  $-\int \partial g_1(y)\phi_1(y)dy = 1/2$ ,  $-\int (1 + 2yg_1(y) + y^2\partial g_1(y))\phi_1(y)dy = 1/2$ , and  $\int (g_1(y) - y\partial g_1(y))\phi_1(y)dy = 0$ .

It remains to note that, as in the last paragraph in the proof of Lemma 5.1, Assumption 2.1.3 ensures that (recall that  $S_0 = \int \psi_f(z)\pi_0(dz)$  for  $f(x') = x'^{\otimes 2}$ : Assumption 2.1.4)

$$\sup_n E \left[ \left( N_n^{c_1} \left| \frac{1}{N_n} \sum_{j=1}^{N_n} X'_{t_{j-1}}{}^{\otimes 2} - S_0 \right| \right)^K \right] < \infty.$$

**Third-order derivatives**

We have

$$\frac{1}{N_n} \sup_a |\partial_a^3 \mathbb{H}_{1,n}(a)| \lesssim \frac{1}{N_n} \sum_{j=1}^{N_n} (1 + |h^{-1} \Delta_j X|),$$

hence  $\sup_n E [\sup_a |N_n^{-1} \partial_a^3 \mathbb{H}_{1,n}(a)|^K] < \infty$ .

**5.1.3 Proof of Lemma 5.3**

We have  $\Delta_{a,n} = r'_{1,n} + o_p(1)$  from the arguments in Sect. 5.1.2 (with  $c < \epsilon'/2$ ). Write  $r'_{1,n} = \sum_{j=1}^{N_n} \zeta_{n,j}$ . Then, the  $(\mathcal{F}_{t_j})$ -martingale difference array  $(\zeta_{n,j})_{j \leq [nT_n]}$  satisfies the Lyapunov condition  $\left( \sum_{j=1}^{N_n} E^{j-1} [|\zeta_{n,j}|^{\delta+2}] = o_p(1) \text{ for } \delta > 0 \right)$ . Also, by using Lemma 5.4 as in Sect. 5.1.2 together with the facts  $\int g_1(y)^2 \phi_1(y) dy = \int (1 + yg_1(y))^2 \phi_1(y) dy = 1/2$  and  $\int g_1(y) (1 + yg_1(y)) \phi_1(y) dy = 0$ , we obtain

$$\begin{aligned} & \frac{1}{N_n} \sum_{j=1}^{N_n} E^{j-1} [\zeta_{n,j}^{\otimes 2}] \\ &= \sigma_0^{-2} \frac{1}{N_n} \sum_{j=1}^{N_n} \left( \begin{array}{cc} (h^{-1} \Delta_j X)^{\otimes 2} g_1(\epsilon_j)^2 & \text{sym.} \\ (h^{-1} \Delta_j X) g_1(\epsilon_j) (1 + \epsilon_j g_1(\epsilon_j)) & (1 + \epsilon_j g_1(\epsilon_j))^2 \end{array} \right) + o_p(1) \\ &= \sigma_0^{-2} \frac{1}{N_n} \sum_{j=1}^{N_n} \left( \begin{array}{cc} X'_{t_{j-1}}{}^{\otimes 2} \int g_1(y)^2 \phi_1(y) dy & 0 \\ 0 & \int (1 + yg_1(y))^2 \phi_1(y) dy \end{array} \right) + o_p(1) \\ &= \Gamma_{a,0} + o_p(1). \end{aligned}$$

It follows that  $\Delta_{a,n} \xrightarrow{\mathcal{L}} N_{q+1}(0, \Gamma_{a,0})$  by the central limit theorem for martingale difference arrays (Shiryaev 1980, Chapter VII.8). The convergence  $\Gamma_{a,n} \xrightarrow{P} \Gamma_{a,0}$  is automatic by the argument in Sect. 5.2; indeed, it holds almost surely by Borel-Cantelli lemma.

**5.1.4 Proof of Lemma 5.4**

As in Masuda (2019, Example 2.7), our proof is based on the explicit form of the characteristic function (1.3). We will proceed similarly to the proof of Masuda (2019, Lemma 2.2), with omitting some details of the calculation.

We write  $g(z)$  for the Lévy density of  $\mathcal{L}(J_1)$  and  $g_0(z) := \pi^{-1} z^{-2}$  for the Lévy density of the standard Cauchy distribution  $t_1$ . Let  $\rho(z) := g(z)/g_0(z) - 1$  so that  $g(z) = g_0(z)(1 +$

$\rho(z)$ ); the function  $\rho$  quantifies how much the  $\mathcal{L}(J_1)$  is far from  $t_1$  around the origin. Then, by Raible (2000, Proposition 2.18) we have

$$\rho(z) = \frac{\pi}{4} (1 - \nu_0) |z| + o(|z|), \quad |z| \rightarrow 0. \tag{5.10}$$

Let  $\varphi_h(u)$  and  $\varphi_0(u)$  denote the characteristic functions of  $\mathcal{L}(h^{-1}J_h)$  and  $t_1$ , respectively; we have  $\varphi_0(u) = \exp \left\{ \int (\cos(uy) - 1) g_0(y) dy \right\} = e^{-|u|}$  and already noted in (2.1) that

$$\varphi_h(u) = \left( \frac{2^{1-\nu/2}}{\Gamma(\nu/2)} \right)^h (|u/h|^{\nu/2} K_{\nu/2}(|u/h|))^h \rightarrow \varphi_0(u), \quad h \rightarrow 0.$$

By straightforward computations, we obtain  $\varphi_h(u) = \exp \left\{ \int (\cos(uy) - 1) h^2 g(hy) dy \right\} = \varphi_0(u) \exp \{ \chi_h(u) \}$  where

$$\chi_h(u) := \int (\cos(uy) - 1) g_0(y) \rho(hy) dy.$$

Pick a  $\delta \in (0, 1)$  for which  $\sup_{|y| \leq \delta} |\rho(y)| \leq 1/2$ . By Sato (1999, Theorem 25.3) we know the equivalence

$$\int_{|z|>1} |z|^r g(z) dz < \infty \iff E[|J_1|^r] < \infty \quad (r > 0).$$

The function  $\rho$  may be unbounded for  $|z| \rightarrow \infty$  when  $\nu_0 \in (0, 1)$ . Now observe that

$$\begin{aligned} |\chi_h(u)| &\leq \int (1 - \cos(uy)) g_0(y) |\rho(hy)| dy \\ &= \int_{|y| \leq \delta/h} (1 - \cos(uy)) g_0(y) |\rho(hy)| dy + \int_{|y| > \delta/h} (1 - \cos(uy)) g_0(y) |\rho(hy)| dy \end{aligned} \tag{5.11}$$

$$\leq \frac{1}{2} \int_{|y| \leq \delta/h} (1 - \cos(uy)) g_0(y) dy + \frac{2}{\pi^2} \int_{|y| > \delta/h} y^{-2} |\rho(hy)| dy. \tag{5.12}$$

The first term in (5.12) can be bounded by  $-2^{-1} \int (\cos(uy) - 1) g_0(y) dy = |u|/2$ . The integral in the second term in (5.12) equals  $h \int_{|y| > \delta} y^{-2} |\rho(y)| dy$ , which is  $O(h)$  if  $\rho$  is bounded. When  $\rho$  is unbounded, we have  $\rho(y) \rightarrow \infty$  as  $|y| \rightarrow \infty$  hence there is a  $c_\rho > \delta$  for which  $\rho(y) \geq 0$  for  $|y| \geq c_\rho$ . In this case,  $\int_{|y| > \delta} y^{-2} |\rho(y)| dy = \int_{|y| \in [\delta, c_\rho]} y^{-2} |\rho(y)| dy + \int_{|y| > c_\rho} y^{-2} \rho(y) dy \lesssim \int_{|y| \in [\delta, c_\rho]} y^{-2} dy + \pi \int_{|y| > c_\rho} (g(y) - g_0(y)) dy < \infty$ , so that the integral in the second term in (5.12) is  $O(h)$ . It follows that  $|\chi_h(u)| \leq |u|/2 + Ch$ . Also, the first term in (5.11) can be bounded by a constant multiple of  $h \{ (1 + u^2) + \log(1/h) \}$ : this can be seen by dividing the domain of integration into  $\{y : |y| \leq \delta/h, |y| > c\}$  and  $\{y : |y| \leq \delta/h, |y| \leq c\}$ , and then proceeding as before with using the fact  $|1 - \cos x| \lesssim x^2$  for the latter subdomain.

Following the same line as in the proof of Masuda (2019, Lemma 5.1), we can conclude that for any  $H > 0$  and  $\xi \in (0, 1)$ ,

$$\int_0^\infty |\varphi_h(u) - \varphi_0(u)| du \leq \int_0^\infty (u^{-\xi} \vee u^H) |\varphi_h(u) - \varphi_0(u)| du \lesssim h \log(1/h). \tag{5.13}$$

This combined with the Fourier inversion formula yields

$$\sup_{y \in \mathbb{R}} |f_h(y) - \phi_1(y)| \lesssim h \log(1/h).$$



As in the proof of Masuda (2019, Eq. (5.1)), we can derive the estimate

$$\sup_{h \in (0,1]} \sup_{M > 0} M^{(1 \wedge \nu_0) - \kappa} \int_{|y| > M} |y|^\kappa f_h(y) dy < \infty$$

valid for any  $\kappa \in (0, 1 \wedge \nu_0)$ , considering the cases  $\nu_0 \geq 1$  and  $\nu_0 \in (0, 1)$  (for which  $\rho$  is unbounded) separately as before.

Let  $(b_n)$  be a positive real sequence such that  $b_n \uparrow \infty$  as  $n \rightarrow \infty$ . To deduce (5.3), we proceed with estimating  $\delta'_n := \int_{|y| > b_n} |\zeta(y)| |f_h(y) - \phi_1(y)| dy$  and  $\delta''_n := \int_{|y| \leq b_n} |\zeta(y)| |f_h(y) - \phi_1(y)| dy$  separately. Since the density  $f_h$  is bounded continuous uniformly in  $h \in (0, 1]$ , we have  $\sup_{h \in (0,1]} f_h(y) \lesssim |y|^{-(1+1 \wedge \nu_0) + \epsilon_0}$  for any  $|y| \geq 1$  and for any small  $\epsilon_0 > 0$ ; recall (Sato 1999, Theorem 25.3) already mentioned above. Since  $\zeta(y)$  is of at most logarithmic-power growth order, there exists a sufficiently small  $\epsilon_1 \in (0, 1 \wedge \nu_0)$  for which

$$\begin{aligned} \delta'_n &\leq \int_{|y| > b_n} |\zeta(y)| f_h(y) dy + \int_{|y| > b_n} |\zeta(y)| \phi_1(y) dy \\ &\lesssim b_n^{-1 \wedge \nu_0 + \epsilon_1} + b_n^{-1 + \epsilon_1} \lesssim b_n^{-1 \wedge \nu_0 + \epsilon_1}. \end{aligned} \tag{5.14}$$

Turning to  $\delta''_n$ , we will show that  $\delta''_n \lesssim b_n^{c''} h \log(1/h)$  for arbitrarily small  $c'' > 0$ . Let  $\psi_h(u) := \log \varphi_h(u)$  and  $\psi_0(u) := \log \varphi_0(u)$ . Fix  $c \in (0, 1]$  arbitrarily. By the exactly same procedure as in Masuda (2019, Eq. (5.4)), we obtain

$$\delta''_n \lesssim (1 + b_n^{\epsilon_2}) \int_{|y| \leq b_n} |f_h(y) - \phi_1(y)| dy \lesssim (1 + b_n^{\epsilon_2}) (\delta''_{1,n}(c) + \delta''_{2,n}(c)),$$

where  $\epsilon_2 > 0$  can be taken arbitrarily small and where

$$\delta''_{1,n}(c) := b_n^c \int_0^\infty u^c |\varphi_h(u) - \varphi_0(u)| |\partial_u \psi_h(u)| du, \tag{5.15}$$

$$\delta''_{2,n}(c) := b_n^c \int_0^\infty u^c |\partial_u \psi_h(u) - \partial_u \psi_0(u)| \varphi_0(u) du. \tag{5.16}$$

We will take a closer look at these quantities through the specific form of  $\varphi_h(u)$ .

Put  $s = \nu_0/2$  in the sequel. We have  $\psi_h(u) = hC_s + sh \log u + sh \log(1/h) + h \log K_s(u/h)$  for  $u > 0$  with some constant  $C_s$  only depending on  $s$ . Since  $\partial_z K_\alpha(z) = -K_{\alpha-1}(z) - (\alpha/z)K_\alpha(z)$ , we obtain the expression  $\partial_u \psi_h(u) - \partial_u \psi_0(u) = 1 - K_{s-1}(u/h)/K_s(u/h) =: \xi_s(u)$ . The function  $x \mapsto x\xi_s(x)$  is smooth on  $(0, \infty)$  and we can deduce that  $\sup_{x > 0} |x\xi_s(x)| < \infty$  as follows:

- We have  $|x\xi_s(x)| \lesssim x^{1 \wedge \nu_0}$  around the origin, by using the property  $K_{r'}(x) \sim \Gamma(|r'|) 2^{|r'-1|} x^{-|r'|}$  for  $x \rightarrow 0$  ( $r' \neq 0$ );
- Moreover,  $x\xi_s(x) = x \left( \frac{1}{8x} + O(x^{-1}) \right) = O(1)$  for  $x \uparrow \infty$  since  $K_s(x) \sim \sqrt{\frac{\pi}{2x}} \left( 1 + \frac{4s^2-1}{8x} + O(x^{-2}) \right)$  for  $x \uparrow \infty$ .

Thus, we have obtained  $|\partial_u \psi_h(u) + 1| = |\xi_s(u/h)| \lesssim h/u$  and  $|\partial_u \psi_h(u)| \lesssim 1 + h/u$  ( $u > 0$ ). Substituting these estimates and (5.13) into (5.15) and (5.16), we obtain  $\delta''_{1,n}(c) \lesssim b_n^c h \log(1/h)$  and  $\delta''_{2,n}(c) \lesssim b_n^c h$ . Hence it follows that  $\delta''_n \lesssim b_n^{c+\epsilon_2} h \log(1/h)$  where we can set  $c + \epsilon_2 > 0$  arbitrarily small as was desired.

Combining (5.14) and the last estimate of  $\delta''_n$  now gives  $\int |\zeta(y)| |f_h(y) - \phi_1(y)| dy \lesssim b_n^{-(1 \wedge \nu_0) + \epsilon_1} + b_n^{c_2 + c} h \log(1/h)$ . This upper bound is minimized for  $b_n = (h \log(1/h))^{-1/\lambda}$

with  $\chi := 1 \wedge v_0 + c + \epsilon_2 - \epsilon_1$ , with which we get

$$\int |f_h(y) - \phi_1(y)| dy \lesssim (h \log(1/h))^{1 - \frac{c+\epsilon_2}{\chi}}.$$

Given any small  $r > 0$ , we can take all of  $\epsilon_1, \epsilon_2, c > 0$  small enough to conclude (5.3). Now (5.4) is trivial under (2.2).

## 5.2 Proof of Theorem 3.1

### 5.2.1 Proof of (3.4): Tail-probability estimate

This section aims to prove (3.4) through (Yoshida 2011, Theorem 3(c)). Define

$$\begin{aligned} \mathbb{Y}_{2,n}(v) &= \frac{1}{T_n} (\mathbb{H}_{2,n}(v) - \mathbb{H}_{2,n}(v_0)), \\ \mathbb{Y}_2(v) &= E[\rho(\epsilon_1; v) - \rho(\epsilon_1; v_0)] = \int \log\left(\frac{f(\epsilon; v)}{f(\epsilon; v_0)}\right) f(\epsilon; v_0) d\epsilon. \end{aligned}$$

**Lemma 5.5** 1. *There exists a positive constant  $c'_0 > 0$  such that  $\mathbb{Y}_2(v) \leq -c'_0|v - v_0|^2$  for every  $v \in \Theta_v$ .*

2. *There exists a constant  $c'_2 \in (0, 1/2]$  such that for every  $K > 0$ ,*

$$\sup_n E \left[ \sup_v \left| T_n^{c'_2} (\mathbb{Y}_{2,n}(v) - \mathbb{Y}_2(v)) \right|^K \right] < \infty.$$

*In particular,  $\sup_v |\mathbb{Y}_{2,n}(v) - \mathbb{Y}_2(v)| \xrightarrow{P} 0$ .*

3. *The consistency holds:  $\hat{v}_n \xrightarrow{P} v_0$ .*

**Proof** The proof of 1 is similar to the case of  $\mathbb{Y}_1(a)$  in Sect. 2, and the consistency 3 is an obvious consequence of 1 and 2.

To prove 2, we fix  $K > 0$  in the rest of this proof. We will repeatedly use the following estimate:

$$\begin{aligned} &\sup_{\epsilon} \sup_v \left( \sup_{k \geq 0} \sup_{l \geq 2} |\partial_{\epsilon}^k \partial_v^l \rho(\epsilon; v)| + \frac{\max_{l \in \{0,1\}} |\partial_v^l \rho(\epsilon; v)|}{1 + \log(1 + \epsilon^2)} \right. \\ &\quad \left. + \sup_{k \geq 1} \max_{l \in \{0,1\}} \left( 1 + |\epsilon|^k \right) |\partial_{\epsilon}^k \partial_v^l \rho(\epsilon; v)| \right) < \infty. \end{aligned} \tag{5.17}$$

Since  $\epsilon_1, \dots, \epsilon_{[T_n]}$  are i.i.d., we have the moment estimate

$$\sup_n E \left[ \left| \frac{1}{\sqrt{T_n}} \sum_{i=1}^{[T_n]} \left( \partial_v^l \rho(\epsilon_i; v) - E[\partial_v^l \rho(\epsilon_1; v)] \right) \right|^K \right] < \infty$$

for any  $v \in \overline{\Theta}_v$  and  $l \in \mathbb{Z}_+$  through the the Sobolev-inequality argument as before. Hence it suffices to show that there exists a constant  $c'_2 \in (0, 1/2]$  such that

$$\sup_n E \left[ \sup_v \left| T_n^{c'_2} \left( \frac{1}{T_n} \sum_{i=1}^{[T_n]} (\rho(\hat{\epsilon}_i; v) - \rho(\epsilon_i; v)) \right) \right|^K \right] < \infty. \tag{5.18}$$

To manage the term “ $\rho(\hat{\varepsilon}_i; \nu) - \rho(\varepsilon_i; \nu)$ ”, we will separately consider it on different four events. Let

$$H_{1,i} := \left\{ |\hat{\varepsilon}_i - \varepsilon_i| \leq \frac{1}{2} |\varepsilon_i| \right\}, \quad i = 1, \dots, [T_n].$$

Writing  $\hat{u}_{\mu,n} = \sqrt{N_n} (\hat{\mu}_n - \mu_0)$  and  $\hat{u}_{\sigma,n} = \sqrt{N_n} (\hat{\sigma}_n - \sigma_0)$ , we have

$$\hat{\varepsilon}_i - \varepsilon_i = \frac{-1}{\hat{\sigma}_n \sqrt{N_n}} (\varepsilon_i \hat{u}_{\sigma,n} + (X_i - X_{i-1}) \cdot \hat{u}_{\mu,n}). \tag{5.19}$$

Let  $(b_n)_n$  be a positive sequence tending to infinity. Then,

$$\begin{aligned} H_{1,i}^c &\subset \left\{ |\hat{u}_{\sigma,n}| |\varepsilon_i| + |X_i - X_{i-1}| |\hat{u}_{\mu,n}| \geq C \sqrt{N_n} |\varepsilon_i| \right\} \\ &\subset \left\{ (|X_i - X_{i-1}| + |\varepsilon_i|) (|\hat{u}_{\mu,n}| + |\hat{u}_{\sigma,n}|) \geq C \sqrt{N_n} |\varepsilon_i| \right\} \\ &\subset \{ |\varepsilon_i| \leq 1 \} \cup \left\{ |\varepsilon_i| \geq 1, |\hat{u}_{\mu,n}| + |\hat{u}_{\sigma,n}| \geq C \sqrt{N_n} \frac{|\varepsilon_i|}{|\varepsilon_i| + b_n} \right\} \\ &\quad \cup \{ |X_i - X_{i-1}| \geq b_n \} \\ &\subset \{ |\varepsilon_i| \leq 1 \} \cup \left\{ |\hat{u}_{\mu,n}| + |\hat{u}_{\sigma,n}| \geq C \frac{\sqrt{N_n}}{b_n} \right\} \cup \{ |X_i - X_{i-1}| \geq b_n \} \\ &=: H_{2,i} \cup H_{3,n} \cup H_{4,i}. \end{aligned}$$

First, for  $H_{1,i}$  and  $H_{2,i}$ , we take a closer look at the right-hand side of the expression

$$\rho(\hat{\varepsilon}_i; \nu) - \rho(\varepsilon_i; \nu) = \left( \int_0^1 \partial_\varepsilon \rho(\varepsilon_i + s(\hat{\varepsilon}_i - \varepsilon_i); \nu) ds \right) (\hat{\varepsilon}_i - \varepsilon_i). \tag{5.20}$$

On  $H_{1,i}$ , we have

$$\inf_{s \in [0,1]} |\varepsilon_i + s(\hat{\varepsilon}_i - \varepsilon_i)| \geq \frac{1}{2} |\varepsilon_i|.$$

Hence by (5.17) and (5.19),

$$\begin{aligned} |\rho(\hat{\varepsilon}_i; \nu) - \rho(\varepsilon_i; \nu)| I_{H_{1,i}} &\lesssim \frac{1}{\sqrt{N_n}} \left( |\hat{u}_{\sigma,n}| + \frac{|X_i - X_{i-1}|}{1 + |\varepsilon_i|} |\hat{u}_{\mu,n}| \right), \\ |\rho(\hat{\varepsilon}_i; \nu) - \rho(\varepsilon_i; \nu)| I_{H_{2,i}} &\lesssim \frac{1}{\sqrt{N_n}} (|\hat{u}_{\sigma,n}| + |X_i - X_{i-1}| |\hat{u}_{\mu,n}|). \end{aligned}$$

The last two displays together with the tail-probability estimate (2.9) and (3.1) imply that

$$E \left[ \sup_\nu \left| \sqrt{T_n} \left( \frac{1}{T_n} \sum_{i=1}^{[T_n]} (\rho(\hat{\varepsilon}_i; \nu) - \rho(\varepsilon_i; \nu)) I_{H_{1,i} \cup H_{2,i}} \right) \right|^K \right] \lesssim \left( \frac{T_n}{N_n} \right)^{\frac{K}{2}} \rightarrow 0. \tag{5.21}$$

Second, we consider  $H_{3,n}$  and  $H_{4,i}$ . To this end, we do not use the expression (5.20) but directly estimate the target expectation. Recalling (2.9), for any  $K_1, K_2 > 0$  we obtain

$$\begin{aligned} \max_{1 \leq i \leq [T_n]} P [H_{3,n} \cup H_{4,i}] &\lesssim \left(\frac{b_n}{\sqrt{N_n}}\right)^{K_1} E \left[ |\hat{u}_{\mu,n}|^{K_1} + |\hat{u}_{\sigma,n}|^{K_1} \right] \\ &\quad + b_n^{-K_2} E \left[ |X_i - X_{i-1}|^{K_2} \right] \\ &\lesssim \left(\frac{b_n}{\sqrt{N_n}}\right)^{K_1} + b_n^{-K_2}. \end{aligned}$$

Now we set  $b_n = N_n^{1/4}$ . Given any  $K > 0$ , under (3.1) we can take sufficiently large  $K_1, K_2 > 0$  so that

$$\max_{1 \leq i \leq [T_n]} P [H_{3,n} \cup H_{4,i}] = o \left( T_n^{-K} \right). \tag{5.22}$$

We will use the following (rough) estimate: there exists a constant  $C' > 0$  only depending on  $K$  and  $v_0$  for which  $\{\log(1 + \epsilon^2)\}^{2K} \leq C' (1 + |\epsilon|^{v_0/2})$  ( $\epsilon \in \mathbb{R}$ ). With this observation together with (5.17) and (5.19), we have

$$\begin{aligned} &\frac{1}{T_n} \sum_{i=1}^{[T_n]} E \left[ \sup_v |\rho(\hat{\epsilon}_i; v) - \rho(\epsilon_i; v)|^{2K} \right]^{1/2} \\ &\lesssim 1 + \frac{1}{T_n} \sum_{i=1}^{[T_n]} \left( E \left[ \{\log(1 + \hat{\epsilon}_i^2)\}^{2K} + \{\log(1 + \epsilon_i^2)\}^{2K} \right] \right)^{1/2} \\ &\lesssim 1 + \frac{1}{T_n} \sum_{i=1}^{[T_n]} \left( E \left[ |\hat{\epsilon}_i|^{v_0/2} + |\epsilon_i|^{v_0/2} \right] \right)^{1/2} \\ &\lesssim 1 + \frac{1}{T_n} \sum_{i=1}^{[T_n]} \left( E \left[ |\epsilon_i|^{v_0/2} + \left( \frac{1}{\sqrt{N_n}} (|\epsilon_i| |\hat{u}_{\sigma,n}| + |X_i - X_{i-1}| |\hat{u}_{\mu,n}|) \right)^{v_0/2} \right] \right)^{1/2} \\ &\lesssim 1. \end{aligned} \tag{5.23}$$

Combining (5.22) and (5.23), Jensen’s inequality and Cauchy-Schwarz inequality yield

$$\begin{aligned} &E \left[ \sup_v \left| \sqrt{T_n} \left( \frac{1}{T_n} \sum_{i=1}^{[T_n]} (\rho(\hat{\epsilon}_i; v) - \rho(\epsilon_i; v)) I_{H_{3,n} \cup H_{4,i}} \right) \right|^K \right] \\ &\lesssim T_n^{K/2} E \left[ \frac{1}{T_n} \sum_{i=1}^{[T_n]} \sup_v |\rho(\hat{\epsilon}_i; v) - \rho(\epsilon_i; v)|^K I_{H_{3,n} \cup H_{4,i}} \right] \\ &\lesssim T_n^{K/2} \frac{1}{T_n} \sum_{i=1}^{[T_n]} E \left[ \sup_v |\rho(\hat{\epsilon}_i; v) - \rho(\epsilon_i; v)|^{2K} \right]^{1/2} P [H_{3,n} \cup H_{4,i}]^{1/2} \\ &\lesssim T_n^{K/2} \max_{1 \leq i \leq [T_n]} P [H_{3,n} \cup H_{4,i}]^{1/2} \frac{1}{T_n} \sum_{i=1}^{[T_n]} E \left[ \sup_v |\rho(\hat{\epsilon}_i; v) - \rho(\epsilon_i; v)|^{2K} \right]^{1/2} \\ &\lesssim o(1). \end{aligned} \tag{5.24}$$

The claim (5.18) with  $c'_2 = 1/2$  follows from (5.21) and (5.24). □

As in the proof of (2.9) in Theorem 2.2, the claim (3.4) follows on showing the next lemma.

**Lemma 5.6** *There exist a constant  $c'_1 \in (0, 1/2)$  such that for every  $K > 0$ ,*

$$\sup_n \left( T_n^{c'_1} |\Gamma_{v,n} - \Gamma_{v,0}| + \sup_v \left| \frac{1}{T_n} \partial_v^3 \mathbb{H}_{2,n}(v) \right| \right) + \sup_n E \left[ |\Delta_{v,n}|^K \right] < \infty,$$

where  $\Gamma_{v,n} := -T_n^{-1} \partial_v^2 \mathbb{H}_{2,n}(v_0)$ .

**Proof** By (3.2), the finiteness of the first (non-random) term is trivial for  $c'_1 \in (0, 1]$ . Turning to the second term, we have

$$\Delta_{v,n} = \frac{1}{\sqrt{T_n}} \sum_{i=1}^{[T_n]} \partial_v \rho(\varepsilon_i; v_0) + \frac{1}{\sqrt{T_n}} \sum_{i=1}^{[T_n]} (\partial_v \rho(\hat{\varepsilon}_i; v_0) - \partial_v \rho(\varepsilon_i; v_0)) \tag{5.25}$$

Let  $K > 0$ . Since  $E[\partial_v \rho(\varepsilon_i; v_0)] = 0$ , the first term is obviously  $L^K(P)$ -bounded. Exactly in the same way as in the proof of (5.18) with  $c'_2 = 1/2$ , the second term converges in  $L^K$  to 0. □

### 5.2.2 Proof of (3.5): Asymptotic normality

To prove the asymptotic normality (3.5), we introduce the concave random function

$$\Lambda_n(u) := \mathbb{H}_{2,n} \left( v_0 + \frac{u}{\sqrt{T_n}} \right) - \mathbb{H}_{2,n}(v_0)$$

defined for  $u \in \{v \in \mathbb{R} : v_0 + v T_n^{-1/2} \in \Theta_v\}$ ; obviously,  $\hat{u}_{v,n} \in \operatorname{argmax} \Lambda_n$ . By means of Hjört (2011, Basic Lemma), we can conclude that

$$\hat{u}_{v,n} = \Gamma_{v,0}^{-1} \Delta_{v,n} + o_p(1) \xrightarrow{\mathcal{L}} N(0, \Gamma_{v,0}^{-1})$$

by showing the locally asymptotically quadratic structure: for each  $u \in \mathbb{R}$ ,

$$\Lambda_n(u) = \Delta_{v,n} u - \frac{1}{2} \Gamma_{v,0} u^2 + o_p(1), \tag{5.26}$$

where the random variable  $\Delta_{v,n} \xrightarrow{\mathcal{L}} N(0, \Gamma_{v,0})$ . By Lemma 5.6 we are left to show the asymptotic normality of  $\Delta_{v,n}$ . But it is trivial since  $\varepsilon \mapsto \partial_v \rho(\varepsilon; v)$  is smooth uniformly in a neighborhood of  $v_0$  so that  $E[\{\partial_v \rho(\varepsilon_1, v_0)\}^2] = -E[\partial_v^2 \rho(\varepsilon_1, v_0)] = \Gamma_{v,0}$ .

### 5.3 Proof of Theorem 3.5

It suffices to show (2.8) and (3.5) individually, the same arguments as in Sect. 3.2 are valid to deduce the asymptotic orthogonality of  $\hat{u}_{a,n}$  and  $\hat{u}_{v,n}$ .

For the consistency of  $\hat{a}_n$ , it suffices to verify  $\sup_a |\mathbb{Y}_{1,n}(a) - \mathbb{Y}_1(a)| \xrightarrow{P} 0$ . In Sect. 5.1.1, we considered the expression

$$\mathbb{Y}_{1,n}(a) - \mathbb{Y}_1(a) = \mathbb{Y}_{1,n}^{**}(a) - \mathbb{Y}_1(a) + \frac{1}{\sqrt{N_n}} \sum_{j=1}^{N_n} \frac{1}{\sqrt{N_n}} \tilde{\zeta}_j(a) + \delta_{11,n}(a) + \delta_{12,n}(a).$$

Assumption 3.4 implies that  $N_n^{-1} \sum_{j=1}^{N_n} |h^{-1}(\Delta_j X - hX'_{t_{j-1}})|^2 = o_p(1)$ , hence  $\sup_a |\delta_{11,n}(a)| = o_p(1)$  from the expression (5.6). Recall (5.5) and (5.7). By the definition

(2.5) we have

$$\sup_a |\zeta_j(a)| \lesssim 1 + \log(1 + \epsilon_j^2) + \log(1 + |X'_{t_{j-1}}|^2) \lesssim 1 + \log(1 + \epsilon_j^2),$$

so that Lemma 5.4 gives  $\sup_a |\delta_{12,n}(a)| = o_p(1)$ . Since we also have  $\sup_a |\partial_a \zeta_j(a)| \lesssim 1 + |X'_{t_{j-1}}| \lesssim 1$ , the Burkholder and Sobolev inequalities yield  $\sup_a \left| \sum_{j=1}^{N_n} N_n^{-1/2} \tilde{\zeta}_j(a) \right| = O_p(1)$ . Finally, from (5.8), we have  $\mathbb{Y}_{1,n}^{**}(a) - \mathbb{Y}_1(a) \xrightarrow{P} 0$  for each  $a$  and moreover,

$$\sup_a |\partial_a \mathbb{Y}_{1,n}^{**}(a)| + \sup_a |\partial_a \mathbb{Y}_1(a)| \lesssim \frac{1}{N_n^{-1}} \sum_{j=1}^{N_n} \left| \frac{1}{h} \Delta_j X \right| + 1 = O_p(1).$$

Hence  $\sup_a |\mathbb{Y}_{1,n}^{**}(a) - \mathbb{Y}_1(a)| \xrightarrow{P} 0$ , followed by  $\sup_a |\mathbb{Y}_{1,n}(a) - \mathbb{Y}_1(a)| \xrightarrow{P} 0$ . For the asymptotic normality, we can follow the same line as in the proof of Lemma 5.3 (Sect. 5.1.3) under Assumption 3.4 except for the following point: concerned with the second term in (5.9) about the first-stage quasi-score  $\Delta_{a,n}$ , we here do not require the moment boundedness (the last condition in Assumption 2.1.2) but just its negligibility. Recall the notation  $R_{n,j} = h^{-1} (\Delta_j X - hX'_{t_{j-1}})$  and  $g_1(\epsilon) = \partial_\epsilon \phi_1(\epsilon) / \phi_1(\epsilon)$ . By the compensation of  $R_{n,j} g_1(\epsilon_j)$  and then applying the Burkholder inequality, it is straightforward to verify the negligibility as follows: for the martingale part  $M'_n := \sum_{j=1}^{N_n} N_n^{-1/2} \{R_{n,j} g_1(\epsilon_j) - E^{j-1} [R_{n,j} g_1(\epsilon_j)]\}$ , we can estimate its quadratic characteristic as

$$\langle M' \rangle_n \lesssim \frac{1}{N_n} \sum_{j=1}^{N_n} E^{j-1} [|R_{n,j}|^2] \xrightarrow{P} 0$$

by Assumption 3.4.3; and moreover, the compensation part is negligible under Assumption 3.4.5. Thus, we conclude (2.8).

To deduce (3.5) under  $\sqrt{N_n}(\hat{a}_n - a_0) = O_p(1)$ , we make use of what we have seen in Sects. 5.2.1 and 5.2.2. The  $u$ -wise asymptotically quadratic structure (5.26) of  $\Lambda_n(\cdot)$  holds since the estimate  $\sup_n |\Gamma_{v,n} - \Gamma_{v,0}| + \sup_{n,v} |T_n^{-1} \partial_v^3 \mathbb{H}_{2,n}(v)| < \infty$  is valid as before. Note that

$$\begin{aligned} \frac{1}{T_n} \sum_{i=1}^{[T_n]} |X_i - X_{i-1}| &= \frac{1}{nT_n} \sum_{i=1}^{[T_n]} \sum_{j \in A_i} \left| \frac{1}{h} \Delta_j X \right| \\ &\lesssim \left( \frac{1}{[nT_n]} \sum_{j=1}^{[nT_n]} \left| \frac{1}{h} (\Delta_j X - hX'_{t_{j-1}}) \right|^2 \right)^{1/2} + 1 = O_p(1). \end{aligned}$$

By (5.17) and (5.19), the second term in the right-hand side of (5.25) can be bounded as follows:

$$\begin{aligned} &\left| \frac{1}{\sqrt{T_n}} \sum_{i=1}^{[T_n]} (\partial_v \rho(\hat{\epsilon}_i; v_0) - \partial_v \rho(\epsilon_i; v_0)) \right| \\ &= \left| \frac{1}{T_n} \sum_{i=1}^{[T_n]} \int_0^1 \partial_\epsilon \partial_v \rho(\epsilon_i + s(\hat{\epsilon}_i - \epsilon_i); v_0) ds \sqrt{N_n} (\hat{\epsilon}_i - \epsilon_i) \sqrt{\frac{T_n}{N_n}} \right| \\ &\lesssim \left( |\hat{\mu}_{\sigma,n}| \frac{1}{T_n} \sum_{i=1}^{[T_n]} |\epsilon_i| + |\hat{\mu}_{\mu,n}| \frac{1}{T_n} \sum_{i=1}^{[T_n]} (1 + |X_i - X_{i-1}|) \right) \sqrt{\frac{T_n}{N_n}} \end{aligned}$$

$$= O_p \left( \sqrt{\frac{T_n}{N_n}} \right) = o_p(1).$$

This leads to (3.5).

### 5.4 Proof of Theorem 3.7

First, we will verify Assumption 3.4 (except for item 4 therein) under Assumption 3.6. First, we show items 1 and 2 partly related to the ergodic behavior of the Markov process  $Y$ . The Lévy measure  $\nu(dz)$  of  $J$  satisfies that  $\nu(\{z : |z| < c\}) > 0$  for each  $c > 0$  (see (5.10)). From the proof of Masuda (2013, Proposition 5.4) under Assumption 5.2(ii) therein, we see that the Local Doeblin (LD) condition holds for  $Y$  (see Kulik (2009) for details). The LD condition implies that, for any constant  $\Delta > 0$ , every compact set is petite for the discrete-time Markov chain  $(Y_{m\Delta})_{m \geq 0}$ . Moreover, the Lévy process  $J$  is centered and the function  $\psi(y) = y^2$  belongs to the domain of the (extended) generator  $\mathcal{A}$  of  $Y$ :

$$\begin{aligned} \mathcal{A}\psi(y) &:= \partial\psi(y) \mu_0 \cdot b(y) + \int (\psi(y + \sigma_0 z) - \psi(y) - \partial\psi(y)\sigma_0 z) g(z) dz \\ &= 2y\mu_0 \cdot b(y) + \frac{\sigma_0^2}{\nu_0 - 2}. \end{aligned}$$

By the latter condition in Assumption 3.6, we can find constants  $K, c > 0$  such that  $\mathcal{A}\psi(y) \leq -c$  for every  $|y| \geq K$ ; as a matter of fact, it suffices that the rightmost side is negative uniformly in  $|y| \geq K$ , but then the condition involves the unknown  $\sigma_0$  and  $\nu_0$  hence inconvenient. Obviously, there exists a constant  $d' = d'(K) > 0$  for which  $\mathcal{A}\psi(y) \leq d$  for every  $|y| \leq K$ . We get  $\mathcal{A}\psi(y) \leq -c + (c + d')I(|y| \leq K)$  for  $y \in \mathbb{R}$ . Hence, it follows from Masuda (2007, Theorem 2.1) that  $Y$  is ergodic, which in particular means that there exists a unique invariant distribution  $\pi_{0,Y}(dy)$  for which  $Y$  satisfies the law of large numbers: for any continuously differentiable  $\pi_{0,Y}$ -integrable  $g : \mathbb{R} \rightarrow \mathbb{R}$ , we have  $T^{-1} \int_0^T g(Y_s) ds \rightarrow \int g(y)\pi_{0,Y}(dy)$  as  $T \rightarrow \infty$ . The last convergence holds for the bounded process  $X' = b(Y)$  as well. It is routine to show that  $N_n^{-1} \sum_{j=1}^{N_n} g \circ b(Y_{t_{j-1}}) \xrightarrow{P} \int g \circ b(y)\pi_{0,Y}(dy)$  for  $g$  being bounded and smooth enough. Thus, items 1 and 2 hold with  $\pi_0(dx') = \pi_{0,Y} \circ b^{-1}(dx')$ .

Turning to item 3, we will show  $\lim_n N_n^{-1} \sum_{j=1}^{N_n} A_{nj} = 0$ , where

$$A_{nj} := E \left[ \left| \frac{1}{h} (\Delta_j X - hX'_{t_{j-1}}) \right|^2 \right] = E \left[ \left| \frac{1}{h} \int_{t_{j-1}}^{t_j} (b(Y_s) - b(Y_{t_{j-1}})) ds \right|^2 \right].$$

Fix  $\epsilon > 0$  and  $H \in (0, 1 \wedge \nu_0) = (0, 1)$ . Pick a  $\delta > 0$  for which  $|b(x) - b(y)| < \sqrt{\epsilon}$  for every  $x, y$  such that  $|x - y| \leq \delta$ . By using the boundedness of  $b$ ,

$$\begin{aligned} A_{nj} &\leq \frac{1}{h} \int_{t_{j-1}}^{t_j} E \left[ |b(Y_s) - b(Y_{t_{j-1}})|^2 \right] ds \\ &\lesssim \frac{1}{h} \int_{t_{j-1}}^{t_j} (P[|Y_s - Y_{t_{j-1}}| > \delta] + E[|b(Y_s) - b(Y_{t_{j-1}})|^2; |Y_s - Y_{t_{j-1}}| \leq \delta]) ds \\ &\leq \frac{1}{h} \int_{t_{j-1}}^{t_j} P[|Y_s - Y_{t_{j-1}}| > \delta] ds + \epsilon \end{aligned}$$

$$\begin{aligned} &\lesssim \frac{1}{h} \int_{t_{j-1}}^{t_j} \left( (s - t_{j-1})^H + E \left[ |J_{s-t_{j-1}}|^H \right] \right) ds + \epsilon \\ &\lesssim O \left( h^H \right) + \sup_{t \in [0, h]} E \left[ |J_t|^H \right] + \epsilon = O \left( h^H \right) + \epsilon. \end{aligned} \tag{5.27}$$

In the last step, we applied (Luschgy and Pagès 2008, Theorem 2(c)) to conclude  $E \left[ \sup_{t \in [0, h]} |J_t|^H \right] \lesssim h^H$ . Since  $\epsilon > 0$  was arbitrary, it follows that  $\lim_n N_n^{-1} \sum_{j=1}^{N_n} A_{nj} = 0$ , hence item 3.

Finally, we verify item 5. Pick any constant  $\delta' \in (0, (1 + \epsilon')/2]$  for  $\epsilon' > 0$  in (2.2). By the essential boundedness of  $b, \partial b$ , and  $g_1$  and by the moment estimate

$$E \left[ |J_s - J_{t_{j-1}}| \right] \leq E \left[ \sup_{t \in [0, h]} |J_t| \right] \lesssim h \log(1/h).$$

due to Luschgy and Pagès (2008, Theorem 3), we have

$$\begin{aligned} &\left| \frac{1}{\sqrt{N_n}} \sum_{j=1}^{N_n} E^{j-1} \left[ \frac{1}{h} \left( \Delta_j X - h X'_{t_{j-1}} \right) g_1 \left( \frac{\Delta_j J}{h} \right) \right] \right| \\ &\lesssim \frac{1}{N_n} \sum_{j=1}^{N_n} \sqrt{N_n} \left( h + \frac{1}{h} \int_j E[|J_s - J_{t_{j-1}}|] ds \right) \\ &\lesssim \sqrt{N_n} (h + h \log(1/h)) \lesssim h^{1-\delta'} \sqrt{N_n} \times h^{\delta'} \log(1/h) \lesssim h^{1-\delta'} \sqrt{N_n} \times h^{\delta'/2}. \end{aligned}$$

Since  $h^{1-\delta'} \sqrt{N_n} \lesssim n^{\delta'-(1+\epsilon')/2} \lesssim 1$ , the last upper bound is  $O_p \left( h^{\delta'/2} \right) = o_p(1)$ , concluding item 5. Thus, we have shown that Assumption 3.4 follows from Assumption 3.6.

We will complete the proof of Theorem 3.7 by showing that  $\sqrt{N_n} (\tilde{a}_n - \hat{a}_n) = o_p(1)$  and  $\sqrt{T_n} (\tilde{v}_n - \hat{v}_n) = o_p(1)$ . Let  $\delta_n(a) := \tilde{\mathbb{H}}_{1,n}(a) - \mathbb{H}_{1,n}(a)$ . Tracing back and inspecting the proof of Theorem 3.5 reveal that it is sufficient for concluding  $\sqrt{N_n} (\tilde{a}_n - \hat{a}_n) = o_p(1)$  to verify

$$\left| \frac{1}{\sqrt{N_n}} \partial_a \delta_n(a_0) \right| + \max_{k \in \{0, 2, 3\}} \sup_a \left| \frac{1}{N_n} \partial_a^k \delta_n(a) \right| = o_p(1).$$

We can write  $\delta_n(a) = \sum_{j=1}^{N_n} \left\{ \log \phi_1 \left( \epsilon'_j(a) \right) - \log \phi_1 \left( \epsilon_j(a) \right) \right\} = \sum_{j=1}^{N_n} B_j(a) (h^{-1} \Delta_j X - b_{j-1})$  for some essentially bounded random functions  $B_j(a)$  smooth in  $a$ . By direct computations and Assumption 3.4.3,

$$\max_{k \in \{0, 2, 3\}} \sup_a \left| \frac{1}{N_n} \partial_a^k \delta_n(a) \right| \lesssim \frac{1}{N_n} \sum_{j=1}^{N_n} |h^{-1} \Delta_j X - b_{j-1}| = o_p(1)$$

for each  $k \in \{0, 2, 3\}$ . In a similar way to (5.27),

$$\begin{aligned} \left| \frac{1}{\sqrt{N_n}} \partial_a \delta_n(a_0) \right| &\lesssim \frac{1}{N_n} \sum_{j=1}^{N_n} \frac{1}{h} \int_{t_{j-1}}^{t_j} \sqrt{N_n} |b(Y_s) - b(Y_{t_{j-1}})| ds \\ &\lesssim \frac{1}{N_n} \sum_{j=1}^{N_n} \frac{1}{h} \int_{t_{j-1}}^{t_j} \sqrt{N_n} |Y_s - Y_{t_{j-1}}| ds \end{aligned}$$



$$\lesssim \frac{1}{N_n} \sum_{j=1}^{N_n} \frac{1}{h} \int_{t_{j-1}}^{t_j} \sqrt{N_n} ((s - t_{j-1}) + |J_s - J_{t_{j-1}}|) ds. \tag{5.28}$$

Therefore, by applying (Luschgy and Pagès 2008, Theorem 3) again,

$$E \left[ \left| \frac{1}{\sqrt{N_n}} \partial_a \delta_n(a_0) \right| \right] \lesssim h\sqrt{N_n} + E \left[ \sup_{t \in [0, h]} |J_t| \right] \lesssim O \left( h \log(1/h) \sqrt{N_n} \right).$$

The upper bound in the last display is  $o(1)$  since  $h\sqrt{N_n} \lesssim n^{-\epsilon'/2}$ .

To show  $\sqrt{T_n}(\tilde{v}_n - \hat{v}_n) = o_p(1)$ , we write  $\tilde{\mathbb{H}}_{2,n}(v) = \sum_{i=1}^{[T_n]} \rho(\tilde{\epsilon}_i; v)$ . From the proof of (3.5) (Sect. 5.2.2),

$$\begin{aligned} \tilde{\Lambda}_n(u) &:= \tilde{\mathbb{H}}_{2,n} \left( v_0 + \frac{u}{\sqrt{T_n}} \right) - \tilde{\mathbb{H}}_{2,n}(v_0) \\ &= \frac{u}{\sqrt{T_n}} \partial_v \tilde{\mathbb{H}}_{2,n}(v_0) - \frac{1}{2} \Gamma_{v,0} u^2 + o_p(1), \end{aligned}$$

Proceeding as in (5.28), we obtain (with recalling (5.17))

$$\begin{aligned} \left| \frac{1}{\sqrt{T_n}} \partial_v \tilde{\mathbb{H}}_{2,n}(v_0) - \Delta_{v,n} \right| &\lesssim \frac{1}{T_n} \sum_{i=1}^{[T_n]} \left( |\sqrt{T_n}(\tilde{\epsilon}_i - \hat{\epsilon}_i)| + |\sqrt{T_n}(\hat{\epsilon}_i - \epsilon_i)| \right) \\ &\lesssim O_p \left( h \log(1/h) \sqrt{T_n} \right) = O_p \left( n^{-\epsilon'/2} (\log n) \frac{T_n}{N_n} \right) = o_p(1). \end{aligned}$$

It follows that  $\sqrt{T_n}(\tilde{v}_n - \hat{v}_n) = o_p(1)$ .

**Acknowledgements** The authors are grateful to the two reviewers for their careful reading, which led to substantial improvements in the presentation and based on which we could fix some insufficient technical arguments. This work was supported by JST CREST Grant Number JPMJCR2115, Japan, by JSPS KAKENHI Grant Number 22H01139, and by JSPS KAKENHI Grant Number 23K13023.

**Author contributions** All authors wrote the main manuscript text and reviewed the manuscript.

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

## Declarations

**Conflict of interest** On behalf of all authors, the corresponding author states 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

Abramowitz M, Stegun IA (1992) (eds) Handbook of mathematical functions with formulas, graphs, and mathematical tables. Dover Publications Inc., New York. Reprint of the 1972 edition

- Adams RA, Fournier JJF (2003) Sobolev spaces, volume 140 of Pure and Applied Mathematics (Amsterdam). Elsevier/Academic Press, Amsterdam, second edition
- Barndorff-Nielsen OE (1998) Processes of normal inverse Gaussian type. *Finance Stoch* 2(1):41–68
- Bertoin J, Doney RA (1997) Spitzer's condition for random walks and Lévy processes. *Ann Inst H Poincaré Probab Stat* 33(2):167–178
- Clément E, Gloter A (2019) Estimating functions for SDE driven by stable Lévy processes. *Ann Inst Henri Poincaré Probab Stat* 55(3):1316–1348
- Clément E, Gloter A (2020) Joint estimation for SDE driven by locally stable Lévy processes. *Electron J Stat* 14(2):2922–2956
- Gloter A, Yoshida N (2021) Adaptive estimation for degenerate diffusion processes. *Electron J Stat* 15(1):1424–1472
- Harvey AC (2013) Dynamic models for volatility and heavy tails, volume 52 of *Econometric Society Monographs*. Cambridge University Press, Cambridge. With applications to financial and economic time series
- HjØrt NL, 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)
- Ivanenko D, Kulik AM, Masuda H (2015) Uniform LAN property of locally stable Lévy process observed at high frequency. *ALEA Lat Am J Probab Math Stat* 12(2):835–862
- Kulik A (2018) Ergodic behavior of Markov processes, vol 67. *De Gruyter Studies in Mathematics*. De Gruyter, Berlin
- Kulik AM (2009) Exponential ergodicity of the solutions to SDE's with a jump noise. *Stochast Process Appl* 119(2):602–632
- Luschgy H, Pagès G (2008) Moment estimates for Lévy processes. *Electron Commun Probab* 13:422–434
- Massing TPG (2019) *Stochastic Properties of Student-Lévy Processes with Applications*. PhD thesis, Universität Duisburg-Essen
- Masuda H (2007) Ergodicity and exponential  $\beta$ -mixing bounds for multidimensional diffusions with jumps. *Stochast Process Appl* 117(1):35–56
- Masuda H (2013) Convergence of Gaussian quasi-likelihood random fields for ergodic Lévy driven SDE observed at high frequency. *Ann Stat* 41(3):1593–1641
- Masuda H (2019) Non-Gaussian quasi-likelihood estimation of SDE driven by locally stable Lévy process. *Stochast Process Appl* 129(3):1013–1059
- Masuda H, Mercuri L, Uehara Y (2022) Noise inference for ergodic Lévy driven SDE. *Electron J Stat* 16(1):2432–2474
- Masuda H, Mercuri L, Uehara Y (2024) Student  $t$ -Lévy regression model in yuima. *ResearchGate preprint* <https://doi.org/10.13140/RG.2.2.26646.15682>
- Neyman J, Scott EL (1948) Consistent estimates based on partially consistent observations. *Econometrica* 16(1):1–32
- Proietti T, Pedregal DJ (2022) Seasonality in high-frequency time series. *Econometr Stat*
- Raible S (2000) Lévy processes in finance: theory, numerics, and empirical facts. PhD thesis, Universität Freiburg i. Br
- Sato K-i (1999) Lévy processes and infinitely divisible distributions, volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge. Translated from the 1990 Japanese original, Revised by the author
- Shiryaev AN (1980) *Probability*. Springer-Verlag, New York, second edition, 1996. Translated from the first Russian edition by R. P. Boas
- Yoshida N (2011) Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. *Ann Inst Stat Math* 63(3):431–479