Electronic Journal of Statistics
Vol. 14 (2020) 4165–4206
ISSN: 1935-7524
https://doi.org/10.1214/20-EJS1769

Rate-optimal estimation of the Blumenthal–Getoor index of a Lévy process*

Fabian Mies

RWTH Aachen University, Institute of Statistics
Wüllnerstr. 3, D-52064 Aachen, Germany
e-mail: mies@stochastik.rwth-aachen.de

Abstract: The Blumenthal–Getoor (BG) index characterizes the jump measure of an infinitely active Lévy process. It determines sample path properties and affects the behavior of various econometric procedures. If the process contains a diffusion term, existing estimators of the BG index based on high-frequency observations achieve rates of convergence which are suboptimal by a polynomial factor. In this paper, a novel estimator for the BG index and the successive BG indices is presented, attaining the optimal rate of convergence. If an additional proportionality factor needs to be inferred, the proposed estimator is rate-optimal up to logarithmic factors. Furthermore, our method yields a new efficient volatility estimator which accounts for jumps of infinite variation. All parameters are estimated jointly by the generalized method of moments. A simulation study compares the finite sample behavior of the proposed estimators with competing methods from the financial econometrics literature.

MSC2020 subject classifications: Primary 62M05; secondary 60G51.
Keywords and phrases: High-frequency, method of moments, jump activity, Fisher information, non-diagonal rate matrix.

Received August 2019.

1. Introduction

Models for continuous time stochastic processes with jumps have gained increased interest in the statistical literature, most prominently in financial econometrics where they are used as a model for asset prices (Andersen, Benzoni and Lund, 2002; Christensen, Oomen and Podolskij, 2014). The jump behavior of these processes X_t can be broadly characterized in terms of the jump activity index, given by

$$\alpha = \inf \left\{ p : \sum_{s \leq T} |\Delta X_s|^p < \infty \right\}. \tag{1.1}$$

Here, $\Delta X_s = X_s - X_{s-}$ denotes the size of a jump at time s. If X_t is a Lévy process, α is also known as the Blumenthal-Getoor index (Blumenthal and Getoor, 1961). The index α depends on the small jumps only, and for semimartingales,

*The results of this article are also part of the author's doctoral thesis (Mies, 2020).

its range is $\alpha \in [0,2]$. Various qualitative properties of the process X_t can be expressed in terms of the jump activity index. If the process has only finitely many jumps in total, then $\alpha = 0$, and if the jumps are of finite variation, we have $\alpha \leq 1$. Conversely, $\alpha < 1$ implies jumps of finite variation. Furthermore, the value of α has implications for various econometric procedures. For example, if the jumps are treated as a nuisance, jump-robust estimation of integrated volatility requires $\alpha < 1$ (Jacod and Reiss, 2014), as well as an efficient drift estimator due to Gloter, Loukianova and Mai (2018). In these applications, a higher jump activity typically induces a non-negligible bias which can not be easily corrected if the jumps are considered as a nuisance. Hence, highly active jumps need to be modeled more explicitly, as done by Amorino and Gloter (2020a) for drift estimation, and by Jacod and Todorov (2014, 2016) for volatility estimation.

As the jump activity index is a central property of infinite activity jump models, it is natural to consider statistical estimation of its precise value. Recent interest in this topic has been initiated by Aït-Sahalia and Jacod (2009), who study the estimation of α based on discrete high-frequency observations $X_{i/n}, i = 1, \ldots, n$, where X is an Itô semimartingale with a non-vanishing diffusion component. They specify (1.1) more precisely by defining α in terms of the spot jump compensator ν_t, assuming that $\nu_t\left((-x,x)^c\right) = r_t|x|^{-\alpha} + \mathcal{O}(|x|^{\delta-\alpha})$ as $|x| \to 0$ for a predictable process r_t, and some $\delta > 0$. The statistical challenge is that, based on discrete observations at a given frequency, the small jumps can hardly be distinguished from the continuous diffusion movement. One solution, originally proposed by Mancini (2006, 2009) for the estimation of volatility, is to introduce a threshold sequence $\tau_n \propto h_n^\omega \to 0$, and consider any increment $X_{i/n} - X_{(i-1)/n} > \tau_n$ to be due to a jump in the interval $((i-1)/n, i/n]$. For the estimation of the jump activity index α, Aït-Sahalia and Jacod (2009) use this approach and consider

$$U(\tau_n) = \sum_{i=1}^n \mathbb{1}\left(\left| X_{\frac{i}{n}} - X_{t_{\frac{i-1}{n}}} \right| > \tau_n \right). \tag{1.2}$$

If $\omega < 1/2$, the contribution of the diffusion towards the statistic $U(\tau_n)$ will be negligible. The jump activity can be identified via the approximate scaling relation $U(\tau_n) \propto \tau_n^{-\alpha}$, and Aït-Sahalia and Jacod (2009) show that this approach lends itself to derive an estimator of α with rate of convergence $n^{\alpha/10}$. Replacing the indicator in (1.2) by a suitable smooth function, Jing et al. (2012) improve this rate to $n^{\alpha/8}$. The method of Jing et al. (2012) may also be extended to a Markovian setting with state-dependent jump activity (Mies, 2021). So far, the best rates for estimating α have been achieved by the estimators of Reiß (2013) for the case that X_t is a Lévy process, and by Bull (2016) for Itô semimartingales. Both authors construct estimators which converge at rate $n^{\alpha/4-\epsilon}$ for arbitrary $\epsilon > 0$. In both cases, the precise form of the estimator depends on the desired rate defect $\epsilon > 0$.

In the considered high-frequency setting, the optimal rate of convergence for estimating α is conjectured to be $n^{\alpha/4}$, up to logarithmic factors. This lower bound is justified by the results of Aït-Sahalia and Jacod (2012), who study the diagonal entries of the Fisher matrix of a fully parametric submodel consisting of the sum of a Brownian motion and a symmetric α-stable Lévy motion. A matching LAN result is not available since the off-diagonal entries have not been studied. This lower bound is discussed in Section 3. It should be highlighted that the achievable rate of convergence for estimating α depends on whether the process contains a non-vanishing diffusion component. If we consider a pure-jump Itô semimartingale, the jump activity index can be estimated at rate $\sqrt{n}$ based on high-frequency observations (Todorov, 2015).

Although the estimators of Reiß (2013) and Bull (2016) almost achieve the optimal rate of convergence, there is so far no procedure which attains the $n^{\alpha/4}$ lower bound, even in the case where X_t is a Lévy process. This issue has also been formulated as an open problem by Reiß (2013). In this paper, we propose a new estimator of α for the Lévy case. If only α is unknown, the estimator achieves the optimal rate of convergence, matching the lower bound of Aït-Sahalia and Jacod (2012). If an additional proportionality factor r needs to be estimated, our estimator is rate-optimal up to a factor of $\log n$ for both r and α. Furthermore, we show that the diagonally rescaled Fisher matrix in the submodel considered by Aït-Sahalia and Jacod (2012) is asymptotically singular for the combined parameter (α, r), and hence we conjecture that our rate of convergence is in fact optimal. Our procedure also yields an efficient estimator of the volatility σ^2 of the diffusion component of X_t in the presence of jumps of infinite variation. Under analogous conditions on the jump behavior, Jacod and Todorov (2014, 2016) have derived a different efficient estimator of volatility which is robust to highly active jumps. Hence, our estimator is an alternative to the method of Jacod and Todorov (2014), although the latter is valid for Itô semimartingales and we restrict our attention to Lévy processes. The proposed estimator is based on the generalized method of moments, and we estimate the jump and the diffusion parameters jointly in a single step as the solution of a system of estimating equations. In the literature, the method of moments has been successfully applied to study various types of stochastic processes, e.g. nonlinear jump-diffusion processes (Jakobsen and Sørensen, 2019), or stochastic differential equations driven by fractional Brownian motion (Barboza and Viens, 2017).

Our model allows for an asymmetric behavior of the small jumps. In particular, for a Lévy process X_t with characteristic triplet (μ, σ^2, ν), we suppose that the Lévy measure ν is locally stable in the sense that, for z close to 0,

$$\nu(dz) \approx \tilde{\nu}(dz) = \sum_{m=1}^{M} \frac{\alpha_m}{|z|^{1+\alpha_m}} \left(r_m^+ \mathbb{1}_{z>0} + r_m^- \mathbb{1}_{z<0} \right) dz. \tag{1.3}$$

Here, M is a natural number, $r_m^\pm \geq 0$, $m = 1, \ldots, M$, and the $0 < \alpha_M < \ldots < \alpha_1 < 2$ are the successive Blumenthal-Getoor indices, as introduced by Aït-Sahalia and Jacod (2012). The approximation in (1.3) will be made precise in the sequel. In particular, the BG index of X_t will be $\alpha = \alpha_1$. We construct an estimator for the parameter vector $\theta \in \mathbb{R}^{3M+1}$ consisting of the volatility σ^2, the indices α_m, and the proportionality factors $r_m^\pm$.

The remainder of this paper is structured as follows. In Section 2, we present our model and the proposed estimator. A central limit theorem is given, establishing the rate $n^{\alpha/4}$. The rate of convergence and related lower bounds are discussed in Section 3. By means of a simulation study (Section 4), we compare the finite sample properties of our method with the jump activity estimators of Bull (2016); Reiß (2013) and the volatility estimator of Jacod and Todorov (2014). Section 5 contains some concluding remarks. All technical results, which might be of independent interest, are outlined in Section 6.1, and the detailed proofs are gathered in Section 6.2.

1.1. Notation

For two real numbers a, b, we denote $a \wedge b = \min(a, b)$, $a \vee b = \max(a, b)$. The indicator function of a set A is denoted as $\mathbb{1}_A$. For a function $f = f(a, b, \ldots)$, $\partial_a f$ denotes the partial derivative w.r.t. a, and for a function $f(\theta) \in \mathbb{R}^m$ with $\theta \in \mathbb{R}^k$, the gradient matrix is denoted by $(\mathrm{D}_\theta f)_{j,l} = \partial_{\theta_l} f_j$. For $\delta > 0$, $B_\delta(0)$ is the ball around 0 with radius δ in $\mathbb{R}^k$, where k is evident from the context. $\boldsymbol{I}_d \in \mathbb{R}^{d\times d}$ denotes the identity matrix. The multivariate normal distribution with covariance matrix Σ and mean 0 is denoted as $\mathcal{N}(0, \Sigma)$, and $\Rightarrow$ denotes weak convergence of probability measures resp. random elements. The expectation operator is $\mathbb{E}$, and dependence upon a parameter θ is denoted as $\mathbb{E}_\theta$.

2. Model and estimator

Consider a univariate Lévy process X_t, $X_0 = 0$, with characteristic triplet (μ, σ^2, ν) for a drift parameter $\mu \in \mathbb{R}$, volatility parameter $\sigma^2 > 0$, and a Lévy measure ν, i.e. $\int(1 \wedge |z|^2)\, \nu(dz) < \infty$. We choose an odd truncation function ξ such that $|\xi| \leq 2$ and $\xi(z) = z$ for $z \in (-1, 1)$. Then X_t admits the Lévy-Itô decomposition

$$\begin{aligned} X_t = \mu t + \sigma B_t + \int_0^t \int (z - \xi(z))\, N(dz, ds) \\ + \int_0^t \int \xi(z)\, (N(dz, ds) - \nu(dz) \otimes ds), \end{aligned} \tag{2.1}$$

where $N(dz, ds)$ is a Poisson point process with intensity measure $\nu(dz) \otimes ds$, and B_t is a standard Brownian motion, independent of N. The value of μ depends on the choice of the truncation function ξ, but for our purposes, it will turn out that μ is negligible anyways. To make the approximation (1.3) precise, we suppose that

$$\begin{aligned} |\nu([x, \infty)) - \tilde{\nu}([x, \infty))| &\leq L|x|^{-\rho}, \quad x \in (0, 1], \\ |\nu(-\infty, x]) - \tilde{\nu}(-\infty, x])| &\leq L|x|^{-\rho}, \quad x \in [-1, 0), \end{aligned} \tag{2.2}$$

for some $L > 0$ and $\rho > 0$. The approximating measure $\tilde{\nu}$ is given by the Lebesgue density

$$\tilde{\nu}(dz) = \sum_{m=1}^{M} \frac{\alpha_m}{|z|^{1+\alpha_m}} \left(r_m^+ \mathbb{1}_{z>0} + r_m^- \mathbb{1}_{z<0} \right) dz, \tag{2.3}$$

for some natural number M and parameters $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_M) \in (0,2)^M$, and $\boldsymbol{r} = (r_1^+, r_1^-, \dots, r_M^+, r_M^-) \in \mathbb{R}_{\geq 0}^{2M}$. The remainder term in (2.2) is treated as a nuisance. In particular, this remainder may still consist of infinite activity jumps. Our main result will require $\rho < \alpha_M$, such that the nuisance jumps are in a sense less active than the Lévy measure $\tilde{\nu}$ and asymptotically negligible. The parameters of the modeled part are summarized as

$$\theta = (\sigma^2, \alpha_1, r_1^+, r_1^-, \dots, \alpha_M, r_M^+, r_M^-) \in \Theta \subset \mathbb{R}^{3M+1}. \tag{2.4}$$

where Θ contains all parameter vectors θ as specified, such that additionally

$$\alpha = \alpha_1 > \alpha_2 > \ldots > \alpha_M > \frac{\alpha}{2}, \qquad r_m^+ + r_m^- > 0,\, i = 1, \dots, M, \quad \sigma^2 \geq 0.$$

The value $\alpha = \alpha_1$ is of central importance. In particular, we need to impose the lower bound $\alpha_M > \alpha/2$ to ensure identifiability of the full parameter vector θ, see Aït-Sahalia and Jacod (2012). Note that the definition (2.3) is the same as given by Jacod and Todorov (2016) for the symmetric case.

In the high-frequency sampling setting considered here, we are given n observations X_{ih_n}, $i = 1, \dots, n$ with observation frequency $h_n \to 0$ such that $nh_n = T$ is constant. Without loss of generality, let $T = 1$ and $h = h_n = 1/n$. Equivalently, we observe the n increments $\Delta_{n,i} X = X_{ih_n} - X_{(i-1)h_n} \sim X_{h_n}$, which constitute a triangular array of random variables with iid rows. The law of X_{h_n} is not fully described by the parameters $(\sigma^2, \boldsymbol{r}, \boldsymbol{\alpha})$ due to the remainder in (2.2). Hence, we approximate it by a fully specified Lévy process $\tilde{Z}_t$ with characteristic triplet $(0, \sigma, \tilde{\nu})$. The process $\tilde{Z}_t$ may be represented as

$$\tilde{Z}_t = \sigma B_t + \sum_{m=1}^{M} S_t^m,$$

where B_t, S_t^m, $m = 1, \dots, M$, are independent Lévy processes, B_t is a standard Brownian motion, and the S_t^m are skewed α_m-stable process with Lévy measure $|z|^{-1-\alpha_m}(r_m^+ \mathbb{1}_{z>0} + r_m^- \mathbb{1}_{z<0})$.

We suggest to estimate the parameter θ via the method of moments. In particular, we choose $3M+1$ functions $f_j : \mathbb{R} \to \mathbb{R}$, $\boldsymbol{f} = (f_1, \dots, f_{3M+1})$, and a suitable scaling factor $u = u_n$, and define $\hat{\theta} = \hat{\theta}_n$ to be a solution of the equation

$$F_n(\hat{\theta}_n) = \left[\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{f}(u_n \Delta_{n,i} X) \right] - \mathbb{E}_{\hat{\theta}_n} \boldsymbol{f}(u_n \tilde{Z}_{h_n}) = 0. \tag{2.5}$$

Here and in the following, $\mathbb{E}_\theta f(\tilde{Z}_h)$ denotes the expectation such that $\tilde{Z}_h$ is determined by the parameter vector θ. Since $\tilde{Z}_h$ is a fully parametric approximation of X_h, the function $F_n(\theta)$ can be computed numerically, such that $\hat{\theta}_n$ is a feasible estimator. To distinguish a generic parameter value from the parameters governing X_t, we denote by θ_0 the true parameter such that (2.2) holds.

To study the limit of $\hat{\theta}_n$, we employ the standard framework for estimating equations as reviewed by Jacod and Sørensen (2018). Under the assumptions imposed below, we show that $\hat{\theta}_n - \theta_0 \approx -(\mathrm{D}_\theta F_n(\theta_0))^{-1} F_n(\theta_0)$, up to negligible terms. In order for $\hat{\theta}_n$ to have good asymptotic properties, the choices of the moment functions $\boldsymbol{f}$ and the scaling factor u_n are crucial. In particular, to derive a central limit theorem for $F_n(\theta_0)$ (see Lemma 6.4), we need to control the sampling variance in (2.5) as well as the bias incurred by approximating X_t by $\tilde{Z}_t$. Furthermore, the asymptotic behavior of $\mathrm{D}_\theta F_n(\theta)$ as $n \to \infty$ needs to be treated (see Lemma 6.5). To this end, the following properties turn out to be sufficient.

Condition (F1)**.** For $j = 1, \ldots, 3M+1$, the functions $f_j \in \mathcal{C}^3(\mathbb{R})$ satisfy $\|f_j^{(k)}\|_\infty < \infty$ for $k = 0, 1, 2, 3$, and $f_j' \in L_1(\mathbb{R})$.

The smoothness imposed by Condition F1 is used to bound the bias incurred by approximating $\mathbb{E}\boldsymbol{f}(uX_{h_n})$ by $\mathbb{E}_\theta \boldsymbol{f}(u\tilde{Z}_{h_n})$, see Corollary 6.3 below. To control the sampling variance, we do not only require smoothness of the employed moment functions, but they further need to be of a specific shape.

Condition (F2)**.** The function f_1 is symmetric and satisfies $f_1(0) = f_1'(0) = 0 \neq f_1''(0)$. The functions f_j, $j = 2, \ldots, 3M+1$, are identically zero on the interval $[-\eta, \eta]$ for some $\eta > 0$.

Additional identifiability conditions are specified in assumption I below. The first moment function f_1 is approximately quadratic near zero, and will serve to identify the volatility σ^2. The functions $f_j(x)$ are smooth thresholds, which distinguish the diffusion from the jump component. An example of suitable moment functions is given in section 4. To ensure that the threshold is effective, we require that $u_n X_{h_n} \to 0$ in probability, i.e. $u_n = o(\sqrt{n})$. By choosing an appropriate scaling sequence as follows, the moments $\mathbb{E} f_j(u_n \tilde{Z}_{h_n})$, $j \geq 2$, will be dominated by the jump component.

Condition (U)**.** $u_n \to \infty$ such that $u_n = \frac{\tau\sqrt{n}}{\sqrt{\log n}}$ for some $\tau < \frac{\eta}{\sigma\sqrt{8}}$.

Although potentially not sharp, the upper bound on the factor τ is required to derive our asymptotic result. For details, see the technical Lemma 6.1 below and the subsequent discussion. When choosing u_n in accordance with condition U, it suffices to use a reasonable upper bound on σ. Furthermore, the simulation results presented in section 4 show that larger values of u_n also perform well in finite samples.

Since f_j vanishes on $[-\eta, \eta]$ for $j = 2, \ldots, 3M+1$, the respective sample moments are only driven by increments such that $\Delta_{n,i} X > \frac{\eta}{u_n}$. Hence, these sample moments correspond to a smooth variant of the statistic (1.2) studied by Aït-Sahalia and Jacod (2009), with threshold $\tau_n \propto \sqrt{\log n}/\sqrt{n}$. The same threshold level has been found to be optimal for the detection of jumps (Figueroa-López and Nisen, 2013, Theorem 4.3) and for the estimation of integrated volatility

(Figueroa-López and Mancini, 2019, Proposition 1). Furthermore, the function f_1 is quadratic near zero, and globally bounded. If we choose f_1 to vanish on $[-1,1]^c$, then the rescaled sample moment $u_n^{-2}\sum_{i=1}^n f_1(u_n\Delta_{i,n}X)$ is a smooth variant of the truncated realized variance estimator introduced by Mancini (2006), with threshold $\tau_n = 1/u_n$. Besides the smooth truncation, a major methodological novelty of our estimator $\hat{\theta}_n$ is to combine the threshold-based estimation of jump and diffusion components into a single system of estimating equations.

To formulate our main result on the asymptotic behavior of $\hat{\theta}$, we introduce the quantities

$$\mathcal{J}_\alpha^\pm g(x) = \alpha \int \frac{g(x+z)-g(x)-g'(x)\xi(z)}{|z|^{1+\alpha}} \mathbb{1}_{\{\pm z>0\}}\,dz, \qquad \alpha\in(0,2),$$

which exist if $\|g\|_\infty, \|g''\|_\infty < \infty$. Furthermore, we introduce the matrices

$$\begin{aligned}
\gamma_{n,m}(\theta) &= \begin{pmatrix} 1 & 0 & 0 \\ -r_m^+ \log u_n & 1 & 0 \\ -r_m^- \log u_n & 0 & 1 \end{pmatrix}, \quad m=1,\dots,M, \\
\Gamma_n(\theta) &= \operatorname{diag}(\boldsymbol{I}_1, \gamma_{n,1},\dots,\gamma_{n,M}) \in \mathbb{R}^{(3M+1)\times(3M+1)}, \\
\bar{\Lambda}_n(\theta) &= \sqrt{h_n}\operatorname{diag}(\sqrt{h_n}^{-1}, u_n^{\alpha_1-\frac{\alpha_1}{2}}, u_n^{\alpha_1-\frac{\alpha_1}{2}}, u_n^{\alpha_1-\frac{\alpha_1}{2}}, u_n^{\alpha_2-\frac{\alpha_1}{2}},\dots \\
&\qquad \dots, u_n^{\alpha_{M-1}-\frac{\alpha_1}{2}}, u_n^{\alpha_M-\frac{\alpha_1}{2}}, u_n^{\alpha_M-\frac{\alpha_1}{2}}, u_n^{\alpha_M-\frac{\alpha_1}{2}}),
\end{aligned}$$

and the matrix $A(\theta)\in\mathbb{R}^{(3M+1)\times(3M+1)}$, given by

$$A(\theta)_{1,1} = f_1''(0)/2, \quad A(\theta)_{1,j} = A(\theta)_{j,1} = 0, j\neq 1,$$

and for $m=1,\dots,M$, $j=2,\dots,3M+1$,

$$\begin{aligned}
A(\theta)_{j,3m-1} &= \partial_{\alpha_m}(r_m^+\mathcal{J}_{\alpha_m}^+ f_j(0) + r_m^-\mathcal{J}_{\alpha_m}^- f_j(0)), \\
A(\theta)_{j,3m} &= \mathcal{J}_{\alpha_m}^+ f_j(0), \qquad A(\theta)_{j,3m+1} = \mathcal{J}_{\alpha_m}^- f_j(0).
\end{aligned}$$

These derivatives exist because $\|f\|_\infty, \|f''\|_\infty$ are finite. Finally, we introduce the symmetric positive semidefinite matrix $\Sigma(\theta)$ given by

$$\begin{aligned}
\Sigma(\theta)_{1,1} &= \frac{\sigma^4 f_1''(0)^2}{2}, \\
\Sigma(\theta)_{1,j} &= \Sigma(\theta)_{j,1} = 0, \qquad j\geq 2, \\
\Sigma(\theta)_{j,k} &= \left(r_1^+\mathcal{J}_{\alpha_1}^+ + r_1^-\mathcal{J}_{\alpha_1}^-\right)(f_j\cdot f_k)(0), \qquad j,k\geq 2.
\end{aligned}$$

If clear from the context, we will omit the dependence on θ. Using this notation, we can formulate the remaining identifiability condition.

Condition (I)**.** For the true parameter θ_0, $A(\theta_0)$ is regular.

Remark 1. Analyzing the degrees of freedom of the equation $|A(\theta)| = 0$ suggests that condition I is, in fact, the generic case. To demonstrate this point, we construct a set of moment functions satisfying the identifiability condition. Consider the case $M = 1$ with $\alpha_m = \alpha$ and $r_m^\pm = r^\pm$, $m = 1$. We can construct a set of moment functions satisfying condition I as follows. Let $f_1 = f$ and g be symmetric functions satisfying conditions F1 such that $f_1''(0) \neq 0$, and g vanishes on $[-1, 1]$. Furthermore, denote $a = \mathcal{J}_\alpha^+ g(0) = \mathcal{J}_\alpha^- g(0)$, and $b = \partial_\alpha \mathcal{J}_\alpha^\pm g(0)$. We set $f_2(x) = g(x), f_3(x) = g(2x)$, and $f_4(x) = g(x)\mathbb{1}_{x>0} + g(2x)\mathbb{1}_{x<0}$. Note that $\mathcal{J}^\pm f_3(0) = 2^\alpha \mathcal{J}^\pm g(0) = 2^\alpha a$, as well as $\mathcal{J}^+ f_4(0) = a$, and $\mathcal{J}^- f_4(0) = 2^\alpha a$. Then one can check that

$$\begin{aligned}A(\theta_0) &= A(\sigma^2, r^+, r^-, \alpha)\\ &= \begin{pmatrix} f''(0)/2 & 0 & 0 & 0 \\ 0 & (r^+ + r^-)b & a & a \\ 0 & 2^\alpha(r^+ + r^-)(b + a\log 2) & 2^\alpha a & 2^\alpha a \\ 0 & r^+ b + r^- a 2^\alpha \log 2 & a & a(1 + 2^\alpha) \end{pmatrix},\end{aligned}$$

with determinant $\det(A) = -\frac{f''(0)}{2}(r^+ + r^-)\, a^3\, 2^\alpha \log 2$. Hence, $A(\theta_0)$ is regular for $(r^+ + r^-) > 0$ and all $\alpha \in (0, 2)$ if g is chosen such that $a \neq 0$. This is in particular the case for the choice of the moment functions for the simulation study in section 4.

The main result of this paper is the consistency and asymptotic normality of $\hat{\theta}_n$, as summarized by the following theorem.

Theorem 2.1. *Let X_t be a Lévy process satisfying (2.2) with some $\rho < \alpha/2$, and parameter vector $\theta_0 \in \Theta$. Let $\boldsymbol{f}$ satisfy assumptions F1 and F2, and be such that $A(\theta_0)$ is regular, and let $u_n \to \infty$ be chosen according to U. Then there exists a sequence of random vectors $\hat{\theta}_n$ solving (2.5) eventually, i.e. $P(F_n(\hat{\theta}_n) = 0) \to 1$. This sequence satisfies $\hat{\theta}_n \to \theta$ in probability as $n \to \infty$, and*

$$\sqrt{n} A(\theta_0) \bar{\Lambda}_n \Gamma_n^{-1}(\theta_0)(\hat{\theta}_n - \theta_0) \Rightarrow \mathcal{N}(0, \Sigma(\theta_0)).$$

The resulting rate of convergence for the BG index $\alpha = \alpha_1$ is thus found to be $(n \log n)^{\frac{\alpha}{4}}$, which improves upon existing estimators and matches the lower bound of Aït-Sahalia and Jacod (2012) up to logarithmic factors. However, the rate matrix of Theorem 2.1 is non-diagonal. The phenomenon of a non-diagonal rate matrix has also been observed in the pure jump case, i.e. $\sigma^2 = 0$, see Brouste and Masuda (2018). Interestingly, a non-diagonal rate matrix also occurs when estimating the Hurst parameter of a fractional Brownian motion in the high-frequency setting, see Brouste and Fukasawa (2018). We further discuss this aspect and the resulting marginal rates of convergence for $\hat{\alpha}_m$ and $\hat{r}_m^\pm$ in the next section. Nevertheless, the matrices Γ_n^{-1}, $A(\theta_0)$, and $\Sigma(\theta_0)$ are block-diagonal, such that the volatility estimator $\hat{\sigma}^2$ is asymptotically independent of the estimator of the jump part.

The presented central limit theorem also holds for the fully specified case without nuisance, i.e. $L = 0$ in (2.2). Even in this parametric case, we find that a simple GMM estimator based on $3M + 1$ fixed moment functions, corresponding to $u_n = 1$, will not achieve the best rate of convergence. A careful construction of the estimating equation (2.5) is thus not only required to handle the nuisance term, but also for the underlying parametric problem itself.

The proposed estimator for α can be contrasted with existing methods in the literature. In an earlier study, Reiß (2013) suggests a test procedure for the value of α based on a statistic T_n^m with tuning parameter $m \in \mathbb{N}$. Therein, it is established that $T_n^m \to Q(\alpha)$ as $n \to \infty$ at rate $n^{\frac{\alpha}{4}-\epsilon(m)}$, and $\epsilon(m) \to 0$ as $m \to \infty$. By inverting the function Q, this approach yields a near-optimal estimator for α. The statistics T_n^m are constructed based on nonlinear sample moments as in (2.5), where the f_j are linear combinations of trigonometric functions, i.e. $f_j(x) = \sum_k w_{k,j} \exp(i\lambda_k x)$. Choosing the weights $w_{k,j}$ carefully such that $\sum_k w_{k,j}\lambda_k^{2p} = 0$ for $p = 1, \ldots, m-1$, Reiß (2013) is able to reduce the variance of the corresponding sample moments. The arbitrarily small defect in the rate of convergence $n^{\alpha/4-\epsilon(m)}$ derived therein is thus due to the sampling variance. In contrast, by choosing the moment functions to vanish near zero according to Condition F2, we obtain a smaller variance of the sample moments.

An alternative estimator achieving the rate $n^{\alpha/4-\epsilon}$ is presented by Bull (2016), which also uses functions which vanish near zero. Therein, the value $\mathbb{E}\boldsymbol{f}(u_n X_{h_n})$ is approximated by a finite series expansion, and extending this expansion reduces the rate defect ϵ. In contrast, we use the approximation $\mathbb{E}\boldsymbol{f}(u_n X_{h_n}) \approx \mathbb{E}\boldsymbol{f}(u_n \tilde{Z}_{h_n})$. Although the latter value is not available in explicit form and needs to be determined numerically, this approach allows us to decrease the bias of the estimating equation further than by any finite series expansion. In particular, we only incur a bias due to approximating the Lévy measure of X_t, but not due to a discretization of the time evolution of the process. Thus, our method effectively circumvents the variance issue of Reiß (2013) and the bias issue of Bull (2016). This allows us to eliminate the polynomial rate defect and achieve a faster rate of convergence.

3. Asymptotic optimality

It is natural to ask whether our proposed estimator is asymptotically optimal. From Theorem 2.1, we find that

$$\sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) \Rightarrow \mathcal{N}(0, 2\sigma^4), \tag{3.1}$$

which matches the optimal estimator in the situation without jumps. That is, $\hat{\sigma}_n^2$ is efficient. In general, jumps of infinite variation reduce the achievable rate of convergence for volatility estimators (Jacod and Reiss, 2014). Here, we are able to recover efficiency by modeling the infinite variation part of the jump measure explicitly via (2.2). The same methodology has been applied by Jacod and Todorov (2014, 2016) to construct an efficient estimator of σ^2. Note that the latter studies treat more general types of semimartingales, while we only derived a result for Lévy processes. In contrast to the existing estimators, which use a multi-step debiasing procedure, we determine $\hat{\sigma}^2$ by a single set of estimating equations. While our approach is conceptually simple, solving the estimating equations (2.5) is computationally expensive. A comparison of the finite sample performance is presented in Section 4.

As the asymptotic variance of the estimators α_m and $r_m^{\pm}$ depends on the choice of $\boldsymbol{f}$, they can not be expected to be variance efficient. Furthermore, they are coupled via Γ_n and via the matrix $A(\theta_0)$, which is in general dense. Inspecting the limit in Theorem 2.1, we find that

$$\begin{aligned} \hat{\alpha}_m - \alpha_m &= \mathcal{O}_P\left(u^{\frac{\alpha_1}{2}-\alpha_m}\right) = \mathcal{O}_P\left((n\log n)^{\frac{\alpha_1}{4}-\frac{\alpha_m}{2}}\right), \\ \hat{r}_m^{\pm} - r_m^{\pm} &= \mathcal{O}_P\left(u^{\frac{\alpha_1}{2}-\alpha_m}\log u\right) = \mathcal{O}_P\left((n\log n)^{\frac{\alpha_1}{4}-\frac{\alpha_m}{2}}\log(n)\right). \end{aligned} \tag{3.2}$$

To assess these rates of convergence, we may compare with the lower bound of Aït-Sahalia and Jacod (2012). Therein, the authors compute the diagonal terms of the Fisher information $\mathcal{I}_\theta^n$ based on n observations of $\tilde{Z}_{1/n}$ for the symmetric case $r_m^+ = r_m^- = r_m$ and $M = 2$. Their analysis of the diagonal entries $\mathcal{I}^n_{\alpha_m,\alpha_m}$ and $\mathcal{I}^n_{r_m,r_m}$ suggests that an asymptotically optimal estimator $(\hat{\alpha}_m^*, \hat{r}_m^*)$ should satisfy

$$\begin{aligned} \hat{\alpha}_m^* - \alpha_m &= \mathcal{O}_P\left((n\log n)^{\frac{\alpha_1}{4}-\frac{\alpha_m}{2}}/\log n\right), \\ \hat{r}_m^* - r_m &= \mathcal{O}_P\left((n\log n)^{\frac{\alpha_1}{4}-\frac{\alpha_m}{2}}\right). \end{aligned} \tag{3.3}$$

Notably, even for $M = 1$, the rates (3.3) are faster than (3.2) by a logarithmic factor.

This difference could potentially be explained by the neglected off-diagonal terms of $\mathcal{I}_\theta$. A similar phenomenon occurs in the pure jump case $\sigma^2 = 0$, $M = 1$, where for any sequence of diagonal matrices D_n, the limit of $D_n \mathcal{I}^n_{(\alpha,r)} D_n$ is singular, see (Masuda, 2015, Thm. 3.4) and (Aït-Sahalia and Jacod, 2008, Thm. 2). Recently, Brouste and Masuda (2018) studied this case, and established the LAN property with a non-diagonal rescaling matrix D_n. They find that the optimal rate of convergence is slower than suggested by the diagonal entries of the Fisher matrix, by a factor of $\log n$. A similar phenomenon is observed when estimating the Hurst parameter of a fractional Brownian motion based on high-frequency observations (Brouste and Fukasawa, 2018). There is no LAN result available for estimation of the BG index in the case $\sigma^2 > 0$, and a full investigation of the LAN property in the present case is out of scope of this paper. Nevertheless, we can adapt the proof of Aït-Sahalia and Jacod (2012) to unveil the off-diagonal entries $\mathcal{I}^n_{\alpha_1,r_1}$. It turns out that the diagonally rescaled Fisher matrix is asymptotically singular, just as in the pure-jump case.

Proposition 3.1. *Let $\mathcal{I}^h$ denote the Fisher information matrix of $\tilde{Z}_h$ with $M = 1$ and $\alpha_1 = \alpha$, $r_1^+ = r_1^- = r$. Then, as $h \to 0$,*

$$\frac{(h\log(1/h))^{\frac{\alpha}{2}}}{h}\begin{pmatrix}1 & 0\\ 0 & \frac{1}{\log(1/h)}\end{pmatrix}\begin{pmatrix}\mathcal{I}_h^{r,r} & \mathcal{I}_h^{r,\alpha}\\ \mathcal{I}_h^{r,\alpha} & \mathcal{I}_h^{\alpha,\alpha}\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & \frac{1}{\log(1/h)}\end{pmatrix} \longrightarrow \frac{2r}{\sigma^\alpha(2-\alpha)^{\frac{\alpha}{2}}}\begin{pmatrix}\frac{1}{r^2} & \frac{1}{2r}\\ \frac{1}{2r} & \frac{1}{4}\end{pmatrix}.$$

In particular, the limiting matrix is singular.

The diagonal entries of the Fisher information matrix should match the optimal rates of convergence in the case where only a single parameter is unknown, e.g. if (σ^2, r_1^+, r_1^-) are known and α_1 should be estimated. In this situation, a natural version of our estimator is to consider only a single moment function f. Analogous to (2.5), for any $m \in \{1, \ldots, M\}$, we may estimate α_m as the solution of

$$\tilde{F}_n(\alpha_m) = \frac{1}{n} \sum_{i=1}^{n} f(u_n \Delta_{n,i} X) - \mathbb{E}_\theta f(u_n \tilde{Z}_h) \stackrel{!}{=} 0. \tag{3.4}$$

With a slight abuse of notation, we may also estimate $r_m^\pm$ by the equation $\tilde{F}_n(r_m^\pm) = 0$. To distinguish jumps and diffusion, we suppose f satisfies the same conditions as $f_2, \ldots, f_{3M+1}$, i.e. it should vanish around zero.

Proposition 3.2. *Let X_t be a Lévy process satisfying (2.2) with some $\rho < \alpha_1/2$, and parameter vector $\theta_0 \in \Theta$. Let f be a non-negative function satisfying F1, and $f(x) = 0$ for $x \in [-\eta, \eta]$, and choose $u_n \to \infty$ such that U holds. Fix some $m \in \{1, \ldots, M\}$, and suppose that $\mathcal{J}_{\alpha_m}^\pm f(0) > 0$. Then there exists a consistent sequence of estimators $\hat{\alpha}_m$ satisfying $P(\tilde{F}_n(\hat{\alpha}_m) = 0) \to 1$, such that $\hat{\alpha}_m \to \alpha_m$ in probability as $n \to \infty$, and*

$$u_n^{\alpha_m - \frac{\alpha_1}{2}} \log(u_n) \left(\hat{\alpha}_m - \alpha_m\right) \Rightarrow \mathcal{N}\left(0, \frac{(r_1^+ \mathcal{J}_{\alpha_1} + r_1^- \mathcal{J}_{\alpha_1}) f^2(0)}{(r_m^+ \mathcal{J}_{\alpha_m}^+ + r_m^- \mathcal{J}_{\alpha_m}^-) f(0)}\right).$$

Under the same conditions, and if all parameters except for r_m^+ resp. r_m^- are known, there exists a consistent sequence of estimators $\hat{r}_m^\pm$ solving the estimating equation eventually, i.e. $P(\tilde{F}_n(\hat{r}_m^\pm) = 0) \to 1$, such that, as $n \to \infty$,

$$u_n^{\alpha_m - \frac{\alpha_1}{2}} \left(\hat{r}_m^\pm - r_m^\pm\right) \Rightarrow \mathcal{N}\left(0, \frac{(r_1^+ \mathcal{J}_{\alpha_1} + r_m^- \mathcal{J}_{\alpha_1}) f^2(0)}{\mathcal{J}_{\alpha_m}^\pm f(0)}\right).$$

Since u_n is of order $\sqrt{n/\log n}$, Proposition 3.2 establishes precisely the rates (3.3). In the setting of Aït-Sahalia and Jacod (2012), in particular $M = 2$, this shows that $\hat{\alpha}_m$ resp. $\hat{r}_m^\pm$ are rate efficient if the remaining parameters θ are known. In contrast, if all parameters θ are unknown, $\hat{\theta}$ achieves the optimal rate of convergence, up to a logarithmic factor. Due to the singularity of the Fisher matrix, we conjecture that the achieved rates (3.2) are in fact optimal.

4. Simulation study

By means of a Monte Carlo study, we compare the finite sample performance of our estimator with the estimators of Reiß (2013) and Bull (2016) for the Blumenthal-Getoor index α, and with the volatility estimator of Jacod and Todorov (2014). The code is available as supplemental material (Mies, 2020). For our simulations, we sample paths of a Lévy process X_t given by

$$X_t = B_t + S_t^{\alpha,\beta} + 0.1 S_t^{0.5,0}. \tag{4.1}$$

We denote by $S_t^{\alpha,\beta}$ the α-stable Lévy motion with skewness parameter $\beta \in (-1,1)$. That is, the characteristic function of $S_t^{\alpha,\beta}$ is given by (see e.g. Zolotarev (1986))

$$\log \mathbb{E} \exp(\boldsymbol{i}\lambda S_t^{\alpha,\beta}) = -t|\lambda|^\alpha \left[1 - \boldsymbol{i} \tan\left(\frac{\pi\alpha}{2}\right) \beta \operatorname{sign}(\lambda)\right].$$

The Lévy measure corresponding to this standardization can be expressed in the form (2.3) with $M=1$, $\frac{r^+ - r^-}{r^+ + r^-} = \beta$, and $(r^+ + r^-) = \frac{1}{\Gamma(1-\alpha)\cos(\pi\alpha/2)}$ if $\alpha \neq 1$. For $\alpha = 1$, the characteristic function may be found as continuous extension. Here, we will set $\beta = -1/3$ and study the cases $\alpha = 1$, $\alpha = 1.3$ and $\alpha = 1.7$. Then (2.2) is satisfied with $\rho = 0.5$, such that $S_t^{0.5,0}$ is a nuisance term, and $\tilde{Z}_t = B_t + S_t^{\alpha,\beta}$. Note that we have $\rho < \frac{\alpha}{2}$ in the cases $\alpha = 1.3$ and $\alpha = 1.7$, such that the nuisance is asymptotically negligible, while we have $\rho = \frac{\alpha}{2}$ for $\alpha = 1$. In view of applications in financial econometrics, we consider the time horizon $T = 1$, and sampling frequencies h between $h = 0.2/23400$ and $h = 60/23400$, corresponding to 0.2 resp. 60 seconds per quote on a trading day of 6.5 hours.

We consider the estimating equation (2.5) for $M = 1$, i.e. $\theta = (\sigma^2, \alpha, r^+, r^-) \in \Theta$. To treat the equation numerically, we compute the moments and its gradients by the Fourier transform $\mathbb{E} f(u\tilde{Z}_h) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\lambda) \mathbb{E}\exp(\boldsymbol{i}\lambda u \tilde{Z}_h)\, d\lambda$, where $\hat{f}(\lambda) = \int_{-\infty}^{\infty} f(x)\exp(-\boldsymbol{i}\lambda x)\, dx$ is the continuous Fourier transform of a function f. Note that $\mathbb{E}\exp(\boldsymbol{i}\lambda\tilde{Z}_h)$ is available in closed form, see e.g. Zolotarev (1986). Using the representation (2.1), we have

$$\begin{aligned}
&\log \mathbb{E}\exp(\boldsymbol{i}\lambda\tilde{Z}_h) \\
&= -h\lambda^2\frac{\sigma^2}{2} - h\boldsymbol{i}\lambda\frac{r_+ - r_-}{1-\alpha} \\
&\qquad - h|\lambda|^\alpha \Gamma(1-\alpha)(r_+ + r_-)\cos\left(\frac{\pi\alpha}{2}\right)\left[1 - \boldsymbol{i}\tan\left(\frac{\pi\alpha}{2}\right)\frac{r_+ - r_-}{r_+ + r_-}\operatorname{sign}(\lambda)\right].
\end{aligned}$$

The drift term in the characteristic exponent depends on the choice of the truncation function $\xi(z)$ which is employed in definition (2.1). This representation of the characteristic exponent is valid for $\xi(z) = z\mathbb{1}_{|z|\leq 1}$. Although there is no canonical specification of the drift of a general Lévy process, this does not impede our estimation procedure since we consider the drift as a nuisance parameter. The employed moment functions $f_1, \ldots, f_4$ are handcrafted to satisfy F1 and F2. In our simulations, we use

$$\begin{aligned}
f_1(x) &= 1 - \exp\left(-\frac{x^2}{2}\right), \\
f_2(x) &= \exp\left(-\frac{300}{(|x|-1)\vee 0}\cdot\exp\left(-\frac{10}{(8-|x|)\vee 0}\right)\right), \\
f_3(x) &= \exp\left(-\frac{300}{(|2x|-1)\vee 0}\cdot\exp\left(-\frac{10}{(8-|2x|)\vee 0}\right)\right), \\
f_4(x) &= \begin{cases} \exp\left(-\frac{300}{(|2x|-1)\vee 0}\cdot\exp\left(-\frac{10}{(8-|2x|)\vee 0}\right)\right), & x \geq 0, \\ \exp\left(-\frac{300}{(|x|-1)\vee 0}\cdot\exp\left(-\frac{10}{(8-|x|)\vee 0}\right)\right), & x < 0. \end{cases}
\end{aligned} \tag{4.2}$$

Note that f_2, f_3, f_4 vanish on $[-1/2, 1/2]$. We use the rescaling factors $u = 1/\sqrt{h|\log h|}$ and $u = 2/\sqrt{h|\log h|}$. Although this choice of u is too large to comply with assumption U, we found it to perform better than smaller values for the given sampling scenario.

Theorem 2.1 guarantees that the estimating equation (2.5) is feasible eventually, as $n \to \infty$. However, depending on the realization of X_t, it might occur for fixed n that there is no parameter value $\theta \in \Theta$ such that $F_n(\theta) = 0$. In order to obtain an estimator in the latter situation, we replace the estimating equation by the equivalent least-squares minimization problem $\hat{\theta} = \arg\min_{\theta\in\Theta} \|W_n^{-1} F_n(\theta)\|^2$, with diagonal weight matrix W_n given by $W_n = \operatorname{diag}(\frac{1}{n}\sum_{i=1}^n \boldsymbol{f}(u_n \Delta_{i,n} X))$. The minimizer is determined numerically, using the methods LBFGS and MMA of the NLopt library (Johnson, 2020), accessed via its R interface. These procedures allow for box constraints on the parameter θ, and we ensure that $\alpha \in (0, 2)$. As an initial value, we choose $0.5 \cdot \theta_0$, which is intentionally distinct from the unknown true parameter value.

The estimators $\hat{\alpha}_{\text{Reiß}}$ of Reiß (2013) and $\hat{\alpha}_{\text{Bull}}$ of Bull (2016) each have a tuning parameter $m \in \mathbb{N}$, and larger values of m increase the rate of convergence. However, smaller values of m can be superior in finite samples. In our simulations, we found that the estimator of Bull performed best when setting $m = 2$ across all simulated frequencies. For the estimator of Reiß, setting $m = 2$ gave the best results for almost all frequencies. The procedure of Bull requires a scaling parameter $\tau_n = Cn^{\frac{m}{2(m+1)}}$. We found the choice $C = 0.5$ to yield good results. Furthermore, the method of Reiß involves a rescaling parameter U_n and two weighting measures w_1, w_2. We choose the weighting measure w_1 to be supported on the set $\{1/m, 2/m, \dots, 1\}$, and w_2 to be supported on the set $\{2/m, 4/m, \dots, 2\}$. The truncation parameter is set to $U = h^{-(1-2m)/(4m-1)}$, as suggested by equation (3.8) in the original article (Reiß, 2013). Finally, the estimator $\hat{\sigma}^2_{\text{JT}}$ of Jacod and Todorov (2014) is implemented as in equation (5.3) therein, with $\zeta = 1.5$ and $u = |\log h|^{\frac{1}{30}}$.

In Table 1, we compare the simulated performance of our moment estimator for α and σ^2 with the estimators of Jacod and Todorov (2014), Reiß (2013), and Bull (2016). It is found that upon choosing $u = 2/\sqrt{h|\log h|}$, the new estimators perform best in the considered setting. However, it should be noted that all benchmarked methods require various tuning parameters. Most notably, all methods require some scaling factors. Furthermore, our new estimator depends on the employed moment functions f_j. It is thus possible that a very careful choice of these parameters might affect the ranking implied by Table 1.

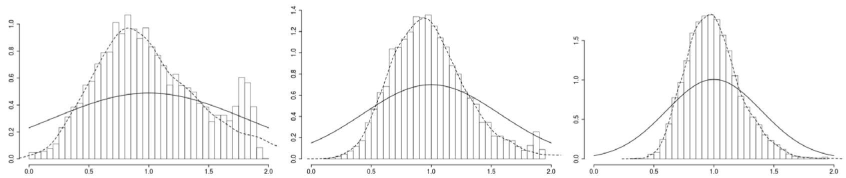
Table 1 also presents the fraction of simulations for which the estimating equations are numerically infeasible, i.e. for which the numerically obtained least-squares minimizer $\hat{\theta}$ satisfies $\|W_n^{-1} F_n(\hat{\theta})\| > 10^{-3}$. As h decreases, we observe that the fraction of feasible equations increases, which is in line with our theoretical results. Moreover, the probability of the estimating equation being feasible is mostly higher for the smaller value of u, even though the larger u leads to a better performance of the estimator $\hat{\theta}$.

TABLE 1
Median absolute errors for the estimation of α and σ^2 in model (4.1), for different estimators, and the proportion of infeasible systems of estimating equations. All values are based on 5000 simulations.

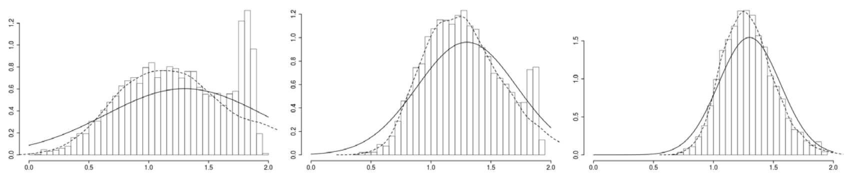
		GMM, $u = \frac{1}{\sqrt{h\|\log h\|}}$			GMM, $u = \frac{2}{\sqrt{h\|\log h\|}}$					
α	h	$\hat{\alpha}$	$\hat{\sigma}^2$	infeas.	$\hat{\alpha}$	$\hat{\sigma}^2$	infeas.	$\hat{\sigma}^2_{\mathrm{JT}}$	$\hat{\alpha}_{\mathrm{Reiß}}$	$\hat{\alpha}_{\mathrm{Bull}}$
1.0	60/23400	0.50	0.24	0.33	0.34	0.10	0.95	0.13	0.65	0.71
1.0	15/23400	0.39	0.10	0.18	0.26	0.05	0.65	0.06	0.45	0.41
1.0	5/23400	0.31	0.05	0.10	0.19	0.02	0.24	0.03	0.33	0.29
1.0	1/23400	0.21	0.02	0.02	0.15	0.01	0.01	0.01	0.22	0.21
1.0	0.2/23400	0.15	0.01	0.00	0.10	0.00	0.00	0.01	0.16	0.16
1.3	60/23400	0.49	0.38	0.43	0.29	0.13	0.99	0.22	0.65	0.79
1.3	15/23400	0.43	0.19	0.30	0.24	0.08	0.90	0.12	0.40	0.54
1.3	5/23400	0.37	0.11	0.21	0.20	0.05	0.57	0.07	0.28	0.36
1.3	1/23400	0.24	0.05	0.08	0.14	0.02	0.05	0.03	0.17	0.24
1.3	0.2/23400	0.14	0.02	0.01	0.09	0.01	0.00	0.01	0.10	0.16
1.7	60/23400	0.27	0.91	0.54	0.09	0.19	1.00	0.75	0.30	1.59
1.7	15/23400	0.22	0.70	0.41	0.05	0.14	1.00	0.59	0.30	0.96
1.7	5/23400	0.18	0.56	0.38	0.04	0.11	0.99	0.43	0.22	0.33
1.7	1/23400	0.16	0.35	0.31	0.06	0.10	0.96	0.22	0.11	0.23
1.7	0.2/23400	0.12	0.19	0.21	0.03	0.05	0.82	0.10	0.06	0.14

The relevance of the scaling factor u may also demonstrated by comparing the simulated distribution of $\hat{\alpha}$ to the asymptotic normal distribution according to Theorem 2.1. For the smaller value $u = 1/\sqrt{h|\log h|}$ (Figure 1), we find that the asymptotic distribution is a good approximation in the cases $\alpha = 1.3$ and $\alpha = 1.7$. Note that the prominent mode near 2 for the case $\alpha = 1.7$ is due to the fact the we restrict the nonlinear minimization to the interval $(0, 2)$. In the case $\alpha = 1$, we find that the simulated distribution of $\hat{\alpha}$ is close to Gaussian, but the convergence of the variance is rather slow. For the larger scaling factor $u = 2/\sqrt{h|\log h|}$ (Figure 2), the asymptotic distribution does not accurately describe the behavior of $\hat{\alpha}$ for $\alpha = 1.7$, as the estimator $\hat{\alpha}$ admits a higher concentration around the true value than suggested by the limit distribution. This observation can be traced back to the nonlinearity of the estimating equation. In particular, based on the estimator $\hat{\theta}$ obtained as a numerical solution of the least-squares problem $\min_\theta \|W_n^{-1} F_n(\theta)\|^2$, we may perform a single Newton iteration to obtain the estimator $\hat{\theta}^* = -(DF_n(\hat{\theta}))^{-1} F_n(\hat{\theta})$. The empirical distribution of the corresponding estimator $\hat{\alpha}^*$ is represented by the dashed line in Figure 1 and 2. Especially in the case $\alpha = 1.7$, we observe that the behavior of $\hat{\alpha}^*$ is well-represented by the asymptotic normal distribution, although the estimation error of $\hat{\alpha}^*$ is larger than that of $\hat{\alpha}$.

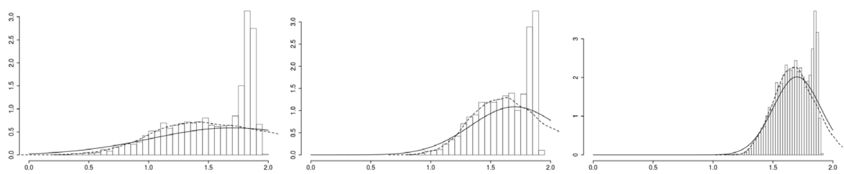
Turning to the estimation of volatility, we have demonstrated that the estimator $\hat{\sigma}^2$ is efficient, and its asymptotic variance depends neither on u nor on $\alpha \in (0, 2)$. However, Table 1 reveals a dependence on both quantities in the simulated finite samples, such that the asymptotic distribution can not be expected to yield a satisfactory approximations in this regime. This defect holds for our proposed estimator as well as for the benchmark method of Jacod and Todorov



(a) $\alpha = 1$, $h = \frac{5}{23400}, \frac{1}{23400}, \frac{0.2}{23400}$ (from left to right)



(b) $\alpha = 1.3$, $h = \frac{5}{23400}, \frac{1}{23400}, \frac{0.2}{23400}$ (from left to right)



(c) $\alpha = 1.7$, $h = \frac{5}{23400}, \frac{1}{23400}, \frac{0.2}{23400}$ (from left to right)

FIG 1. *Simulated distributions of the proposed estimator $\hat{\alpha}$ (histogram) and of the linearized estimator $\hat{\alpha}^*$ (dashed) for $u = 1/\sqrt{h \log |h|}$, based on* 5000 *simulations, and the asymptotic Gaussian distribution as suggested by Theorem 2.1 (solid).*

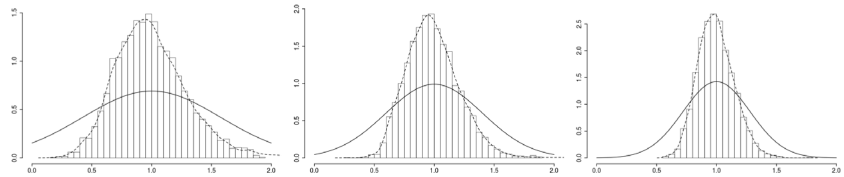
(2014), and it is bigger for large values of α.

The increments of the process (4.1) have infinite variance, which might not be realistic for applications in financial econometrics. Moreover, the jump component is rather large in comparison to the Brownian motion. Therefore, we repeat our analysis for the process

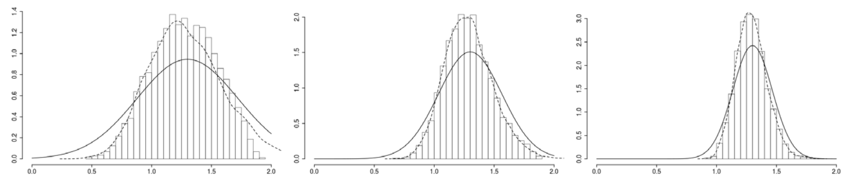
$$Y_t = 0.05\sqrt{\tfrac{2}{3}}B_t + 0.05\sqrt{\tfrac{1}{3}}J_t, \tag{4.3}$$

where J_t is a normal inverse Gaussian Lévy process. The Lévy process J_t has Blumenthal–Getoor index $\alpha = 1$ with proportionality factors $r^+ = r^- = \frac{1}{\pi}$, and its Lévy measure has exponentially decaying tails, see (Cont and Tankov, 2004, Sec. 4.4). Moreover, J_t is standardized such that $\mathrm{Var}(J_1) = 1$. The normal inverse Gaussian process has previously been applied for financial modeling, e.g. by Rydberg (1997) and Barndorff-Nielsen (1997).

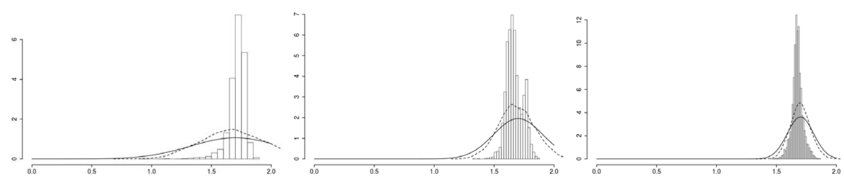
With the same tuning parameters as specified above, the estimators $\hat{\alpha}_{\text{Reiß}}$ and $\hat{\alpha}_{\text{Bull}}$ do not yield satisfactory results for the process Y_t, and we choose not to perform an exhaustive search for optimal tuning parameters. Hence, in Table 2, we only report the performance of the remaining estimators. For the estimating equations, we choose $u = 30/\sqrt{h|\log(h)|}$. The proposed estimator $\hat{\alpha}$ is found



(a) $\alpha = 1$, $h = \frac{5}{23400}, \frac{1}{23400}, \frac{0.2}{23400}$ (from left to right)



(b) $\alpha = 1.3$, $h = \frac{5}{23400}, \frac{1}{23400}, \frac{0.2}{23400}$ (from left to right)



(c) $\alpha = 1.7$, $h = \frac{5}{23400}, \frac{1}{23400}, \frac{0.2}{23400}$ (from left to right)

FIG 2. *Simulated distributions of the proposed estimator $\hat{\alpha}$ (histogram) and of the linearized estimator $\hat{\alpha}^*$ (dashed) for $u = 2/\sqrt{h \log |h|}$, based on 5000 simulations, and the asymptotic Gaussian distribution as suggested by Theorem 2.1 (solid).*

TABLE 2
Median absolute errors for the estimation of α and σ^2 in model (4.3), for different estimators. Errors for $\hat{\sigma}^2$ and $\hat{\sigma}^2_{JT}$ are scaled by $1/\sigma^2$. All values are based on 5000 simulations.

h	$\hat{\alpha}$	$\frac{\hat{\sigma}^2}{\sigma^2}$	infeasible	$\frac{\hat{\sigma}^2_{\mathrm{JT}}}{\sigma^2}$
60/23400	0.89	0.95	0.99	0.26
15/23400	0.86	0.14	0.88	0.24
5/23400	0.56	0.05	0.67	0.24
1/23400	0.33	0.01	0.27	0.17
0.2/23400	0.21	.005	0.05	0.09

to be consistent, and the probability of the estimating equations being feasible increases as the h tends to zero. Furthermore, the new volatility estimator $\hat{\sigma}^2$ is found to yield smaller errors than the benchmark $\hat{\sigma}^2_{\mathrm{JT}}$ in this situation.

5. Conclusion

The method of moments is a standard technique for estimation in parametric models. The results of this paper show that the theory of estimating equations is also applicable in the non-standard setting of high-frequency observations of a Lévy process, even if the jump measure is specified semi-parametrically. Upon careful choice of the nonlinear moment functions, the resulting estimator for the Blumenthal-Getoor index is shown to be rate-optimal up to logarithmic factors, and is conjectured to be in fact rate-optimal. The corresponding volatility estimator is efficient, even in the presence of jumps of infinite variation, and may serve as an alternative to existing efficient estimators. An appealing feature of our approach is that the jumps and the diffusion component of the process are considered jointly by a single set of estimating equations.

While our theoretical results allow for a broad range of possible nonlinearities $\boldsymbol{f}$ and scaling factors u, choosing a specific estimator may be difficult in practice. Our numerical results indicate that the scaling factor u has a non-trivial influence on the error of estimation as well as the accuracy of distributional approximations. Hence, future work might investigate the optimal choice of these tuning parameters. A natural solution would be to replace the moment equations by maximum likelihood estimation. However, a major obstacle is the fact that no closed expression for the likelihood function of the process is available. This aspect also impedes the derivation of an exact statistical lower bound via the theory of local asymptotic normality.

As Lévy processes may be regarded as prototypical semimartingales, we expect that our methodology could be extended to more complex, non-stationary models which are common in financial mathematics. For example, the proposed estimators could be applied locally to obtain an estimator of the corresponding spot characteristics of the process. Making this intuition mathematically precise requires further analysis.

Financial time-series of asset prices at high frequencies are often influenced by so-called market microstructure, which is often modeled as measurement error. Unreported simulations show that our estimators are not robust against microstructure noise. In practice, the statistical effect of market microstructure may be reduced by considering lower frequencies, e.g. 1 minute or more per price. However, the simulation results of Section 4 show that the estimation of the BG index α incurs large errors in this regime, which is due to the rather slow rate of convergence intrinsic to the problem of estimating the BG index in the presence of a diffusion term. This highlights the need for an explicit handling of observational noise, as studied by Jing, Kong and Liu (2011). To the best of our knowledge, statistical lower bounds for the estimation of the BG index under noisy observations are not known.

6. Technical tools

In this section, we present the proofs of Theorem 2.1 and Propositions 3.1 and 3.2. Preliminary technical results are presented in Subsection 6.1, as they might be of independent interest, in particular Lemma 6.1 and Corollary 6.3. The detailed proofs are presented in Subsection 6.2.

6.1. Preliminary results

To study the asymptotic behavior of the estimating equation (2.5) by standard techniques (see e.g. Jacod and Sørensen (2018)), we need

- a central limit theorem for the term $\frac{1}{n}\sum_{i=1}^n \boldsymbol{f}(u_n\Delta_{n,i}X) - \mathbb{E}_\theta \boldsymbol{f}(u_n\tilde{Z}_h)$, and
- properties of the derivatives $\mathrm{D}_\theta \mathbb{E}_\theta \boldsymbol{f}(u_n\tilde{Z}_h)$.

To determine asymptotic variances, as well as for some technical steps of the following proofs, it is useful to derive some explicit approximations of $\mathbb{E}\boldsymbol{f}(u_n\tilde{Z}_h)$.

Lemma 6.1. *Let $f \in \mathcal{C}^2$ be such that f, f' and f'' are bounded and $f(0) = 0$, and let $\tilde{X}_t$ be a Lévy process with characteristic triplet $(\mu, \sigma^2, \tilde{\nu})$. Let $u = u_t$ be a sequence of real numbers such that $u_t \to \infty$ as $t \to 0$.*

(i) If $f(x) = 0$ for $|x| \leq \eta$, then for any $\lambda \in (0,1)$ such that $u \leq \frac{(1-\lambda)\eta}{\sigma\sqrt{8t|\log t|}}$, as $t \to 0$,

$$\mathbb{E}f(u\tilde{X}_t) = o(tu^\alpha) + tu^\alpha \left[r_1^+ \tilde{\mathcal{J}}_\alpha^+ f(0) + r_1^- \tilde{\mathcal{J}}_\alpha^- f(0)\right],$$

where

$$\tilde{\mathcal{J}}_\alpha^\pm f(x) = \int \alpha \mathbb{1}_{\pm z>0} \frac{\left[f(x+z) - f(x) - f'(x)\mathbb{1}_{|z|\leq 1}\right]}{|z|^{1+\alpha}}\, \tilde{\nu}(dz).$$

(ii) If, alternatively, $f(0) = 0$ but $f''(0) \neq 0$, then for any $u = o(1/\sqrt{t})$

$$\mathbb{E}f(u\tilde{X}_t) = tu^2 \frac{\sigma^2}{2} f''(0) + o(tu^2).$$

(iii) If $f(0) = 0, f''(0) = 0$ but $f^{(4)} \neq 0$, and $f^{(3)}, f^{(4)}$ are bounded, then for any $u = o(1/\sqrt{t})$

$$\mathbb{E}f(u\tilde{X}_t) = t^2u^4 \frac{\sigma^4}{8} f^{(4)}(0) + o(t^2u^4) + \mathcal{O}(tu^\alpha).$$

(iv) If $f(0) = 0$ and $\mu = 0$, $\sigma^2 = 0$, then there exists a constant $\tilde{C}$ bounded uniformly on compacts, such that for all f and all $u > 1$, $t \geq 0$,

$$\mathbb{E}f(u\tilde{X}_t) \leq tu^{\alpha\vee 1}(1 + \log(u))\left(\|f\|_\infty + \|f'\|_\infty + \|f''\|_\infty\right)\tilde{C}.$$

All $\mathcal{O}(\cdot)$ and $o(\cdot)$ terms are bounded resp. vanishing uniformly on compacts in Θ.

The case (i), which is exploited in the proofs several times, imposes a subtle upper bound on u. Although this bound need not be sharp, the Lemma will not hold for $u = \tau/\sqrt{t|\log t|}$ if τ is too large. To make this plausible, note that for an α-stable process S_t^α, the probability $P(|S_t^\alpha| \geq \eta\sqrt{t|\log t|}/\tau)$ tends to zero as $t \to 0$, roughly polynomially in t. On the other hand, for the Brownian motion,

$P(|B_t| > \eta\sqrt{t|\log t|}/\tau) = P(|B_1| > \eta\sqrt{|\log t|}/\tau) \to 0$ polynomially as well, but the polynomial order of this decay will depend on the specific value of τ. For the jump term to dominate, as in case (i) of Lemma 6.1, τ must be small. The uniformity w.r.t. θ of the previous results will be used later on to derive the consistency of the estimator.

Another ingredient to obtain a central limit theorem is a bias bound, i.e. a bound on the error of approximating $\mathbb{E}\boldsymbol{f}(u_n\Delta_{n,i}X)$ by $\mathbb{E}_\theta\boldsymbol{f}(u_n\tilde{Z}_h)$. For two random variables X and Y, recall the definition of the 1-Wasserstein metric d_W and the total variation distance d_{TV} given by

$$d_{TV}(X,Y) = \sup_{g:\|g\|_\infty\leq 1} |\mathbb{E}g(X) - \mathbb{E}g(Y)|,$$
$$d_W(X,Y) = \sup_{g:\|g'\|_\infty\leq 1} |\mathbb{E}g(X) - \mathbb{E}g(Y)|,$$

where the supremum is taken over all bounded resp. Lipschitz continuous, measurable functions $g:\mathbb{R}\to\mathbb{R}$. These distances are used in the proof of the following Lemma, which quantifies the error of approximation implied by the local stability assumption (2.2).

Lemma 6.2. *Let $X_t, \tilde{X}_t$ be two Lévy processes with characteristic triplets given by (μ,σ^2,ν) and $(\mu,\sigma^2,\tilde{\nu})$, respectively. Suppose furthermore that for some $\rho\in(0,1\wedge\alpha)$,*

$$|\nu((z,\infty)) - \tilde{\nu}((z,\infty))| \leq L|z|^{-\rho}, \quad z\in(0,1),$$
$$|\nu((-\infty,z)) - \tilde{\nu}((-\infty,z))| \leq L|z|^{-\rho}, \quad z\in(-1,0).$$

There exists a constant $\tilde{C}$ depending on L, ρ, and θ, such that for any differentiable function $f:\mathbb{R}\to\mathbb{R}$, and all $u\geq 1$,

$$\left|\mathbb{E}f(uX_t) - \mathbb{E}f(u\tilde{X}_t - ut\bar{\zeta})\right| \leq \tilde{C}(\|f\|_\infty + \|f'\|_\infty + \|f'\|_{L_1})(tu^\rho + t^2u^{\alpha+1}), \tag{6.1}$$

where $\bar{\zeta} = \int \xi(z)(\nu-\tilde{\nu})(dz) \in \mathbb{R}$. The constant $\tilde{C}$ is bounded on compacts in $\theta\in\Theta$, $\rho\in(0,1\wedge\alpha)$, and $L\geq 0$.

Corollary 6.3. *Let $f\in\mathcal{C}^3$ such that f,f',f'',f''' are bounded and $f'\in L_1$, and suppose that there exists $\eta>0$ such that $f(x)=f(-x)$ for $|x|\leq\eta$. Let $X_t,\tilde{X}_t$ be two Lévy processes with characteristic triplets (μ,σ^2,ν) and $(0,\sigma^2,\tilde{\nu})$, respectively. Suppose that ν, $\tilde{\nu}$ satisfy the conditions of Lemma 6.2. Then, for any $\lambda\in(0,1)$, and for any sequence $u=u_t\to\infty$ as $t\to 0$, such that $u\leq\frac{(1-\lambda)\eta}{\sigma\sqrt{8t|\log t|}}$,*

$$|\mathbb{E}f(uX_t) - \mathbb{E}f(u\tilde{X}_t)| \leq \tilde{C}\left(tu^\rho + t^2u^{2\vee(\alpha+1)}\right)(1+\log(u))). \tag{6.2}$$

The constant $\tilde{C}$ is bounded on compacts in $\mu\in\mathbb{R}$, $\theta\in\Theta$, $\rho\in(0,1\wedge\alpha)$, and $L\geq 0$.

Note that the presented result of 6.3 can not be directly formulated in terms of d_{TV} or d_W, distinguishing it from the results of Mariucci and Reiß (2018). An alternative bound on the total variation distance between X_t and $\tilde{Z}_t$ is presented by (Clément and Gloter, 2019, Proposition 4) and (Amorino and Gloter, 2020b, Proposition 2), stating that $d_{TV}(X_t, \tilde{Z}_t) \le C t^{1\wedge\frac{1}{\alpha}} \log(t)$ as $t \to 0$. Their assumptions on the Lévy measure $\nu(dz)$ imply that our condition (2.2) holds, with $\rho \le (\alpha - 1) \vee 0$. Thus, if $\alpha > 1$ and $u \ll t^{-1/2}$, our bound (6.2) is sharper since $tu^{\alpha-1} \ll t^{\frac{3}{2}-\frac{\alpha}{2}} \ll t^{\frac{1}{\alpha}}$. In the case $\alpha \le 1$, our bound is of the same order of magnitude as the one presented by Clément and Gloter (2019) and Amorino and Gloter (2020b). Furthermore, our result may also be applied in the case $\rho > \alpha - 1$. However, we impose additional smoothness assumptions upon the considered function f, which is suitable for our statistical purposes because the moment functions are chosen by the statistician.

To state the remaining technical results, introduce the notation

$$\begin{aligned}\Lambda_n(\theta) &= \operatorname{diag}(hu^2, hu^{\alpha_1}, hu^{\alpha_1}, hu^{\alpha_1}, \ldots \\ &\qquad \ldots, hu^{\alpha_M}, hu^{\alpha_M}, hu^{\alpha_M}) \in \mathbb{R}^{(3M+1)\times(3M+1)}, \\ \tilde{\Lambda}_n(\theta) &= \operatorname{diag}(hu^2, \sqrt{hu^{\alpha_1}}, \ldots, \sqrt{hu^{\alpha_1}}) \in \mathbb{R}^{(3M+1)\times(3M+1)},\end{aligned}$$

such that

$$\begin{aligned}\bar{\Lambda}_n(\theta) &= \tilde{\Lambda}_n^{-1}(\theta)\Lambda_n(\theta) \\ &= \sqrt{h}\,\operatorname{diag}(\sqrt{h}^{-1}, u^{\alpha_1-\frac{\alpha_1}{2}}, u^{\alpha_1-\frac{\alpha_1}{2}}, u^{\alpha_1-\frac{\alpha_1}{2}}, \ldots \\ &\qquad \ldots, u^{\alpha_M-\frac{\alpha_1}{2}}, u^{\alpha_M-\frac{\alpha_1}{2}}, u^{\alpha_M-\frac{\alpha_1}{2}}).\end{aligned}$$

Corollary 6.3 and Lemma 6.1 allow us to derive the following central limit theorem for the estimated moments. In particular, we use Lemma 6.1 to control the sampling variance, and Corollary 6.3 to control the bias.

Lemma 6.4. *Let $nh_n = T = 1$ constant, i.e. $h_n = 1/n$, and choose $u_n \to \infty$ according to U. Let $\boldsymbol{f}$ satisfy F1 and F2, and suppose that the Lévy process X_t satisfies (2.2) with some $\rho < \alpha/2$. Then, as $n \to \infty$,*

$$\tilde{\Lambda}_n^{-1}(\theta)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[\boldsymbol{f}(u_n\Delta_{n,i}X) - \mathbb{E}_\theta \boldsymbol{f}(u_n\tilde{Z}_h)\Big] \quad \Rightarrow \quad \mathcal{N}\left(0, \Sigma(\theta)\right).$$

Note that the rate of convergence for the first moment f_1 is slower than for $f_j, j \ge 2$. This is due to our special choice of $f_j, j \ge 2$, which vanish near zero. Hence, these moments are primarily driven by the jump component, which is of a smaller order than the diffusion term. On the other hand, the jump parameters $\alpha_m, r_m^{\pm}$ are harder to identify, i.e. $\partial_{\alpha_m}\mathbb{E}_\theta \boldsymbol{f}(u\tilde{Z}_h) \ll \partial_{\sigma^2}\mathbb{E}_\theta \boldsymbol{f}(u\tilde{Z}_h)$. This is established in the following Lemma.

Lemma 6.5. *Let $f \in \mathcal{C}^3(\mathbb{R})$ be such that f, f', f'', f''' are bounded. Let $\tilde{X}_t$ be a Lévy process with characteristic triplet $(0, \sigma^2, \tilde{\nu})$, parameterized by θ as in (2.4).*

Then, as $h \to 0$, $u \to \infty$, such that $hu^2 \to 0$,

$$\begin{aligned}
\partial_{\sigma^2} \mathbb{E}_\theta f(u\tilde{X}_h) &= h\frac{u^2}{2} f''(0) + o(hu^2), \\
\partial_{r_m^\pm} \mathbb{E}_\theta f(u\tilde{X}_h) &= hu^{\alpha_m} \mathcal{J}_{\alpha_m}^\pm f(0) + o(hu^{\alpha_m}) + \mathcal{O}\left(hu^{\alpha_m \vee 1} \log u\right) \mathbb{E}_\theta f'(u\tilde{X}_h), \\
\partial_{\alpha_m} \mathbb{E}_\theta f(u\tilde{X}_h) &= hu^{\alpha_m} (\log u) \left[r_m^+ \mathcal{J}_{\alpha_m}^+ f(0) + r_m^- \mathcal{J}_{\alpha_m}^- f(0)\right] \\
&\quad + o(hu^{\alpha_m} \log u) + \mathcal{O}\left(hu^{\alpha_m \vee 1} (\log u)^2\right) \mathbb{E}_\theta f'(u\tilde{X}_h),
\end{aligned} \tag{6.3}$$

and,

$$\begin{aligned}
&\left(\partial_{\alpha_m} - \log(u)\left(r_m^+ \partial_{r_m^+} + r_m^- \partial_{r_m^-}\right)\right) \mathbb{E}_\theta f(u\tilde{X}_h) \\
&= hu^{\alpha_m} \partial_{\alpha_m} \left[r_m^+ \mathcal{J}_{\alpha_m}^+ f(0) + r_m^- \mathcal{J}_{\alpha_m}^- f(0)\right] + o(hu^{\alpha_m}) \\
&\quad + \mathcal{O}\left(hu^{\alpha_m \vee 1} (\log u)^2\right) \mathbb{E}_\theta f'(u\tilde{X}_h).
\end{aligned} \tag{6.4}$$

Moreover, if f vanishes on $[-\eta, \eta]$ and u satisfies Condition U,

$$\partial_{\sigma^2} \mathbb{E}_\theta f(u\tilde{X}_h) = o(hu^\alpha). \tag{6.5}$$

All terms of the form $\mathcal{O}(\cdot)$ and $o(\cdot)$ are bounded resp. vanishing uniformly on compacts in Θ.

Corollary 6.6. *Let $\boldsymbol{f}$ satisfy F1 and F2, and let $\tilde{X}_t$ be a Lévy process with characteristic triplet $(0, \sigma, \tilde{\nu})$, parameterized by θ as in (2.4). Then, as $h = \frac{1}{n} \to 0$, $u_n \to \infty$, such that $u_n = o(\sqrt{h})$,*

$$\tilde{\Lambda}_n^{-1}(\theta) \left[\mathrm{D}_\theta \mathbb{E}_\theta \boldsymbol{f}(u_n \tilde{X}_h)\right] \Gamma_n(\theta) \bar{\Lambda}_n^{-1}(\theta) \to A(\theta). \tag{6.6}$$

This convergence holds uniformly on compacts in $\theta \in \Theta$.

These results allow us to establish the consistency of $\hat{\theta}_n$. We do not consider global uniqueness of the solution of the estimating equation (2.5). Hence, we only obtain the existence of a consistent sequences of random variables satisfying the equation.

Lemma 6.7 (Consistency)**.** *Let X_t be a Lévy process satisfying (2.2) with some $\rho < \alpha/2$, and parameter vector θ_0. Let $\boldsymbol{f}$ satisfy assumptions F1, F2, and I, and let $u_n \to \infty$ be chosen according to U. There exists a sequence of random vectors $\hat{\theta}_n$ solving (2.5) eventually, i.e. $P(F_n(\hat{\theta}_n) = 0) \to 1$, such that $\hat{\theta}_n \to \theta$ in probability as $n \to \infty$. This sequence is eventually unique in the sense that, i.e. for any other sequence $\hat{\theta}_n^*$ solving the estimating equation, and which satisfy $\|\hat{\theta}_n^* - \theta_0\| = \mathcal{O}_P(1/|\log u_n|^2)$, it holds $P(\hat{\theta}_n \neq \hat{\theta}_n^*) \to 0$.*

To obtain a central limit theorem for $\hat{\theta}_n$, we may apply a Taylor expansion to obtain the representation

$$\hat{\theta}_n - \theta_0 \approx -\left[\widetilde{\mathrm{D}_\theta \boldsymbol{f}}\right]^{-1} \frac{1}{n} \left[\sum_{i=1}^n \boldsymbol{f}(u_n \Delta_{n,i} X) - \mathbb{E}_{\theta_0} \boldsymbol{f}(u_n \tilde{Z}_h)\right],$$

where $\widetilde{\mathrm{D}\boldsymbol{f}}_{j,k} = \partial_{\theta_k} \mathbb{E}_{\tilde{\theta}^j} f_j(u_n \tilde{Z}_h)$ for some $\tilde{\theta}^j$ on the line segment between θ_0 and $\hat{\theta}_n$, for $j = 1, \ldots, 3M+1$. This standard approach allows to establish Theorem 2.1, as detailed in Subsection 6.2.

6.2. Proofs

Proof of Lemma 6.1. At the price of changing the term μ, we may assume w.l.o.g. that $\xi(z) = z\mathbb{1}_{|z|\leq 1}$. In view of the Lévy-Itô decomposition (2.1), we write

$$\begin{aligned} u\tilde{X}_t &= u\mu t + u\sigma B_t + \int uz \left(N(dz, ds) - \mathbb{1}_{|z|\leq \frac{1}{u}} \tilde{\nu}(dz) \otimes ds \right) \\ &\quad + t \int uz (\mathbb{1}_{|z|\leq \frac{1}{u}} - \mathbb{1}_{|z|\leq 1}) \tilde{\nu}(dz) \\ &= u\mu t + u\sigma B_t + J_t^u + ut\mu_u \end{aligned}$$

where N is a Poisson counting measure with intensity $\tilde{\nu}(dz) \otimes ds$, and J_t^u denotes the corresponding integral term. The explicit form of $\tilde{\nu}$ allows for computation of μ_u, as

$$\begin{aligned} |\mu_u| &\leq \int_{\frac{1}{u}}^1 \sum_{m=1}^M \alpha_m (r_m^+ + r_m^-) |z|^{-\alpha_m} \, dz \\ &\leq 2 \sum_{m=1}^M (r_m^+ + r_m^-) \int_{\frac{1}{u}}^1 |z|^{-\alpha_1} \, dz \\ &= 2 \sum_{m=1}^M (r_m^+ + r_m^-) \frac{u^{\alpha_1 - 1} - 1}{1 - \alpha_1} \quad \leq 2 \sum_{m=1}^M (r_m^+ + r_m^-) u^{(\alpha_1 - 1)\vee 0} \log(u). \end{aligned}$$

In the last inequality, we used the mean value theorem for the function $a \mapsto u^a$. Note that the bounds holds for $\alpha_1 \neq 1$ directly, and for $\alpha_1 = 1$ by continuity. This bound on μ_u will be used in the sequel.

To derive the claims of the Lemma, we start with a rough bound for the probability

$$\begin{aligned} P(|u\tilde{X}_t| > \eta) &\leq P\left(|ut(\mu + \mu_u)| > \frac{1-\lambda}{2} \eta \right) \\ &\quad + P\left(|\sigma u B_t| > \frac{1-\lambda}{2} \eta \right) + P\left(|J_t^u| > \lambda \eta \right), \quad \lambda \in (0,1). \end{aligned} \tag{6.7}$$

The first term tends to zero identically as $t \to 0$, because $|ut\mu_u| = \mathcal{O}(u^{\alpha_1 \vee 1} t) \to 0$. To study the jump term, choose a bounded, smooth function $g(x) \geq \mathbb{1}_{|x|\geq \lambda\eta}$ such that $g(0) = g'(0) = 0$. Then by Itô's formula, and a substitution in the integral, we obtain

$$
\begin{aligned}
&P\left(|J_t^u| > \lambda\eta\right) \\
&\le \mathbb{E} g(J_t^u) \\
&= \int_0^t \int \mathbb{E}\left[g(J_s^u + uz) - g(J_s^u) - g'(J_s^u) uz \mathbb{1}_{|z|\le \frac{1}{u}}\right] \tilde{\nu}(dz)\, ds \\
&\le \sum_{m=1}^M (r_m^+ + r_m^-)\alpha_m u^{\alpha_m} \int_0^t \int \frac{\mathbb{E}\left|g(J_s^u + z) - g(J_s^u) - g'(J_s^u) z \mathbb{1}_{|z|\le 1}\right|}{|z|^{1+\alpha_m}}\, dz\, ds \\
&\le \tilde{C} t u^{\alpha}(\|g\|_\infty + \|g''\|_\infty),
\end{aligned}
$$

for a constant $\tilde{C}$ depending on $\boldsymbol{\alpha}, \boldsymbol{r}$ and is bounded on compacts in these parameters. The function g can be chosen such that the latter term is finite. Thus, $P(|J_t^u| > \lambda\eta) = \mathcal{O}(u^\alpha t)$, uniformly on compacts in $\boldsymbol{\alpha}, \boldsymbol{r}$.

For the Gaussian term in (6.7), we employ the tail bound $P(|X| > x) \le \frac{1}{x\sqrt{2\pi}} \exp(-x^2/2)$, for $X \sim \mathcal{N}(0,1)$. In our situation, we obtain

$$
P\left(|B_1| > \frac{(1-\lambda)\eta}{2\sigma u\sqrt{t}}\right) \;\le\; \frac{2\sigma u\sqrt{t}}{(1-\lambda)\eta\sqrt{2\pi}} \exp\left(\frac{-\eta^2(1-\lambda)^2}{8\sigma^2 u^2 t}\right).
$$

Now let $a = a_t > 0$ be such that $u = \frac{(1-\lambda)\eta}{\sqrt{a}\sigma\sqrt{8t|\log t|}}$. Then

$$
P\left(|B_1| > \frac{(1-\lambda)\eta}{2\sigma u\sqrt{t}}\right) \;\le\; \frac{\exp(a\log t)}{2\sqrt{a\pi|\log t|}} = \frac{t^a}{2\sqrt{a\pi|\log t|}}.
$$

If $a = a_t \ge \underline{a}$ for some constant $\underline{a} > 1$, i.e. $u \le \frac{(1-\lambda)\eta}{\underline{a}\sigma\sqrt{8t|\log t|}}$, the latter bound is of order less than $\mathcal{O}(u^\alpha t)$, uniformly on compacts. Since we prove claim (i) for all $\lambda \in (0,1)$, requiring $a = a_t \ge \underline{a}$ is not a restriction. In particular,

$$
P(|u\tilde{X}_t| > \eta) \le \tilde{C} t u^\alpha. \tag{6.8}
$$

Note that the latter inequality does not hold if $u = \tau/\sqrt{-t\log t}$ for a proportionality factor τ which is too large.

If $u = u_t$ is larger than $(1-\lambda)/(\sigma\sqrt{8t|\log t|})$, but $u = o(1/\sqrt{t})$, the bound on $P(|J_t^u| > \lambda\eta)$ remains unchanged, while we still obtain $P(|u\sigma B_t| > \eta) \to 0$ uniformly on compacts. Thus, if we only suppose $u = o(1/\sqrt{t})$, we have $P(|u\tilde{X}_t| > \tilde{\eta}) \to 0$ uniformly on compacts, for any $\tilde{\eta} > 0$, but with a slower rate.

To obtain an asymptotically exact value, we plug the former rough bound (6.8) into Itô's formula. In case (i), we have

$$
\begin{aligned}
\mathbb{E} f(u\tilde{X}_t) = \mathbb{E}\int_0^t \Big[&\frac{u^2\sigma^2}{2} f''(u\tilde{X}_s) + (\mu + \mu_u) u f'(u\tilde{X}_s) \\
&+ \int \left(f(u\tilde{X}_s + uz) - f(u\tilde{X}_s) - uz\mathbb{1}_{|uz|\le 1} f'(u\tilde{X}_s)\right) \tilde{\nu}(dz)\Big] ds
\end{aligned}
$$

$$
\begin{aligned}
&= \int_0^t \frac{u^2\sigma^2}{2} \mathbb{E} f''(u\tilde{X}_s) + (\mu + \mu_u) u \mathbb{E} f'(u\tilde{X}_s)\, ds \\
&\quad + \sum_{m=1}^M u^{\alpha_m} \int_0^t \left[r_m^+ \mathbb{E}\mathcal{J}_{\alpha_m}^+ f(u\tilde{X}_s) + r_m^- \mathbb{E}\mathcal{J}_{\alpha_m}^- f(u\tilde{X}_s) \right] ds \\
&= u^2 t\, \mathcal{O}(u^\alpha t) + \sum_{m=1}^M u^{\alpha_m} \int_0^t \left[r_m^+ \mathbb{E}\mathcal{J}_{\alpha_m}^+ f(u\tilde{X}_s) + r_m^- \mathbb{E}\mathcal{J}_{\alpha_m}^- f(u\tilde{X}_s) \right] ds.
\end{aligned} \tag{6.9}
$$

Here, we used $\mathbb{E} f''(u\tilde{X}_s) \leq \|f''\|_\infty P(|u\tilde{X}_s| > \eta) = \mathcal{O}(u^\alpha t)$ as f vanishes on $[-\eta, \eta]$. We moreover used that $\mathbb{E} f'(u\tilde{X}_s) = \mathcal{O}(u^\alpha t)$, and $\mu_u u = \mathcal{O}(u^2)$ as established previously. These upper bounds hold uniformly on compacts in Θ. To proceed, note that $\mathcal{J}_\alpha^\pm f$ is a bounded continuous function, since

$$
|\mathcal{J}_\alpha^\pm f(x)| \leq 2\|f\|_\infty \int_{|z|\geq 1} \frac{\alpha}{|z|^{1+\alpha}} dz + \|f''\|_\infty \int_{|z|\leq 2} \frac{\alpha |z|^2}{|z|^{1+\alpha}} dz,
$$

which is furthermore bounded uniformly on compacts in α. By virtue of this boundedness, $u\tilde{X}_s \xrightarrow{P} 0$ implies $\mathbb{E}\mathcal{J}_{\alpha_m}^\pm f(u\tilde{X}_s) = \mathcal{J}_{\alpha_m}^\pm f(0) + o(1)$. To ensure that this last approximation holds uniformly on compacts in Θ, note that $\|(\mathcal{J}_{\alpha_m}^\pm f)'\|_\infty = \|\mathcal{J}_{\alpha_m}^\pm f'\|_\infty$ is also bounded, such that it suffices to control $\mathbb{E}(|u\tilde{X}_s| \wedge 1)$ uniformly. But we already established that for any η, $P(|u\tilde{X}_s| > \eta) \to 0$ uniformly on compacts in Θ. Hence,

$$
\begin{aligned}
\mathbb{E} f(u\tilde{X}_t) &= u^2 t \mathcal{O}(u^\alpha t) + \sum_{m=1}^M u^{\alpha_m} \int_0^t \left[r_m^+ \mathbb{E}\mathcal{J}_{\alpha_m}^+ f(u\tilde{X}_s) + r_m^- \mathbb{E}\mathcal{J}_{\alpha_m}^- f(u\tilde{X}_s) \right] ds \\
&= o(u^\alpha t) + \sum_{m=1}^M u^{\alpha_m} (r_m^+ \mathcal{J}_{\alpha_m}^+ f(0) + r_m^- \mathcal{J}_{\alpha_m}^- f(0)) \\
&= o(u^\alpha t) + u^\alpha t \left[r_1^+ \mathcal{J}_\alpha^+ f(0) + r_1^- \mathcal{J}_\alpha^- f(0) \right],
\end{aligned}
$$

uniformly on compacts in $\sigma^2, \boldsymbol{\alpha}, \boldsymbol{r}$. This proves the first claim.

In case (ii), i.e. $f(0) = 0, f''(0) \neq 0$, a different term dominates in (6.9). We obtain

$$
\begin{aligned}
\mathbb{E} f(u\tilde{X}_t) &= \int_0^t \frac{u^2\sigma^2}{2} \mathbb{E} f''(u\tilde{X}_s)\, ds + \mathcal{O}(u^\alpha t) \\
&= \mathcal{O}(t u^\alpha) + \frac{u^2 t}{2} \left(f''(0) + o(1) \right),
\end{aligned}
$$

uniformly on compacts in Θ.

For case (iii), i.e. $f''(0) = 0, f^{(4)}(0) \neq 0$, we may apply the result of case (ii) to obtain $\mathbb{E} f''(u\tilde{X}_t) = \frac{u^2 t \sigma^2}{2} f^{(4)}(0) + o(u^2 t)$, and hence

$$
\mathbb{E} f(u\tilde{X}_t) = \int_0^t \frac{u^2\sigma^2}{2} \mathbb{E} f''(u\tilde{X}_s) ds + \mathcal{O}(u^\alpha t)
$$

$$
\begin{aligned}
&= \int_0^t \frac{u^4\sigma^4}{4} s f^{(4)}(0) ds + \mathcal{O}(u^\alpha t) + o(u^4 t^2) \\
&= \frac{u^4 t^2 \sigma^4}{8} f^{(4)}(0) ds + \mathcal{O}(u^\alpha t) + o(u^4 t^2).
\end{aligned}
$$

For the last claim, we use Itô's formula again. Recall that the truncation function satisfies $\xi(z) = z$ for $|z| \le 1$, and $|\xi(z)| \le 2$. Then

$$
\begin{aligned}
&\mathbb{E} f(u\tilde{X}_t) \\
&= \mathbb{E} \int_0^t \int \left[f(u(\tilde{X}_s + z)) - f(u\tilde{X}_s) - u f'(u\tilde{X}_s)\xi(z) \right] \tilde{\nu}(dz) \\
&\le 2\, t \|f\|_\infty \tilde{\nu}\left(\left(-\frac{1}{u}, \frac{1}{u} \right)^c \right) + 2\, tu \|f'\|_\infty \tilde{\nu}((-1,1)^c) \\
&\quad + tu \|f'\|_\infty \int_{(-1,1)\setminus(-\frac{1}{u},\frac{1}{u})} |z|\, \tilde{\nu}(dz) + tu^2 \|f''\|_\infty \int_{-\frac{1}{u}}^{\frac{1}{u}} z^2\, \tilde{\nu}(dz) \\
&\le t\tilde{C} \left(\|f\|_\infty u^\alpha + u\|f'\|_\infty + u\|f'\|_\infty (u^{\alpha-1} + 1) + u^2 \|f''\|_\infty u^{\alpha-2} \right) (1 + \log(u)) \\
&\le t\tilde{C} u^{\alpha \vee 1} (1 + \log(u)) \left(\|f\|_\infty + \|f'\|_\infty + \|f''\|_\infty \right)
\end{aligned}
$$

The additional factor $\log(u)$ is introduced to cover the special case $\alpha = 1$ when computing the integral $\int_{1/u}^1 |z|^{-\alpha} dz$. □

Proof of Lemma 6.2. Choose some $0 < \epsilon < \frac{1}{u}$. The process X_t may be decomposed by virtue of the Lévy-Itô decomposition as

$$
\begin{aligned}
X_t &= \mu t + \sigma B_t + \int_0^t \int (z - \xi(z))\, N(dz, ds) + \int_0^t \int \xi(z)\, (N - \nu)(dz, ds) \\
&= \mu t + \sigma B_t + J_t^1 + J_t^2 + J_t^3 - t\zeta_\epsilon, \\
J_t^1 &= \int_0^t \int_{[-\epsilon,\epsilon]} \xi(z)(N-\nu)(dz, dt), \\
J_t^2 &= \sum_{s \le t} \Delta X_s \mathbb{1}_{\epsilon < |\Delta X_s| \le \frac{1}{u}}, \\
J_t^3 &= \sum_{s \le t} \Delta X_s \mathbb{1}_{\frac{1}{u} < |\Delta X_s|}, \\
\zeta_\epsilon &= \int_{|z| > \epsilon} \xi(z)\, \nu(dz),
\end{aligned}
$$

where $(N - \nu)$ is a compensated homogeneous Poisson point process with intensity measure $\nu(dz)$, such that J_t^1 is a martingale. For $\tilde{X}_t$, we have the analogous decomposition $\tilde{X}_t = \mu t + \sigma \tilde{B}_t + \tilde{J}_t^1 + \tilde{J}_t^2 + \tilde{J}_t^3 + t\tilde{\zeta}_\epsilon$. Since we are only interested in expected values, we may assume without loss of generality that X_t and $\tilde{X}_t$

are defined on the same probability space, and that $\tilde{B}_t = B_t$. Moreover,

$$\begin{aligned}\zeta_\epsilon - \tilde{\zeta}_\epsilon &= \int_{|z|>\epsilon} \xi(z)\,(\nu-\tilde{\nu})(dz) \\ &= \int_{\epsilon<|z|<1} z\,(\nu-\tilde{\nu})(dz) + \int_{|z|>1} \xi(z)\,(\nu-\tilde{\nu})(dz).\end{aligned}$$

The second integral is finite. Furthermore, integrating by parts,

$$\int_\epsilon^1 z(\nu-\tilde{\nu})(dz) = \int_\epsilon^1 \left[\nu((z,1]) - \tilde{\nu}((z,1])\right] dz + \epsilon\left[\nu((\epsilon,1]) - \tilde{\nu}((\epsilon,1])\right],$$

which has a limit as $\epsilon \to 0$ if $\rho < 1$. Thus, there exists a real number $\bar{\zeta}$ such that $\zeta_\epsilon - \tilde{\zeta}_\epsilon \to \bar{\zeta}$ as $\epsilon \to 0$.

By subadditivity of the total variation distance and the Wasserstein distance,

$$\begin{aligned}&\left|\mathbb{E}f(uX_t) - \mathbb{E}f(u\tilde{X}_t - ut\bar{\zeta})\right| \\ &\leq \left|\mathbb{E}f\left(uX_t\right) - \mathbb{E}f\left(u((\mu-\tilde{\zeta}_\epsilon-\bar{\zeta})t + \sigma B_t + \tilde{J}_t^1 + \tilde{J}_t^2 + J_t^3)\right)\right| \\ &\quad + \left|\mathbb{E}f\left(u((\mu-\tilde{\zeta}_\epsilon-\bar{\zeta})t + \sigma B_t + \tilde{J}_t^1 + \tilde{J}_t^2 + J_t^3)\right) - \mathbb{E}f\left(u(\tilde{X}_t - t\bar{\zeta})\right)\right| \\ &\leq u\|f'\|_\infty \left(t|\bar{\zeta} - (\zeta_\epsilon - \tilde{\zeta}_\epsilon)| + d_W(J_t^1, \tilde{J}_t^1) + d_W(J_t^2, \tilde{J}_t^2)\right) \\ &\quad + \left|\mathbb{E}f\left(u((\mu-\tilde{\zeta}_\epsilon-\bar{\zeta})t + \sigma B_t + \tilde{J}_t^1 + \tilde{J}_t^2 + J_t^3)\right) - \mathbb{E}f\left(u(\tilde{X}_t - t\bar{\zeta})\right)\right|.\end{aligned} \tag{6.10}$$

We treat all terms in (6.10) individually.

Part (i) The small jumps can be handled by noting

$$d_W(J_t^1, \tilde{J}_t^1) \leq \mathbb{E}|J_t^1| + \mathbb{E}|\tilde{J}_t^1| \leq \sqrt{\mathbb{E}|J_t^1|^2} + \sqrt{\mathbb{E}|\tilde{J}_t^1|^2}. \tag{6.11}$$

Since J_t^1 and $\tilde{J}_t^1$ have bounded jumps, we have $\mathbb{E}|J_t^1|^2, \mathbb{E}|\tilde{J}_t^1|^2 \to 0$ as $\epsilon \to 0$. Furthermore, $|\bar{\zeta} - (\zeta_\epsilon - \tilde{\zeta}_\epsilon)| \to 0$ as $\epsilon \to 0$.

Part (ii) As a next step, we study the medium sized jumps J_t^2. Consider the slightly more general process

$$J_t^{(a,b]} = \sum_{s\leq t} \Delta X_s \mathbb{1}_{a<|\Delta X_s|\leq b},$$

for $0 < a < b < 1$. Let $\tilde{J}_t^{(a,b]}$ be defined analogously based on $\tilde{X}_t$. These are compound Poisson processes, which can be written as

$$J_t^{(a,b]} = \sum_{i=1}^{N_t} U_i, \qquad \tilde{J}_t^{(a,b]} = \sum_{i=1}^{\tilde{N}_t} \tilde{U}_i,$$

where N_t is a Poisson counting process with intensity $\eta((a,b]) = \nu([-b,-a) \cup (a,b])$, and the U_i are iid random variables with distribution $\frac{\nu(dz)\mathbb{1}(a<|z|\le b)}{\eta((a,b])}$. Vice versa, the same holds for $\tilde{N}_t$ and $\tilde{U}_i$ with $\tilde{\eta}((a,b]) = \tilde{\nu}([-b,-a) \cup (a,b])$. Then Theorem 10 and Proposition 3 of Mariucci and Reiß (2018) for $p = 1$, yield

$$d_W(J_t^{(a,b]}, \tilde{J}_t^{(a,b]}) = d_W\left(\sum_{i=1}^{N_t} U_i, \sum_{i=1}^{\tilde{N}_t} \tilde{U}_i\right) \le t\eta((a,b])d_W(U_1,\tilde{U}_1) + t\,|\eta((a,b]) - \tilde{\eta}((a,b])|\,\mathbb{E}|\tilde{U}_1|. \quad (6.12)$$

We compute

$$\begin{aligned}
&\mathbb{E}|\tilde{U}_1| \\
&= \frac{1}{\tilde{\eta}((a,b])}\int_{|z|\in(a,b]} z\,\tilde{\nu}(dz) \qquad (6.13)\\
&= \frac{1}{\tilde{\eta}((a,b])}\left[a\tilde{\nu}((a,b]) + \int_a^b \tilde{\nu}((z,b])dz + a\tilde{\nu}([-b,-a)) + \int_a^b \tilde{\nu}([-b,-z))dz\right]\\
&= a + \int_a^b \frac{\tilde{\eta}((z,b])}{\tilde{\eta}((a,b])}\,dz.
\end{aligned}$$

Recall that $\tilde{\eta}((z,b]) = \sum_{m=1}^M (r_m^+ + r_m^-)(|z|^{-\alpha_m} - b^{-\alpha_m})$. Then there exists a constant $\tilde{C}$ which is bounded on compacts in Θ and L, such that for $z < b/2$, and $\alpha = \alpha_1$,

$$\frac{1}{\tilde{C}}|z|^{-\alpha} \le \tilde{\eta}((z,b]) \le \tilde{C}|z|^{-\alpha}. \quad (6.14)$$

In particular, for a potentially different constant $\tilde{C}$, and some $\tilde{\alpha}$ between α and 1,

$$\begin{aligned}
\mathbb{E}|\tilde{U}_1| &\le a + \tilde{C}a^\alpha \int_a^b |z|^{-\alpha}\,dz \\
&= a + \tilde{C}a^\alpha b^{1-\alpha}\frac{(\frac{a}{b})^0 - (\frac{a}{b})^{1-\alpha}}{1-\alpha} \\
&= a + \tilde{C}a^\alpha b^{1-\alpha}\left(\frac{a}{b}\right)^{1-\tilde{\alpha}}|\log(a/b)| \\
&\le a + \tilde{C}|\log a|\,a\left(\frac{a}{b}\right)^{\alpha-\tilde{\alpha}} \\
&\le a + \tilde{C}|\log a|\,a\left(\frac{a}{b}\right)^{(\alpha-1)\wedge 0} \quad \le \tilde{C}|\log a|\,a^{1\wedge\alpha}.
\end{aligned}$$

For $\alpha = 1$, this bound holds by continuity. Here and in the following, the constant $\tilde{C}$ may vary from line to line, and is bounded on compacts in θ, L, and ρ.

Furthermore, since ν and $\tilde{\nu}$ are sufficiently similar,

$$\begin{aligned}\eta((a,b]) &= \nu((a,\infty)) + \nu((-\infty,-a)) - \nu([-b,b]^c) \\ &= \tilde{\eta}((a,b]) + \xi,\end{aligned}$$

for $|\xi| \leq 2L(a^{-\rho} + b^{-\rho}) \leq 4La^{-\rho}$. Thus, the second term in (6.12) is of order $\mathcal{O}(ta^{(1\wedge\alpha)-\rho})$. Moreover, $|\eta((a,b])| \leq \tilde{C}(a^{-\alpha} + a^{-\rho}) = \mathcal{O}(a^{-\alpha})$ for small a, since $\rho < \alpha$.

We now consider the distance $d_W(U_1, \tilde{U}_1)$ occurring in (6.12), which can be expressed in terms of their cumulative distribution functions as

$$\begin{aligned} & d_W(U_1,\tilde{U}_1) \\ &= \int_{-b}^{b} \Big| P(U_1 \leq v) - P(\tilde{U}_1 \leq v) \Big| \, dv \\ &= \int_{-b}^{-a} \Big| P(U_1 \leq v) - P(\tilde{U}_1 \leq v) \Big| \, du + \int_{a}^{b} \Big| P(U_1 > v) - P(\tilde{U}_1 > v) \Big| \, dv \\ &\quad + 2a \Big| P(U_1 \leq -a) - P(\tilde{U}_1 \leq -a) \Big| . \end{aligned} \tag{6.15}$$

In the second step, we use that U_1 is supported on $[-b,-a) \cup (a,b]$. For $-b \leq v \leq -a$, and $b \leq 1$, it holds

$$\begin{aligned} & \Big| P(U_1 \leq v) - P(\tilde{U}_1 \leq v) \Big| \\ &= \left| \frac{\nu([-b,v])}{\eta((a,b])} - \frac{\tilde{\nu}([-b,v])}{\tilde{\eta}((a,b])} \right| \\ &\leq \left| \frac{1}{\eta((a,b])} - \frac{1}{\tilde{\eta}((a,b])} \right| \tilde{\nu}([-b,v]) \\ &\qquad + \frac{1}{\eta((a,b])} \left| \nu([-b,v]) - \tilde{\nu}([-b,v]) \right| \\ &\leq \tilde{C}|v|^{-\alpha} \frac{|\eta((a,b]) - \tilde{\eta}((a,b])|}{[\eta((a,b]) \wedge \tilde{\eta}((a,b])]^2} + \frac{|\nu([-b,v]) - \tilde{\nu}([-b,v])|}{\eta((a,b]) \wedge \tilde{\eta}((a,b])}. \end{aligned} \tag{6.16}$$

Recall that $|\eta((a,b]) - \tilde{\eta}((a,b])| = \mathcal{O}(a^{-\rho})$. Furthermore, the assumed similarity of ν and $\tilde{\nu}$ implies that $|\nu([-b,v]) - \tilde{\nu}([-b,v])| \leq L(|v|^{-\rho} + b^{-\rho}) \leq 2L|v|^{-\rho}$, and

$$\eta((a,b]) \wedge \tilde{\eta}((a,b]) \geq \tilde{\eta}((a,b]) - 2La^{-\rho} \geq \frac{a^{-\alpha}}{\tilde{C}} \tag{6.17}$$

as $a \to 0$, whenever $b \geq 2a$. In this case, for $-b \leq v \leq -a$,

$$\Big| P(U_1 \leq v) - P(\tilde{U}_1 \leq v) \Big| \leq \tilde{C}|v|^{-\alpha} a^{2\alpha-\rho} + \tilde{C}|v|^{-\rho} a^{\alpha} \leq \tilde{C}|v|^{-\rho} a^{\alpha}. \tag{6.18}$$

The analogous bound holds for $|P(U_1 > v) - P(\tilde{U}_1 > v)|$, when $a \leq v \leq b$. Now plug (6.18) into expression (6.15) for the Wasserstein distance, to obtain for $a \to 0$ and $a \leq \frac{b}{2}$,

$$d_W(U_1,\tilde{U}_1) \leq \tilde{C} \left(a^{\alpha} b^{1-\rho} + a^{\alpha+1-\rho} \right),$$

where we used $\rho < 1$. Using (6.12) and $|\eta((a,b]) - \tilde{\eta}((a,b])| \leq \tilde{C}a^{-\rho}$, we may hence bound,

$$\begin{aligned} d_W(J_t^{(a,b]}, \tilde{J}_t^{(a,b]}) &\leq \tilde{C}t \left(a^{-\alpha}(a^{\alpha}b^{1-\rho} + a^{\alpha+1-\rho}) + a^{(1\wedge\alpha)-\rho} \right) \\ &\leq \tilde{C}t \left(b^{1-\rho} + a^{(1\wedge\alpha)-\rho}|\log a| \right), \end{aligned} \tag{6.19}$$

This upper bound will be exploited in the rest of the proof. In particular, for $J_t^2 = J_t^{(\epsilon,1/u]}$ and ϵ small enough,

$$\begin{aligned} d_W(J_t^2, \tilde{J}_t^2) &\leq \tilde{C}t(u^{\rho-1} + \epsilon^{(1\wedge\alpha)-\rho}) \\ &\leq 2\tilde{C}tu^{\rho-1}, \end{aligned} \tag{6.20}$$

since $(1\wedge\alpha) - \rho > 0$ and $u \geq 1$. Note that, here, the suitable choice of ϵ depends on u, but we will let $\epsilon \to 0$ for any u in step (v) below.

Part (iii) It remains to study the term in (6.10) due to the large jumps. Here, our approach is slightly different as we will not (only) bound a metric distance between J_t^3 and $\tilde{J}_t^3$. Define

$$f_{u,t}(x) = \mathbb{E}f(u(x + t(\mu - \tilde{\zeta}_\epsilon - \bar{\zeta}) + \sigma B_t + \tilde{J}_t^1 + \tilde{J}_t^2)),$$

and we consider $\left|\mathbb{E}f_{u,t}(J_t^3) - \mathbb{E}f_{u,t}(\tilde{J}_t^3)\right|$, as suggested by (6.10). Since J_t^3 is a Lévy process, Itô's formula yields

$$\begin{aligned} \mathbb{E}f_{u,t}(J_t^3) &= f_{u,t}(0) + \int_0^t \mathbb{E}\mathcal{J}^3 f_{u,t}(J_s^3)ds, \\ \mathcal{J}^3 g(x) &= \int_{[-\frac{1}{u},\frac{1}{u}]^c} [g(x+z) - g(x)]\ \nu(dz), \end{aligned} \tag{6.21}$$

i.e., $\mathcal{J}^3$ is the infinitesimal generator of J_t^3. Analogously, we denote by $\tilde{\mathcal{J}}^3$ the generator of $\tilde{J}_t^3$. Then integration by parts yields, for any $x \in \mathbb{R}$,

$$\begin{aligned} &\left| \int_{(1/u,\infty)} [f_{u,t}(x+z) - f_{u,t}(x)]\ (\nu - \tilde{\nu})(dz) \right| \\ &= \left| \left[f_{u,t}\left(x + \frac{1}{u}\right) - f_{u,t}(x) \right] [\nu((1/u,\infty)) - \tilde{\nu}((1/u,\infty))] \right. \\ &\quad \left. + \int_{\frac{1}{u}}^{\infty} [\nu((z,\infty)) - \tilde{\nu}((z,\infty))]\, f'_{u,t}(x+z)\, dz \right| \\ &\leq 2\|f\|_\infty Lu^{\rho} + \int_{\frac{1}{u}}^{1} Lz^{-\rho}|f'_{u,t}(x+z)|\, dz \\ &\quad + [\nu((1,\infty)) + \tilde{\nu}((1,\infty))] \int_1^{\infty} |f'_{u,t}(x+z)|\, dz \end{aligned}$$

$$\leq \tilde{C}\|f\|_\infty u^{\alpha-\delta} + \tilde{C}u^\rho \int_{\frac{1}{u}}^1 |f'_{u,t}(x+z)|\,dz + \tilde{C}\int_1^\infty |f'_{u,t}(x+z)|\,dz.$$

The same bound holds for the range of integration $z \in (-\infty, -1/u)$, such that

$$\left|\mathcal{J}^3 f_{u,t}(x) - \tilde{\mathcal{J}}^3 f_{u,t}(x)\right| \leq \tilde{C}u^\rho \left(\|f\|_\infty + \int_{-\infty}^\infty |f'_{u,t}(z)|\,dz\right).$$

Now note that,

$$f'_{u,t}(x) = u\,\mathbb{E}f'\left(u(x + t(\mu+\zeta_0) + \sigma B_t + \tilde{J}_t^1 + \tilde{J}_t^2)\right),$$

such that by Fubini's theorem,

$$\begin{aligned}\int_{-\infty}^\infty |f'_{u,t}(z)|\,dz &\leq \mathbb{E}\int_{-\infty}^\infty u\left|f'\left(u(z + t(\mu+\zeta_0) + \sigma B_t + \tilde{J}_t^1 + \tilde{J}_t^2)\right)\right|\,dz \\ &= \mathbb{E}\int_{-\infty}^\infty |f'(v)|dv \qquad = \|f'\|_{L_1(\mathbb{R})},\end{aligned}$$

where we performed a linear substitution in the second step. Hence,

$$\left|\mathcal{J}^3 f_{u,t}(x) - \tilde{\mathcal{J}}^3 f_{u,t}(x)\right| \leq \tilde{C}u^\rho(\|f\|_\infty + \|f'\|_{L_1}). \tag{6.22}$$

Using this in (6.21),

$$\begin{aligned}\mathbb{E}f_{u,t}(J_t^3) &= f_{u,t}(0) + \int_0^t \mathbb{E}\tilde{\mathcal{J}}^3 f_{u,t}(J_s^3)\,ds + \mathcal{O}(u^\rho t(\|f\|_\infty + \|f'\|_{L_1})) \\ &= f_{u,t}(0) + \int_0^t \mathbb{E}\tilde{\mathcal{J}}^3 f_{u,t}(\tilde{J}_s^3) + \mathcal{O}\left(|\mathbb{E}\tilde{\mathcal{J}}^3 f_{u,t}(J_s^3) - \mathbb{E}\tilde{\mathcal{J}}^3 f_{u,t}(\tilde{J}_s^3)|\right)\,ds \\ &\quad + \mathcal{O}(u^\rho t(\|f\|_\infty + \|f'\|_{L_1})) \\ &= \mathbb{E}f_{u,t}(\tilde{J}_t^3) + \mathcal{O}\left(u^\rho t(\|f\|_\infty + \|f'\|_{L_1})\right) \\ &\quad + \mathcal{O}\left(\|\tilde{\mathcal{J}}^3 f_{u,t}\|_\infty \int_0^t d_{TV}\left(J_s^{(1,\infty)}, \tilde{J}_s^{(1,\infty)}\right)ds\right) \\ &\quad + \mathcal{O}\left(\|(\tilde{\mathcal{J}}^3 f_{u,t})'\|_\infty \int_0^t d_W\left(J_s^{(\frac{1}{u},1]}, \tilde{J}_s^{(\frac{1}{u},1]}\right)ds\right).\end{aligned} \tag{6.23}$$

We now study the latter two terms.

Part (iv) The total variation distance can be bounded by noting that $J_t^{(1,\infty)}$ and $\tilde{J}_t^{(1,\infty)}$ admit only finitely many jumps. The number of their jumps is Poisson distributed, such that

$$\begin{aligned}d_{TV}(J_t^{(1,\infty)}, 0) = 1 - P(J_t^{(1,\infty)} = 0) &= 1 - \exp\left[-t\nu((-1,1)^c)\right] \\ &\leq t\,\nu((-1,1)^c).\end{aligned}$$

In particular,

$$d_{TV}(J_t^{(1,\infty)}, \tilde{J}_t^{(1,\infty)}) \leq t\,[\nu((-1,1)^c) + \tilde{\nu}((-1,1)^c)] \leq t\tilde{C}. \tag{6.24}$$

Moreover,

$$\begin{aligned}\left|\tilde{\mathcal{J}}^3 f_{u,t}(x)\right| &\leq \int_{[-\frac{1}{u},\frac{1}{u}]^c} |f_{u,t}(x+z) - f_{u,t}(x)|\tilde{\nu}(dz) \\ &\leq \|f_{u,t}\|_\infty \tilde{\nu}([-1/u,1/u]^c) \\ &\leq \tilde{C}u^\alpha \|f\|_\infty. \end{aligned} \tag{6.25}$$

Via the same argument, we also obtain

$$\left|\frac{d}{dx}\tilde{\mathcal{J}}^3 f_{u,t}(x)\right| = \left|\int_{[-\frac{1}{u},\frac{1}{u}]^c} (f'_{u,t}(x+z) - f'_{u,t}(x))\tilde{\nu}(dz)\right| \leq \tilde{C}u^{\alpha+1}\|f'\|_\infty. \tag{6.26}$$

From (6.19), we know that

$$\int_0^t d_W\left(J_s^{(\frac{1}{u},1]}, \tilde{J}_s^{(\frac{1}{u},1]}\right)\,ds \leq \tilde{C}\int_0^t s\,ds \leq \tilde{C}t^2.$$

In combination with (6.23), we thus obtain

$$\begin{aligned}\left|\mathbb{E}f_{u,t}(J_t^3) - \mathbb{E}f_{u,t}(\tilde{J}_t^3)\right| &\leq tu^\rho \tilde{C}(\|f\|_\infty + \|f'\|_\infty + \|f'\|_{L_1}) \\ &\quad + \tilde{C}t^2u^\alpha\|f\|_\infty + \tilde{C}t^2u^{\alpha+1}\|f'\|_\infty \\ &\leq \tilde{C}(\|f\|_\infty + \|f'\|_{L_1})(tu^\rho + t^2u^{\alpha+1}). \end{aligned} \tag{6.27}$$

Part (v) Now putting (6.11), (6.20), and (6.27) into (6.10), and letting $\epsilon \to 0$,

$$\left|\mathbb{E}f(uX_t) - \mathbb{E}f(u\tilde{X}_t - ut\bar{\zeta})\right| \leq \tilde{C}(\|f\|_\infty + \|f'\|_\infty + \|f'\|_{L_1})(tu^\rho + t^2u^{\alpha+1}). \tag{6.28}$$

It can be checked that the upper bounds which are summarized in the constant $\tilde{C}$ all satisfy the desired uniformity on compacts in $\boldsymbol{\alpha}$, $\boldsymbol{r}$, L, and $\rho - \alpha < 0$. This concerns the lines (6.14), (6.16), (6.17), (6.22), (6.24), (6.25), (6.26). □

Proof of Corollary 6.3. A Taylor expansion yields, for any $a \in \mathbb{R}$,

$$|\mathbb{E}f(u(\tilde{X}_t + ta)) - \mathbb{E}f(u\tilde{X}_t)| \leq |uta\mathbb{E}f'(u\tilde{X}_t)| + \|f''\|_\infty t^2u^2a^2.$$

We denote $\tilde{X}_t = \sigma B_t + \tilde{J}_t$, where $\tilde{J}_t$ is the purely discontinuous component of $\tilde{X}$. Introduce for any function g the notation $g_{[u]}(x) = \mathbb{E}g(u\sigma B_t + x)$. Then for any k-th derivative, $\|g_{[u]}^{(k)}\|_\infty \leq \|g^{(k)}\|_\infty$. In particular, by Lemma 6.1 (iv),

$$|\mathbb{E}f'(u\tilde{X}_t) - \mathbb{E}f'(u\sigma B_t)| = |\mathbb{E}f'_{[u]}(u\tilde{J}_t) - f'_{[u]}(0)| \\ \leq tu^\alpha(1+\log(u))\left(\|f'\|_\infty + \|f''\|_\infty + \|f'''\|_\infty\right)\tilde{C}.$$

Since f is symmetric on $[-\eta,\eta]$, the derivative f' is odd on $[-\eta,\eta]$. We may write $f'(x) = f'(x)\mathbb{1}_{|x|\leq\eta} + f'(x)\mathbb{1}_{|x|>\eta} = \bar{f}_1(x) + \bar{f}_2(x)$. Then $\bar{f}_1$ is odd, such that $\mathbb{E}\bar{f}_1(u\sigma B_t) = 0$. Hence,

$$|f'_{[u]}(0)| = |\mathbb{E}f'(u\sigma B_t)| = \left|\mathbb{E}f'(u\sigma B_t)\mathbb{1}_{|u\sigma B_t|>\eta}\right| \quad \leq \|f'\|_\infty P(|u\sigma B_t| > \eta).$$

If u is chosen as specified, the latter term is of order tu^α, see Lemma 6.1 (i). We have thus shown that $|\mathbb{E}f'(u\tilde{X}_t)| = \mathcal{O}(tu^\alpha \log u)$. This yields

$$\begin{aligned} &|\mathbb{E}f(u(\tilde{X}_t + ta)) - \mathbb{E}f(u\tilde{X}_t)| \\ &\leq t^2 u^{2\vee(\alpha+1)}(1+\log(u))\left(\|f'\|_\infty + \|f''\|_\infty + \|f'''\|_\infty\right)(|a| + |a|^2)\tilde{C}. \end{aligned} \tag{6.29}$$

Moreover, $|\mathbb{E}f(uX_t) - \mathbb{E}f(u(\tilde{X}_t + t\mu - t\bar{\zeta}))| \leq \tilde{C}(tu^\rho + t^2u^{\alpha+1})$ from Lemma 6.2. Applying (6.29) for the drift $a = \mu - \bar{\zeta}$, this yields (6.2). □

Proof of Lemma 6.4. All summands $\boldsymbol{f}(u_n\Delta_{n,i}X)$ are iid and bounded. Furthermore, $\tilde{\Lambda}_n^{-1}/\sqrt{n} \to 0$, such that the Lindeberg-Feller condition for triangular arrays of independent r.v.s is satisfied (Durrett, 2005, Thm. 2.4.5). Moreover, the bias is of order $|\mathbb{E}\boldsymbol{f}(u_n\Delta X_{t_i}) - \mathbb{E}\boldsymbol{f}(u_n\tilde{Z}_h)| = \mathcal{O}(h_n u_n^\rho)$ by Corollary 6.3. If $\rho < \alpha/2$, this is small enough to ensure $\tilde{\Lambda}_n^{-1}\sqrt{n}[\mathbb{E}\boldsymbol{f}(\Delta_{n,i}X) - \mathbb{E}_\theta\boldsymbol{f}(u_n\tilde{Z}_h)] = o(1)$. Hence, the bias is asymptotically negligible.

It thus suffices to check the asymptotic covariance structure. Denote $f_{j,k}(x) = f_j(x)f_k(x)$. Then $f_{j,k}$ is smooth and vanishes on $[-\eta,\eta]$ unless $j = 1 = k$. Moreover, $f_{1,1}(0) = f'_{1,1}(0) = f''_{1,1}(0) = 0$ and $f_{1,1}^{(4)}(0) = 6f''_1(0)^2$. Corollary 6.3 and Lemma 6.1 yield

$$\begin{aligned} \mathbb{E}f_{j,k}(u_n\Delta_{n,i}X) &= \mathbb{E}_\theta f_{j,k}(u_n\tilde{Z}_h) + \mathcal{O}(h_n u_n^\rho) \\ &= u_n^\alpha h\left(r_1^+\mathcal{J}_\alpha^+ f_{j,k}(0) + r_1^-\mathcal{J}_\alpha^- f_{j,k}(0)\right) + o(u_n^\alpha h),\ (j,k) \neq (1,1), \\ \mathbb{E}f_{1,1}(u_n\Delta_{n,i}X) &= \frac{3}{4}u_n^4h^2\sigma^4 f''_1(0)^2 + o(u_n^4h^2) + \mathcal{O}(u_n^\alpha h) + \mathcal{O}(u_n^\rho h) \\ &= \frac{3}{4}\sigma^4u_n^4h^2f''_1(0)^2 + o(u_n^4h^2). \end{aligned}$$

To compute the asymptotic covariance, we further determine

$$\begin{aligned} \left(\mathbb{E}f_1(u_n\Delta_{n,i}X)\right)^2 &= \left(hu^2\frac{\sigma^2}{2}f''_1(0) + o(hu^2) + \mathcal{O}(hu^\rho)\right)^2 \\ &= \frac{h^2u^4\sigma^4}{4}f''_1(0)^2 + o(h^2u^4), \end{aligned}$$

and for $j \geq 2, k \geq 1$,

$$\mathbb{E}f_j(u_n\Delta_{n,i}X)\,\mathbb{E}f_k(u_n\Delta_{n,i}X) = \mathcal{O}(u_n^\alpha h)\cdot\mathcal{O}(u_n^2h) = o(u_n^\alpha h).$$

These approximations can be summarized as

$$\begin{aligned}
&\operatorname{Cov}(\boldsymbol{f}(u_n\Delta_{n,i}X))_{j,k} \\
&= \begin{cases} \frac{\sigma^4}{2}u_n^4h^2f''(0)^2 + o(u_n^4h^2), & j=k=1 \\ u_n^\alpha h\left(r_1^+\mathcal{J}_\alpha^+ f_{j,k}(0) + r_1^-\mathcal{J}_\alpha^- f_{j,k}(0)\right) + o(u_n^\alpha h), & \text{otherwise,} \end{cases}
\end{aligned}$$

This scaling behavior yields $\operatorname{Cov}_\theta(\tilde{\Lambda}_n^{-1}(\theta)\boldsymbol{f}(u_n\Delta_{n,i}X)) \to \Sigma(\theta)$ as $n \to 0$, and thus the desired central limit theorem. □

Proof of Lemma 6.5. First, assume f to be a Schwartz function with Fourier transform $\hat{f}(\lambda) = \frac{1}{2\pi}\int f(x)e^{i\lambda x}\,dx$. Then

$$\mathbb{E}f(u\tilde{X}_h) = \int \hat{f}(\lambda/u)e^{-h\psi_\theta(\lambda)}\,d\lambda,$$

where ψ_θ is the Lévy symbol of $\tilde{X}_h$, i.e. $\mathbb{E}_\theta\exp(i\lambda\tilde{X}_h) = \exp(-h\psi_\theta(\lambda))$. In particular, for any entry θ_j of the parameter vector θ,

$$\partial_{\theta_j}\mathbb{E}_\theta f(u\tilde{X}_h) = -h\int \hat{f}(\lambda)\left(\partial_{\theta_j}\psi_\theta(u\lambda)\right)e^{-h\psi_\theta(u\lambda)}\,d\lambda.$$

Integration and differentiation may be exchanged because f is a Schwartz function and ψ has polynomial growth. In particular, via the Lévy-Khintchine formula, the Lévy symbol may be determined as

$$\begin{aligned}
\psi_\theta(u\lambda) &= \frac{u^2\sigma^2\lambda^2}{2} + \int\left[e^{iu\lambda z} - 1 - iu\lambda\xi(z)\right]\tilde{\nu}(dz) \\
&= \frac{u^2\sigma^2\lambda^2}{2} - i\lambda\int_{|z|\geq\frac{1}{u}}[u\xi(z) - \xi(uz)]\,\tilde{\nu}(dz) \\
&\quad + \sum_{m=1}^M \alpha_m u^{\alpha_m}\lambda^{\alpha_m}\int\frac{e^{iz}-1-i\xi(z)}{|z|^{1+\alpha_m}}\left(r_m^+\mathbb{1}_{z>0} + r_m^-\mathbb{1}_{z<0}\right)dz.
\end{aligned}$$

The second term appears because the Lévy measure $\tilde{\nu}$ is allowed to be asymmetric. In its expression, we used that $\xi(z) = z$ for $z \in (-1,1)$, and denote

$$\begin{aligned}
\overline{\xi}_u &= \int_{|z|\geq\frac{1}{u}}[u\xi(z) - \xi(uz)]\,\tilde{\nu}(dz) \\
&= u\int_{|z|\geq 1}\xi(z)\,\tilde{\nu}(dz) + u\int_{\frac{1}{u}\leq|z|<1} z\,\tilde{\nu}(dz) - \int_{|z|\geq\frac{1}{u}}\xi(uz)\,\tilde{\nu}(dz). \qquad (6.30)
\end{aligned}$$

Hence, by inverting the Fourier transform,

$$\begin{aligned}
&\partial_{\theta_j}\mathbb{E}_\theta f(u\tilde{X}_h) \\
&= h\,\mathbb{E}_\theta\left[\partial_{\theta_j}\left(\frac{\sigma^2u^2}{2}f'' + \sum_{m=1}^M u^{\alpha_m}(r_m^+\mathcal{J}_{\alpha_m}^+ f + r_m^-\mathcal{J}_{\alpha_m}^-) - \overline{\xi}_u f'\right)(u\tilde{X}_h)\right]. \qquad (6.31)
\end{aligned}$$

So far, we assumed f to be a Schwartz function, but the right hand side of (6.31) makes sense whenever $f \in \mathcal{C}^2$. We can extend the whole equation (6.31) to this case by approximating f suitably with a sequence of Schwartz functions f_n, such that $\sup_{|x|\leq K} |f_n^{(k)}(x) - f^{(k)}(x)| \to 0$ as $n \to \infty$ for each $K > 0$, and $k = 0, 1, 2$, and $\sup_n \|f_n^{(k)}\|_\infty < \infty$. Hence, standard arguments allow us to pass to the limit on both sides of the equation (6.31)

To handle the asymmetry term $\bar{\xi}_u$, we exploit (6.30) to derive

$$\begin{aligned}\left|\partial_{r_m^\pm} \bar{\xi}_u\right| &\leq u\|\xi\|_\infty \int_1^\infty \alpha_m |z|^{-1-\alpha_m} dz + u \int_{\frac{1}{u}}^1 \alpha_m |z|^{-\alpha_m}\, dz \\ &\quad + \|\xi\|_\infty \int_{\frac{1}{u}}^\infty \alpha_m |z|^{-1-\alpha_m} dz \\ &\leq u\|\xi\|_\infty + u\left|\int_{\frac{1}{u}}^1 \alpha_m |z|^{-\alpha_m}\, dz\right| + \|\xi\|_\infty u^{\alpha_m}.\end{aligned}$$

The second integral can be bounded as follows. For any $\epsilon \in (0,1)$ and any $p \neq 1$, there is a $\tilde{p}$ between p and 1 such that

$$\begin{aligned}\left|\int_\epsilon^1 |z|^{-p}\, dz\right| &= \frac{1}{|1-p|}|\epsilon^{1-p} - \epsilon^0| \\ &= \frac{|1-p|}{|1-p|}|\epsilon^{1-\tilde{p}} \log(\epsilon)| \quad \leq |\log \epsilon| \epsilon^{(1-p)\wedge 0}.\end{aligned}$$

By continuity, the same bound holds for $p = 1$. Thus, we obtain

$$\begin{aligned}\left|\partial_{r_m^\pm} \bar{\xi}_u\right| &\leq u\|\xi\|_\infty + \alpha_m |\log u| u^{\alpha_m \vee 1} + \|\xi\|_\infty u^{\alpha_m} \\ &\leq \tilde{C} u^{\alpha_m \vee 1}(1 + |\log u|).\end{aligned}$$

Similarly,

$$\begin{aligned}\left|\partial_{\alpha_m} \bar{\xi}_u\right| &\leq u\|\xi\|_\infty (r_m^+ + r_m^-) \int_1^\infty \frac{\alpha_m |\log z| + 1}{|z|^{1+\alpha_m}}\, dz \\ &\quad + u(r_m^+ + r_m^-) \int_{\frac{1}{u}}^1 \frac{\alpha_m |\log z| + 1}{|z|^{\alpha_m}}\, dz \\ &\quad + \|\xi\|_\infty (r_m^+ + r_m^-) \int_{\frac{1}{u}}^\infty \frac{\alpha_m |\log z| + 1}{|z|^{1+\alpha_m}}\, dz \\ &\leq \tilde{C}(1 + |\log u|)^2 u^{\alpha_m \vee 1}.\end{aligned}$$

Note also that $\partial_{\sigma^2} \overline{\xi}_u = 0$.

For specific partial derivatives, we thus have shown that

$$\begin{aligned}\partial_{\sigma^2}\mathbb{E}_\theta f(u\tilde{X}_h) &= h\frac{u^2}{2}\mathbb{E}_\theta f''(u\tilde{X}_h),\\ \partial_{r_m^\pm}\mathbb{E}_\theta f(u\tilde{X}_h) &= hu^{\alpha_m}\mathbb{E}_\theta \mathcal{J}^\pm_{\alpha_m} f(u\tilde{X}_h) + \mathcal{O}\left(hu^{\alpha_m\vee 1}\log u\right)\mathbb{E}_\theta f'(u\tilde{X}_h),\\ \partial_{\alpha_m}\mathbb{E}_\theta f(u\tilde{X}_h) &= hu^{\alpha_m}\mathbb{E}_\theta\left(\frac{d}{d\alpha_m}(r_m^+\mathcal{J}^+_{\alpha_m}f + r_m^-\mathcal{J}^-_{\alpha_m}f)(u\tilde{X}_h)\right)\\ &\quad + hu^{\alpha_m}\log u\mathbb{E}_\theta\left((r_m^+\mathcal{J}^+_{\alpha_m}f + r_m^-\mathcal{J}^-_{\alpha_m}f)(u\tilde{X}_h)\right)\\ &\quad + \mathcal{O}\left(hu^{\alpha_m\vee 1}(\log u)^2\right)\mathbb{E}_\theta f'(u\tilde{X}_h).\end{aligned} \tag{6.32}$$

For fixed f such that $\|f^{(k)}\|_\infty < \infty, k \leq 3$, the three functions f'', $\mathcal{J}^\pm_{\alpha_m} f$, and $\partial_{\alpha_m}\mathcal{J}^\pm_{\alpha_m} f$, are bounded and Lipschitz continuous, uniformly on compacts in θ. Moreover, $P_\theta(|u\tilde{X}_h| > \eta) \to 0$ uniformly on compacts in Θ for any η, as established in the proof of Lemma 6.1. Therefore, $\mathbb{E}_\theta f''(u\tilde{X}_h) \to f''(0)$ uniformly on compacts as $h \to 0$, as well as $\mathbb{E}_\theta\mathcal{J}^\pm_{\alpha_m} f(u\tilde{X}_h) \to \mathcal{J}^\pm_{\alpha_m} f(0)$ and $\mathbb{E}_\theta\partial_{\alpha_m}\mathcal{J}^\pm_{\alpha_m} f(u\tilde{X}_h) \to \partial_{\alpha_m}\mathcal{J}^\pm_{\alpha_m} f(0)$. This completes the proof of (6.3), and (6.4) follows analogously by applying a linear transformation to (6.32). Finally, (6.5) is a consequence of (6.32) upon noting that $\mathbb{E}_\theta f''(u\tilde{X}_h) = \mathcal{O}(hu^\alpha)$, see Lemma 6.1. ☐

Proof of Corollary 6.6. Since f_1' is bounded, (6.3) shows that

$$|\partial_{r_m^\pm}\mathbb{E}_\theta f_1(u\tilde{X}_h)| = o(hu^2), \quad |\partial_{\alpha_m}\mathbb{E}_\theta f_1(u\tilde{X}_h)| = o(hu^2).$$

This corresponds to the entries $A(\theta)_{1,k} = 0$ for $k \geq 2$. For $j \geq 2$, we have $\mathbb{E}_\theta f_j'(u\tilde{X}_h) = \mathcal{O}(hu^\alpha)$ by virtue of Lemma 6.1, since f_j vanishes near zero. Hence, since $\alpha_m > \alpha/2$ and $u \leq \mathcal{O}(\sqrt{h})$,

$$\mathcal{O}\left(hu^{\alpha_m\vee 1}(\log u)^2\right)\mathbb{E}_\theta f_j'(u\tilde{X}_h) = \mathcal{O}(h^2u^{\alpha+(\alpha_m\vee 1)}(\log(u)^2)) \leq o(hu^{\alpha_m}).$$

This corresponds to the entries $A(\theta)_{j,k} = 0$ for $j \geq 2$, $k \geq 2$. Moreover, $\partial_{\sigma^2}\mathbb{E}_\theta f_j(u\tilde{Z}_h) = o(hu^\alpha)$. In combination with Lemma 6.5, this suffices to establish the convergence (6.6). ☐

Proof of Lemma 6.7. Denote the estimating equation (2.5) as $F_n(\hat{\theta}_n) = 0$, for

$$F_n(\theta) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{f}(u_n\Delta_{n,i}X) - \mathbb{E}_{\hat{\theta}_n}\boldsymbol{f}(u_n\tilde{Z}_{h_n}). \tag{6.33}$$

Let θ_0 be the true parameters, and reparameterize $\theta = \theta_0 + \Gamma_n(\theta_0)\bar{\Lambda}_n^{-1}(\theta_0)T$ for $T = \bar{\Lambda}_n(\theta_0)\Gamma_n^{-1}(\theta_0)(\theta - \theta_0)$, and let

$$\bar{F}_n(T) = \tilde{\Lambda}_n^{-1}(\theta_0)F_n\left(\theta_0 + \Gamma_n(\theta_0)\bar{\Lambda}_n^{-1}(\theta_0)T\right).$$

This is well defined whenever $T \in B_{d_n}(0)$, for $d_n = c\sqrt{h_n}u_n^{\alpha_M - \frac{\alpha_1}{2}}/(\log u_n)^3 \to 0$, and $c > 0$ sufficiently small. In this reparameterized model, we need to show that there exists a sequence of random vectors $\hat{T}_n \in B_{d_n}$ such that $\bar{F}_n(\hat{T}_n) = 0$ for large n. This will imply that $\|\hat{\theta}_n - \theta_0\| \leq C/(\log u_n)^2$ for a sufficiently large factor C.

We know from Lemma 6.4 that

$$\bar{F}_n(0) = \tilde{\Lambda}_n^{-1}(\theta_0)F_n(\theta_0) = \mathcal{O}_P\left(\frac{1}{\sqrt{n}}\right) = o(d_n).$$

Furthermore,

$$\begin{aligned}\mathrm{D}_T\bar{F}_n(T) &= \begin{pmatrix}\partial_{T_1} & \dots & \partial_{T_{3M+1}}\end{pmatrix} \bar{F}_n(T) \\ &= \tilde{\Lambda}_n^{-1}(\theta_0)\mathrm{D}_\theta F_n(\theta_0 + \Gamma_n\bar{\Lambda}_n^{-1}T)\Gamma_n(\theta_0)\bar{\Lambda}_n^{-1}(\theta_0).\end{aligned}$$

By Corollary 6.6, $\tilde{\Lambda}_n^{-1}(\theta)\mathrm{D}_\theta F_n(\theta)\Gamma_n(\theta)\bar{\Lambda}_n^{-1}(\theta) \to A(\theta)$ locally uniformly, and it can be checked that $\theta \mapsto A(\theta)$ is continuous. Moreover, the definitions of $\tilde{\Lambda}_n, \bar{\Lambda}_n$, and Γ_n readily yield, as $n \to \infty$,

$$\begin{aligned}&\sup_{T\in\overline{B}_{d_n}(0)} \|\tilde{\Lambda}_n^{-1}(\theta_0)\tilde{\Lambda}_n(\theta_0 + \Gamma_n(\theta_0)\bar{\Lambda}_n^{-1}(\theta_0)T) - \mathbf{I}_{3M+1}\| \\ &\leq \sup_{\|\theta-\theta_0\|\leq\frac{C}{(\log u_n)^2}} \|\tilde{\Lambda}_n^{-1}(\theta_0)\tilde{\Lambda}_n(\theta) - \mathbf{I}_{3M+1}\| \to 0, \\ &\sup_{T\in\overline{B}_{d_n}(0)} \|\bar{\Lambda}_n^{-1}(\theta_0)\bar{\Lambda}_n(\theta_0 + \Gamma_n(\theta_0)\bar{\Lambda}_n^{-1}(\theta_0)T) - \mathbf{I}_{3M+1}\| \to 0, \\ &\sup_{T\in\overline{B}_{d_n}(0)} \|\Gamma_n^{-1}(\theta_0 + \Gamma_n(\theta_0)\bar{\Lambda}_n^{-1}(\theta_0)T)\Gamma_n(\theta_0) - \mathbf{I}_{3M+1}\| \to 0.\end{aligned} \tag{6.34}$$

Here, we denote by $\|\cdot\|$ the spectral norm of a matrix, i.e. $\|A\|^2$ is the largest absolute eigenvalue of the symmetrized matrix A^TA, and $\mathbf{I}_d$ denotes the $d \times d$ identity matrix. Thus,

$$\sup_{T\in\overline{B}_{d_n}(0)} \|\mathrm{D}_T\bar{F}_n(T) - A(\theta_0)\| \to 0.$$

Now we apply (Jacod and Sørensen, 2018, Lemma 6.2) to establish the existence of a solution $\hat{T}_n \in B_{d_n}(0)$ of the equation $\bar{F}_n(\hat{T}_n) = 0$. Let $\lambda = \frac{1}{2}\|A(\theta_0)^{-1}\|^{-1}$, and denote by C_n the event

$$C_n = \left\{\sup_{T\in\overline{B}_{d_n}(0)} \|\mathrm{D}_T\bar{F}_n(T) - A(\theta_0)\| \leq \lambda\right\} \quad \cap \quad \left\{\left\|\bar{F}_n(0)\right\| \leq \lambda d_n\right\}.$$

Since the first set is deterministic, and since $\|\bar{F}_n(0)\|/d_n \xrightarrow{P} 0$, we have $P(C_n) \to 1$. On the set C_n, it holds that $0 \in \overline{B}_{\lambda d_n}(\bar{F}_n(0))$. Then Lemma 6.2 of Jacod and Sørensen (2018) with $y = 0, f = \bar{F}_n$ and $r = d_n$, states that there exists a unique point $\hat{T}_n \in \overline{B}_{d_n}(0)$ which solves $\bar{F}_n(\hat{T}_n) = 0$.

Returning to the original parametrization, we conclude there exists a random variable $\hat{\theta}_n$ such that with probability at least $P(C_n) \to 1$, $\hat{\theta}_n$ solves the estimating equation and $\hat{\theta}_n - \theta_0 \in \Gamma_n(\theta_0)\bar{\Lambda}_n^{-1}(\theta_0)\overline{B}_{d_n}(0)$, i.e.

$$\|\hat{\theta}_n - \theta_0\| = \mathcal{O}_P(1/|\log u_n|^2). \tag{6.35}$$

Now let $\hat{\theta}_n^*$ be another sequence of random variables satisfying $F_n(\hat{\theta}_n^*) = 0$, and such that $\|\hat{\theta}_n^* - \theta_0\| = \mathcal{O}_P(1/|\log u_n|^2)$. Theorem 2.1 below establishes that this consistent sequence $\hat{\theta}_n^*$ indeed converges at a faster polynomial rate, which in particular yields $\|\hat{\theta}_n^* - \theta_0\| = o_P(1/|\log u_n|^2)$. Thus, $\hat{T}_n^* = \Gamma_n(\theta_0)^{-1}\bar{\Lambda}_n^{-1}(\theta_0)(\hat{\theta}_n^* - \theta_0) \in \overline{B}_{d_n}(0)$ eventually. Hence, the uniqueness of $\hat{T}_n$ on $\overline{B}_{d_n}(0)$ implies the $\hat{T}_n = \hat{T}_n^*$ and thus the uniqueness of $\hat{\theta}_n$, i.e. $P(\hat{\theta}_n^* \neq \hat{\theta}_n) = P(\hat{T}_n^* \neq \hat{T}_n) \to 0$. □

Proof of Theorem 2.1. Denote the estimating equation as $F_n(\theta) = 0$, for $F_n(\theta)$ as in (6.33). The mean value theorem yields

$$0 = \tilde{\Lambda}_n^{-1}(\theta_0)F_n(\hat{\theta}_n) = \tilde{\Lambda}_n^{-1}(\theta_0)F_n(\theta_0) + \left[\tilde{\Lambda}_n^{-1}\widetilde{F_n}\Gamma_n\bar{\Lambda}_n^{-1}\right]\bar{\Lambda}_n\Gamma_n^{-1}(\hat{\theta}_n - \theta_0),$$

where $(\widetilde{F}_n)_{j,k} = \partial_{\theta_k}(F_n)_j(\tilde{\theta}^j)$ for some $\tilde{\theta}^j$ on the line segment between θ_0 and $\hat{\theta}_n$. Denote by $R_n \subset \Omega$ the event that $A_n = \tilde{\Lambda}_n(\theta_0)^{-1}\widetilde{F_n}\Gamma_n(\theta_0)\bar{\Lambda}_n(\theta_0)^{-1}$ is regular, and introduce furthermore the matrices

$$A_n^j = \tilde{\Lambda}(\theta_0)^{-1}\mathrm{D}_\theta F_n(\tilde{\theta}^j)\Gamma_n(\theta_0)\bar{\Lambda}_n(\theta_0)^{-1}, \quad j = 1, \dots, 3M+1.$$

That is, the j-th row of A_n and A_n^j coincide, $(A_n)_{j,k} = (A_n^j)_{j,k}$. Now note that $\|\tilde{\theta}^j - \theta_0\| \leq \|\hat{\theta}_n - \theta_0\| = \mathcal{O}_P(1/(\log u_n)^2)$, by (6.35). Moreover, for any $C > 0$, as in (6.34),

$$\begin{aligned} &\sup_{\|\theta-\theta_0\|\leq \frac{C}{(\log u_n)^2}} \|\tilde{\Lambda}_n^{-1}(\theta_0)\tilde{\Lambda}_n(\theta) - \mathbf{I}_{3M+1}\| \\ &\sup_{\|\theta-\theta_0\|\leq \frac{C}{(\log u_n)^2}} \|\bar{\Lambda}_n^{-1}(\theta_0)\bar{\Lambda}_n(\theta) - \mathbf{I}_{3M+1}\| \\ &\sup_{\|\theta-\theta_0\|\leq \frac{C}{(\log u_n)^2}} \|\Gamma_n^{-1}(\theta_0)\Gamma_n(\theta) - \mathbf{I}_{3M+1}\|. \end{aligned} \tag{6.36}$$

Together with the locally uniform convergence of Corollary 6.6, this yields $A_n^j \xrightarrow{P} A(\theta_0)$ for each j, and thus $A_n \xrightarrow{P} A(\theta_0)$.

In particular, $P(R_n) \to 1$, and on the set R_n, we may rewrite

$$\sqrt{n}\bar{\Lambda}_n\Gamma_n^{-1}(\theta_0)(\hat{\theta}_n - \theta_0) = -\sqrt{n}A_n^{-1}\tilde{\Lambda}_n^{-1}F_n(\theta_0).$$

But $\sqrt{n}\tilde{\Lambda}_n^{-1}F_n(\theta_0) \Rightarrow \mathcal{N}(0, \Sigma(\theta_0))$ by Lemma 6.4, and $A_n^{-1} \to A^{-1}(\theta)$ in probability, such that Slutsky's lemma completes the proof. □

Proof of Proposition 3.1. We show how to adjust the proof of Aït-Sahalia and Jacod (2012) to consider the off-diagonal entries. Denote by φ_α the density of a symmetric α-stable random variable, standardized to have Lévy measure $\alpha|x|^{-1-\alpha}dx$. This is the same parametrization as implied by (2.3). Furthermore, let φ be the density of a standard normal distribution. Then the probability density of $\tilde{Z}_h$ is given by the convolution

$$p_h(x) = \int \frac{1}{\sqrt{\sigma^2 h}}\varphi\left(\frac{x - (rh)^{\frac{1}{\alpha}}y}{\sqrt{\sigma^2 h}}\right)\varphi_\alpha(y)\,dy.$$

Now introduce the terms

$$w_h = (rh)^{\frac{1}{\alpha}}/\sqrt{\sigma^2 h}, \qquad v_h = \frac{1}{\alpha(2-\alpha)}\left(2 + \frac{\log(r/\sigma^2)}{\log(1/w_h)}\right),$$

and

$$\begin{aligned}
S_h(x) &= \int \varphi(x - w_h y)\varphi_\alpha(y)\, dy = \sqrt{\sigma^2 h}\cdot p_h(x\sqrt{\sigma^2 h}),\\
R_h^0(x) &= \frac{1}{w_h^\alpha}\int \varphi(x - w_h y)(\varphi_\alpha(y) + y\partial_y\varphi_\alpha(y))\, dy\\
&= \frac{-r\alpha\sqrt{\sigma^2 h}}{w_h^\alpha}\,\frac{d}{dr} p_h(x\sqrt{\sigma^2 h}),\\
R_h^1(x) &= \frac{1}{w_h^\alpha \log(1/w_h)}\int \varphi(x - w_h y)\partial_\alpha\varphi_\alpha(y)\, dy,\\
\rightsquigarrow &\qquad w_h^\alpha\log(1/w_h)R_h^1(x) - w_h^\alpha v_h\log(1/w_h)R_h^0(x)\\
&= \sqrt{\sigma^2 h}\frac{d}{d\alpha} p_h(x\sqrt{\sigma^2 h}),\\
J_h^{l,m} &= \int \frac{R_h^l(x)R_h^m(x)}{S_h(x)}dx, \qquad l,m\in\{0,1\}.
\end{aligned}$$

Some technical integral transformations, explained in more detail by Aït-Sahalia and Jacod (2012) (cf. (A.3) therein), establish that

$$\begin{aligned}
\mathcal{I}_h^{r,r} &= \frac{w_h^{2\alpha}}{r^2\alpha^2}J_h^{0,0},\\
\mathcal{I}_h^{\alpha,\alpha} &= \int \frac{w_h^{2\alpha}\log(1/w_h)^2(R_h^1(x) - v_h R_h^0(x))^2}{S_h(x)}dx\\
&= w_h^{2\alpha}\log(1/w_h)^2(J_h^{1,1}(x) - 2v_h J_h^{1,0}(x) + v_h^2 J_h^{0,0}(x)),\\
\mathcal{I}_h^{\alpha,r} &= \int \frac{w_h^{2\alpha}\frac{-R_h^0(x)}{r\alpha}\log(1/w_h)\left(R_h^1(x) - v_h R_h^0(x)\right)}{S_h(x)}\\
&= \frac{w_h^{2\alpha}\log(1/w_h)}{r\alpha}\left(v_h J_h^{0,0}(x) - J_h^{1,0}(x)\right).
\end{aligned}$$

The main workload of the proof given by Aït-Sahalia and Jacod (2012) derives the limiting behavior of $J_h^{l,m}$ as $h\to 0$. They show that

$$J_h^{0,0}/\psi_h \to \alpha^4, \quad J_h^{1,0}\to\alpha^3, \quad J_h^{1,1}\to\alpha^2,$$

where

$$\psi_h = \frac{2\sigma^\alpha}{r\alpha^2(2-\alpha)^{\frac{\alpha}{2}}}\,\frac{1}{h^{1-\frac{\alpha}{2}}\log(1/h)^{\frac{\alpha}{2}}}.$$

Using furthermore that $v_h \to \frac{2}{\alpha(2-\alpha)}$, this yields

$$\begin{pmatrix} \frac{r\alpha}{w_h^\alpha\sqrt{\psi_h}} & 0\\ 0 & \frac{1}{w_h^\alpha\log(1/w_h)\sqrt{\psi_h}}\end{pmatrix}\begin{pmatrix}\mathcal{I}_h^{r,r} & \mathcal{I}_h^{r,\alpha}\\ \mathcal{I}_h^{r,\alpha} & \mathcal{I}_h^{\alpha,\alpha}\end{pmatrix}\begin{pmatrix} \frac{r\alpha}{w_h^\alpha\sqrt{\psi_h}} & 0\\ 0 & \frac{1}{w_h^\alpha\log(1/w_h)\sqrt{\psi_h}}\end{pmatrix}$$

$$\longrightarrow \quad \begin{pmatrix} \alpha^4 & \frac{\alpha^4}{2-\alpha} \\ \frac{\alpha^4}{2-\alpha} & \frac{\alpha^4}{(2-\alpha)^2} \end{pmatrix}.$$

Some straightforward manipulations show that

$$\begin{aligned} &\frac{(h\log(1/h))^{\frac{\alpha}{2}}}{h} \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\log(1/h)} \end{pmatrix} \begin{pmatrix} \mathcal{I}_h^{r,r} & \mathcal{I}_h^{r,\alpha} \\ \mathcal{I}_h^{r,\alpha} & \mathcal{I}_h^{\alpha,\alpha} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\log(1/h)} \end{pmatrix} \\ \longrightarrow &\frac{2r}{\sigma^\alpha \alpha^2 (2-\alpha)^{\frac{\alpha}{2}}} \begin{pmatrix} \frac{\alpha^2}{r^2} & \frac{\alpha^4}{2-\alpha}\frac{1}{r\alpha}\frac{2-\alpha}{2\alpha} \\ \text{sym} & \frac{(2-\alpha)^2}{4\alpha^2}\frac{\alpha^4}{(2-\alpha)^2} \end{pmatrix} \\ = &\frac{2r}{\sigma^\alpha (2-\alpha)^{\frac{\alpha}{2}}} \begin{pmatrix} \frac{1}{r^2} & \frac{1}{2r} \\ \frac{1}{2r} & \frac{1}{4} \end{pmatrix} \end{aligned}$$

This limiting matrix is singular. The off-diagonal entry $\mathcal{I}_h^{\alpha,r}$ has not been considered by Aït-Sahalia and Jacod (2012). □

Proof of Proposition 3.2. Denote the true parameter by $\alpha_{0,m}$ and $r^\pm_{0,m}$, respectively. By Lemma 6.5, we have as $n \to \infty$, $h = 1/n \to 0$,

$$\begin{aligned} \frac{1}{hu_m^\alpha \log u} \partial_{\alpha_m} \tilde{F}_n(\alpha_m) &\to r_m^+ \mathcal{J}^+_{\alpha_m} f(0) + r_m^- \mathcal{J}^-_{\alpha_m} f(0), \\ \frac{1}{hu^{\alpha_m}} \partial_{r_m^\pm} \tilde{F}_n(r_m^\pm) &\to \mathcal{J}^\pm_{\alpha_m} f(0). \end{aligned}$$

This convergence holds uniformly on compacts in Θ. The limits are positive because $r_m^+ + r_m^- > 0$ by the definition of Θ, and $\mathcal{J}^\pm_{\alpha_m} f(0) > 0$ by assumption. Moreover, Lemma 6.4 also holds for $\tilde{F}_n$, i.e.

$$nu_n^{-\alpha_1/2} \tilde{F}_n(\theta_0) \Rightarrow \mathcal{N}\left(0, (r_1^+ \mathcal{J}_{\alpha_1} + r_1^- \mathcal{J}_{\alpha_1}) f^2(0)\right). \tag{6.37}$$

Thus, the existence of a consistent sequence of estimators follows along the same lines as Lemma 6.7.

For the central limit theorem, we use the mean value theorem to obtain, for a value $\tilde{\alpha}_m$ between $\alpha_{0,m}$ and $\hat{\alpha}_m$,

$$0 = \tilde{F}_n(\hat{\alpha}_m) = \tilde{F}_n(\alpha_{0,m}) + \partial_{\alpha_m} \tilde{F}_n(\tilde{\alpha}_m)(\hat{\alpha}_m - \alpha_{0,m}).$$

In particular, $(\hat{\alpha}_m - \alpha_{0,m}) = -(\partial_{\alpha_m} \tilde{F}_n(\tilde{\alpha}_m))^{-1} \tilde{F}_n(\alpha_{0,m})$. Just as in the proof of Theorem 2.1, we may use the convergence of $\partial_{\alpha_m} \tilde{F}_n(\alpha_m)$ and the central limit theorem (6.37) to derive the asymptotic distribution of $\hat{\alpha}_m$ by means of Slutsky's Lemma. Analogously for $r_m^\pm$. □

Supplementary Material

Supplement to "Rate-optimal estimation of the Blumenthal–Getoor index of a Lévy process"
(doi: 10.1214/20-EJS1769SUPP; .zip).

References

Aït-Sahalia, Y. and Jacod, J. (2008). Fisher's information for discretely sampled Lévy processes. *Econometrica* **76** 727–761. MR2433480

Aït-Sahalia, Y. and Jacod, J. (2009). Estimating the degree of activity of jumps in high frequency data. *The Annals of Statistics* **37** 2202–2244. MR2543690

Aït-Sahalia, Y. and Jacod, J. (2012). Identifying the successive Blumenthal–Getoor indices of a discretely observed process. *The Annals of Statistics* **40** 1430–1464. MR3015031

Amorino, C. and Gloter, A. (2020a). Contrast function estimation for the drift parameter of ergodic jump diffusion process. *Scandinavian Journal of Statistics* **47** 279–346. MR4157142

Amorino, C. and Gloter, A. (2020b). Unbiased truncated quadratic variation for volatility estimation in jump diffusion processes. *Stochastic Processes and their Applications* **130** 5888–5939. MR4140022

Andersen, T. G., Benzoni, L. and Lund, J. (2002). An empirical investigation of continuous-time equity return models. *The Journal of Finance* **57** 1239–1284.

Barboza, L. A. and Viens, F. G. (2017). Parameter estimation of Gaussian stationary processes using the generalized method of moments. *Electronic Journal of Statistics* **11** 401-439. MR3611508

Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of statistics* **24** 1–13. MR1436619

Blumenthal, R. M. and Getoor, R. K. (1961). Sample functions of stochastic processes with stationary independent increments. *Journal of Mathematics and Mechanics* **10** 493–516. MR0123362

Brouste, A. and Fukasawa, M. (2018). Local asymptotic normality property for fractional Gaussian noise under high-frequency observations. *The Annals of Statistics* **46** 2045–2061. MR3845010

Brouste, A. and Masuda, H. (2018). Efficient estimation of stable Lévy process with symmetric jumps. *Statistical Inference for Stochastic Processes* **21** 289–307. MR3824969

Bull, A. D. (2016). Near-optimal estimation of jump activity in semimartingales. *The Annals of Statistics* **44** 58–86. MR3449762

Christensen, K., Oomen, R. C. A. and Podolskij, M. (2014). Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics* **114** 576–599.

Clément, E. and Gloter, A. (2019). Estimating functions for SDE driven by stable Lévy processes. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* **55** 1316–1348. MR4010937

Cont, R. and Tankov, P. (2004). *Financial modelling with jump processes*. Chapman and Hall/CRC. MR2042661

Durrett, R. (2005). *Probability: Theory and Examples*. Brooks/Cole – Thomson Learning, Belmont, USA. MR1068527

Figueroa-López, J. E. and Mancini, C. (2019). Optimum thresholding using mean and conditional mean squared error. *Journal of Econometrics* **208** 179–210. MR3906967

Figueroa-López, J. E. and Nisen, J. (2013). Optimally thresholded realized power variations for Lévy jump diffusion models. *Stochastic Processes and their Applications* **123** 2648–2677. MR3054540

Gloter, A., Loukianova, D. and Mai, H. (2018). Jump filtering and efficient drift estimation for Lévy-driven SDEs. *The Annals of Statistics* **46** 1445–1480. MR3819106

Jacod, J. and Reiss, M. (2014). A remark on the rates of convergence for integrated volatility estimation in the presence of jumps. *The Annals of Statistics* **42** 1131–1144. MR3224283

Jacod, J. and Sørensen, M. (2018). A review of asymptotic theory of estimating functions. *Statistical Inference for Stochastic Processes* **21** 415–434. MR3824976

Jacod, J. and Todorov, V. (2014). Efficient estimation of integrated volatility in presence of infinite variation jumps. *The Annals of Statistics* **42** 1029–1069. MR3210995

Jacod, J. and Todorov, V. (2016). Efficient Estimation of Integrated Volatility in Presence of Infinite Variation Jumps with Multiple Activity Indices. In *The Fascination of Probability, Statistics and their Applications* 317–341. Springer International Publishing. MR3495691

Jakobsen, N. M. and Sørensen, M. (2019). Estimating functions for jump-diffusions. *Stochastic Processes and their Applications* **129** 3282–3318. MR3985563

Jing, B.-Y., Kong, X.-B. and Liu, Z. (2011). Estimating the jump activity index under noisy observations using high-frequency data. *Journal of the American Statistical Association* **106** 558–568. MR2847970

Jing, B.-Y., Kong, X.-B., Liu, Z. and Mykland, P. (2012). On the jump activity index for semimartingales. *Journal of Econometrics* **166** 213–223. MR2862961

Johnson, S. G. (2020). The NLopt nonlinear-optimization package. `http://github.com/stevengj/nlopt`.

Mancini, C. (2006). Estimating the integrated volatility in stochastic volatility models with Lévy type jumps Technical Report, Università di Firenze.

Mancini, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics* **36** 270–296. MR2528985

Mariucci, E. and Reiss, M. (2018). Wasserstein and total variation distance between marginals of Lévy processes. *Electronic Journal of Statistics* **12** 2482–2514. MR3833470

Masuda, H. (2015). Parametric estimation of Lévy processes. In *Levy Matters IV* 179–286. MR3364258

Mies, F. (2020). Supplement to “Rate-optimal estimation of the Blumenthal–Getoor index of a Lévy process”. DOI:10.1214/20-EJS1769SUPP.

Mies, F. (2020). High-frequency inference for stochastic processes with jumps of infinite activity. RWTH Aachen University. DOI:10.18154/RWTH-2020-07022.

Mies, F. (2021). Estimation of state-dependent jump activity and drift for Markovian semimartingales. *Journal of Statistical Planning and Inference* **210** 114–140. MR4102181

Reiss, M. (2013). Testing the characteristics of a Lévy process. *Stochastic Processes and their Applications* **123** 2808–2828. MR3054546

Rydberg, T. H. (1997). The normal inverse Gaussian Lévy process: simulation and approximation. *Communications in statistics. Stochastic models* **13** 887–910. MR1482297

Todorov, V. (2015). Jump activity estimation for pure-jump semimartingales via self-normalized statistics. *The Annals of Statistics* **43** 1831–1864. MR3357880

Zolotarev, V. M. (1986). *One-dimensional Stable Distributions*. American Mathematical Society. MR0854867