Estimation of the characteristics of a Lévy process observed at arbitrary frequency

Johanna Kappus

Institute of Mathematics
Humboldt-Universität zu Berlin
`kappus@math.hu-berlin.de`

Markus Reiß

Institute of Mathematics
Humboldt-Universität zu Berlin
`mreiss@math.hu-berlin.de`

January 14, 2010

Abstract

A Lévy process is observed at time points of distance Δ until time T. We construct an estimator of the Lévy-Khinchine characteristics of the process and derive optimal rates of convergence simultaneously in T and Δ. Thereby, we encompass the usual low- and high-frequency assumptions and obtain also asymptotics in the mid-frequency regime.

1   Introduction

Lévy processes are the main building blocks for stochastic continuous-time jump models, which become more and more popular in applications. One important task is thus to provide estimation methods for the characteristics of a Lévy process.

There exist two fundamentally different estimation approaches, depending on the nature of observations. If we can assume high-frequency observations of the Lévy process, we can discretize a natural estimator based on continuous-time observations, where the jumps and the diffusion part are observed directly [9, 7, 5]. Alternatively, the low-frequency setting is considered where the observation distance does not tend to zero and even asymptotically we cannot observe the diffusion and the jumps directly. Not surprisingly, in that case we face a more complicated inference problem leading to a deconvolution-type inverse problem [14, 4, 12, 11]. A very similar structure occurs in the estimation for Lévy-Ornstein-Uhlenbeck processes [13] and in the calibration of financial derivatives (European options) to Lévy models [2].

Here, we want to bridge the gap between high- and low-frequency estimation methods by allowing the time distance Δ between observations to remain constant or to converge to zero at an arbitrary speed. First results into that direction have been obtained by [4] for specific models. In any case,

1

the observation time T tends to infinity because only this allows identification of the drift and the jump part in the limit. We extend the approach for general Lévy processes by [14] to arbitrary observation distances Δ.

First we introduce the setup in Section 2. Then in Section 3 we propose our estimator based on a minimum-distance criterion. The correct distance relies upon uniform convergence properties of the empirical characteristic function. The main result is an asymptotic upper bound for the estimator of the jump measure. Particularly interesting is the fact that we recover simultaneously the convergence rates for the high- and low-frequency setup, without any prescription for the estimator. As a minimax lower bound proves, also our intermediate (mid-frequency) risk bounds are asymptotically optimal. All proofs are postponed to Section 4.

2 Statistical model and estimation strategy

A Lévy process $(X_t,\, t \geqslant 0)$ is observed at the n equidistant time points $\Delta, \cdots, n\Delta = T$. It is well known that the characteristic function of X_Δ has the form

$$\varphi_\Delta(u) = \mathbb{E}\left[e^{iuX_\Delta}\right] = e^{\Delta\Psi(u)},$$

where the characteristic exponent Ψ reads as

$$\Psi(u) = iu\bar{b} - \frac{\sigma^2}{2}u^2 + \int \left(e^{iux} - 1 - iux1(|x| \leq 1)\right)\nu(\,\mathrm{d}x),$$

with volatility $\sigma \geq 0$, drift $\bar{b} \in \mathbb{R}$ and jump measure ν, where ν is a σ-finite Borel measure on $\mathbb{R}$ with $\int_{\mathbb{R}\setminus\{0\}}(x^2 \wedge 1)\nu(\,\mathrm{d}x) < \infty$. Throughout the text we shall assume that X_1 has finite moments up to order $4 + \delta$ for some positive constant δ. Then we even have (cf. Thm. 25.3 [15])

$$\int_{\mathbb{R}\setminus\{0\}} x^2\nu(\,\mathrm{d}x) < \infty.$$

We can thus give the following reparametrization of the characteristic exponent in terms of the finite measure $\nu_\sigma(\,\mathrm{d}x) := \sigma^2\delta_0(\,\mathrm{d}x) + x^2\nu(\,\mathrm{d}x)$ and $b := \bar{b} + \int x\,1(|x| > 1)\nu(\,\mathrm{d}x)$:

$$\Psi(u) = iub + \int_{\mathbb{R}} \frac{\left(e^{iux} - 1 - iux\right)}{x^2}\nu_\sigma(\,\mathrm{d}x),$$

where the integrand is continuously extended to $-u^2/2$ at $x = 0$. The Lévy process is fully described by the parameters b (which is equal to the mean value of X_1) and ν_σ. The motivation for considering the above parametrization comes from the following fundamental result (see e.g. Theorem 8.7 in [15]):

2

2.1 Proposition. *Let $\mathbb{P}_{(b,\nu_\sigma)}$ and $\left(\mathbb{P}_{(b_n,\nu_{\sigma_n})}\right)_{n\in\mathbb{N}}$ denote infinitely divisible laws with the corresponding characteristics. Then weak convergence $\mathbb{P}_{(b_n,\nu_{\sigma_n})} \Rightarrow \mathbb{P}_{(b,\nu_\sigma)}$ takes place if and only if $b_n \to b$ and $\nu_{\sigma_n} \Rightarrow \nu_\sigma$.*

Using the fact that the increments of a Lévy process are independent and identically distributed, we can define the empirical characteristic function

$$\hat{\varphi}_{\Delta,T}(u) := \frac{1}{n} \sum_{k=1}^{n} e^{iu(X_{k\Delta} - X_{(k-1)\Delta})}. \tag{2.1}$$

Pointwise convergence of $\hat{\varphi}_{\Delta,T}$ to φ_Δ suggests to choose the estimators of the parameters of interest such that the corresponding characteristic function approximately minimizes the distance to the empirical characteristic function. Consequently, we define

$$\left(\hat{b}_{\Delta,T}, \hat{\nu}_{\sigma_{\Delta,T}}\right) := \mathrm{arginf}_{(\tilde{b},\tilde{\nu}_\sigma)} \, \mathrm{d}\left(\hat{\varphi}_{\Delta,T}, \varphi_\Delta\left(\bullet; \tilde{b}, \tilde{\nu}_\sigma\right)\right) \tag{2.2}$$

for an appropriate choice of the metric d. It was shown in [14] that for equidistant observations with Δ fixed, the estimators of b and ν_σ defined according to (2.2) are strongly consistent under rather general conditions on the choice of the metric d. Moreover, optimal rates of convergence are obtained if b and ν_σ are chosen to fit the weighted empirical characteristic function and its first and second derivative.

The motivation for considering not only the characteristic function, but also its derivatives comes from the fact that the Fourier transform of the finite measure ν_σ can be expressed as

$$\mathcal{F}\nu_\sigma(u) := \int_\mathbb{R} e^{iux} \nu_\sigma(\,\mathrm{d}x) = -\Psi''(u),$$

which gives

$$\mathcal{F}\nu_\sigma(u) = \frac{\varphi_1'(u)^2}{\varphi_1(u)^2} - \frac{\varphi_1''(u)}{\varphi_1(u)}, \tag{2.3}$$

and in terms of Δ:

$$\mathcal{F}\nu_\sigma(u) = \frac{1}{\Delta}\left(\frac{\varphi_\Delta'(u)^2}{\varphi_\Delta(u)^2} - \frac{\varphi_\Delta''(u)}{\varphi_\Delta(u)}\right) \tag{2.4}$$

Note that by formula (2.3) and (2.4) there is a strong resemblance of the problem of estimating ν_σ with a deconvolution problem. The optimal rates of convergence depend on the decay behaviour of the characteristic function.

To obtain an estimator which is rate optimal for $T \to \infty$ with arbitrary observation distance Δ, the appropriate choice of a distance function will have to depend on Δ. Because of the moment sizes $\mathbb{E}[X_\Delta^{2k}] = O(\Delta^{1\wedge k})$ (see p.9 for a proof), it turns out that the distance function

$$d_\Delta(\varphi, \psi) := \sum_{k=0}^{2} \Delta^{-\frac{1\wedge k}{2}} \| \varphi^{(k)} - \psi^{(k)} \|_{L^\infty(w)} \tag{2.5}$$

3

is appropriate, where

$$\|f\|_{L^\infty(w)} := \sup_{u \in \mathbb{R}} |f(u)| \, w(u)$$

for a weight function $w : \mathbb{R} \to \mathbb{R}^+$ specified later. Since we cannot guarantee that the infimum is always attained, our estimators $\hat{b}_{\Delta,T}$ and $\hat{\nu}_{\sigma_{\Delta,T}}$ are chosen such that

$$d_\Delta \left(\hat{\varphi}_{\Delta,T}, \varphi_\Delta \left(\bullet; \hat{b}_{\Delta,T}, \hat{\nu}_{\sigma_{\Delta,T}} \right) \right) \leqslant \inf_{(b,\nu_\sigma)} d_\Delta \left(\hat{\varphi}_{\Delta,T}, \varphi_\Delta \left(\bullet; b, \nu_\sigma \right) \right) + \varepsilon_T \quad (2.6)$$

with $\varepsilon_T = o\left(\Delta^{1/2} T^{-\frac{1}{2}} \right)$. In what follows, we will use the notation

$$\varphi_{\Delta,T} := \varphi_\Delta \left(\bullet; \hat{b}_{\Delta,T}, \hat{\nu}_{\sigma_{\Delta,T}} \right).$$

3 Rate optimality of the estimation procedure

3.1 Convergence of the empirical characteristic function

The main technical tool needed to prove rate optimality in T and Δ is the following result giving control of the weighted empirical characteristic function on the whole real line *uniformly in* Δ. In an abstract sense, the statement below will tell us that the Donsker property holds for the empirical characteristic function uniformly over the class of distributions $(\mathbb{P}_\Delta)_{\Delta \leq 1}$, where $\mathbb{P}_\Delta$ denotes the distribution of X_Δ.

Let the normalized version of the k-th derivative of the empirical characteristic function process be defined by

$$C_{\Delta,T}^{(k)}(u) := n^{-\frac{1}{2}} \Delta^{-\frac{k \wedge 1}{2}} \sum_{j=1}^n \frac{\mathrm{d}^k}{\mathrm{d}u^k} \left(e^{iu(X_{j\Delta} - X_{(j-1)\Delta})} - \mathbb{E}\left[e^{iuX_\Delta} \right] \right). \quad (3.1)$$

We can now formulate the main result of this section, which is proved in Section 4.

3.1 Theorem. *For $k \in \mathbb{N}_0$ let X be a Lévy process with finite $(2k+\gamma)$-th moment and choose $w(u) = (\log(e+ \mid u \mid))^{-1/2-\delta}$ for some constants $\gamma, \delta > 0$. Then for $C_{\Delta,T}^{(k)}$, defined by (3.1), we have*

$$\sup_{n \geq 1, \Delta \leq 1} \mathbb{E}\left[\parallel C_{\Delta,T}^{(k)} \parallel_{L^\infty(w)} \right] < \infty.$$

With the distance d_Δ defined according to (2.5), the above theorem tells us that in terms of T, the empirical characteristic function $\hat{\varphi}_{\Delta,T}$ satisfies

$$\mathbb{E}\left[\Delta^{-\frac{1}{2}} d_\Delta \left(\hat{\varphi}_{\Delta,T}, \varphi_\Delta \right) \right] = O(T^{-\frac{1}{2}}). \quad (3.2)$$

An application of the triangle inequality gives

$$d_\Delta\left(\varphi_{\Delta,T}, \varphi_\Delta\right) \leq 2d_\Delta\left(\hat{\varphi}_{\Delta,T}, \varphi_\Delta\right) + o(\Delta^{1/2}T^{-1/2}),$$

so (3.2) remains true if we replace the empirical characteristic function $\hat{\varphi}_{\Delta,T}$ by the minimum distance fit $\varphi_{\Delta,T}$.

3.2 Asymptotic risk bounds

We are now ready to prove upper bounds for convergence in probability. We consider in particular the following decay scenarios for the characteristic function:

a) The characteristic function of X_Δ satisfies

$$|\varphi_\Delta(u)| \geq Ce^{-\Delta c|u|^\alpha} \tag{3.3}$$

for some $0 \leq \alpha \leq 2$ and $C, c > 0$. This is equivalent to stating that X_Δ posesses at most a *supersmooth* density with parameters c and α (if a density exists at all).

Any infinitely divisible distribution having nonzero Gaussian part is supersmooth with $\alpha = 2$. Examples of distributions which are supersmooth with $\alpha < 2$ are tempered stable laws with index of stability α (e.g. [6], Chapter 4.5). Note that stable distributions do not fit in our setting, as they do not match the required moment condition. Normal inverse Gaussian processes which fulfill (3.3) with $\alpha = 1$ have been used for financial modelling, see e.g. [1]. Another example of processes in finance matching condition (3.3) with $\alpha = 1$ are Meixner-processes, see e.g. [16].

b) We have at most polynomial decay of the characteristic function:

$$|\varphi_\Delta(u)| \geq C\left(1 + |u|\right)^{-\Delta\beta} \tag{3.4}$$

for $C > 0$ and $\beta \geqslant 0$. This means that if X_Δ possesses a density at all, this can be no smoother than *ordinary smooth* with parameter β.

Typical examples of infinitely divisible random variables with ordinary smooth densities are Gamma distributions. Compound Poisson distributions, which do not posess a distributional density, fulfill (3.4) for $\beta = 0$. Another typical example of processes fulfilling (3.4) are variance gamma processes, which have been used to model the logarithm of stock prices, see, for example [3].

Inspired by the weak convergence in Proposition 2.1, the performance of the estimator of the finite measure ν_σ is measured by an integral criterion. For $s > 0$ define the space of test functions

$$F_s := \left\{ f \in L^1(\mathbb{R}) : \int |\mathcal{F}f(u)|\,(1+|u|)^s\,\mathrm{d}u < 1. \right\}.$$

The corresponding loss for an estimator $\hat{\nu}_\sigma$ of ν_σ is then defined to be

$$\ell_s\left(\hat{\nu}_\sigma, \nu_\sigma\right) := \sup_{f \in F_s} \left| \int f\,\mathrm{d}\nu_\sigma - \int f\,\mathrm{d}\hat{\nu}_\sigma \right|.$$

3.2 Theorem. *Assume* $\mathbb{E}\left[|X_\Delta|^{4+\gamma}\right] < \infty$ *for some* $\gamma > 0$. *Let* $\hat{\nu}_{\sigma_{\Delta,T}}$ *and* $\hat{b}_{\Delta,T}$ *be defined according to (2.6). Then*

$$\mathbb{E}\left[|\hat{b}_{\Delta,T} - b|\right] = O\left(T^{-\frac{1}{2}}\right).$$

For $\hat{\nu}_{\sigma_{\Delta,T}}$, *we obtain the following rates of convergence in probability:*

a) *For distributions with tail behaviour* $|\varphi_\Delta(u)| \geq Ce^{-\Delta c|u|^\alpha}$ *we have*

$$\ell_s\left(\hat{\nu}_{\sigma_{\Delta,T}}, \nu_\sigma\right) = O_\mathbb{P}\left(\left(\frac{\log T}{\Delta}\right)^{-\frac{s}{\alpha}} \vee T^{-\frac{1}{2}}\right).$$

Especially, the parametric rate $T^{-\frac{1}{2}}$, *is attained for* $T \to \infty$ *and simultaneously* $\Delta_T \to 0$ *provided*

$$\Delta_T = O\left(T^{-\frac{\alpha}{2s}}\log T\right).$$

b) *For distributions with tail behaviour* $|\varphi_\Delta(u)| \geq C(1+|u|)^{-\Delta\beta}$ *we have*

$$\ell_s\left(\hat{\nu}_{\sigma_{\Delta,T}}, \nu_\sigma\right) = O_\mathbb{P}\left(T^{-\frac{s}{2\Delta\beta}}\left(\log(e+T)\right)^{\frac{s(1/2+\delta)}{\Delta\beta}} \vee T^{-\frac{1}{2}}\right).$$

Especially, the parametric rate $T^{-\frac{1}{2}}$, *is attained for* $T \to \infty$ *under the non-asymptotic condition*

$$\Delta_T < \frac{s}{\beta}.$$

By standard parametric theory, all parameters cannot be estimated at a better rate than $T^{-1/2}$. Therefore the next result shows that our rates of convergence are minimax optimal (at least up to a logarithmic factor for (b)) within a nonparametric class.

6

3.3 Theorem (Minimax lower bounds)**.** *Let us introduce the following non-parametric classes for ν_σ:*

$$
\begin{aligned}
\mathcal{A}(C, c, \alpha) : \quad &= \quad \left\{ \nu_\sigma : \; |\varphi_\Delta(u)| \geq C e^{-\Delta c |u|^\alpha} \right\} \\
\mathcal{B}(C, \beta) : \quad &= \quad \left\{ \nu_\sigma : \; |\varphi_\Delta(u)| \geq C (1 + |u|)^{-\Delta \beta} \right\}.
\end{aligned}
$$

Then we obtain the following minimax lower bounds uniformly for $|b| \leq B$, where B is some positive constant:

$$
\exists \varepsilon > 0 : \liminf_{\substack{T \to \infty \\ \Delta_T \in (0,1]}} \inf_{\overline{\nu}_{\sigma_{\Delta_T}, T}} \sup_{\nu_\sigma \in \mathcal{A}(C, c, \alpha)} \mathbb{P}_{b, \nu_\sigma} \left(\left(\left(\frac{\log T}{\Delta_T} \right)^{\frac{s}{\alpha}} \wedge T^{\frac{1}{2}} \right) \ell_s \left(\overline{\nu}_{\sigma_{\Delta_T}, T}, \nu_\sigma \right) > \varepsilon \right) > 0,
$$

$$
\exists \varepsilon > 0 : \liminf_{\substack{T \to \infty \\ \Delta_T \in (0,1]}} \inf_{\overline{\nu}_{\sigma_{\Delta}, T}} \sup_{\nu_\sigma \in \mathcal{B}(C, \beta)} \mathbb{P}_{b, \nu_\sigma} \left(T^{\frac{s}{2\Delta_T \beta} \wedge \frac{1}{2}} \ell_s \left(\overline{\nu}_{\sigma_{\Delta_T}, T}, \nu_\sigma \right) > \varepsilon \right) > 0,
$$

where the infimum is taken over all estimators $\overline{\nu}_{\sigma_{\Delta_T}, T}$ of ν_σ based on observations of X with distance Δ_T up to time T.

The proof follows along the same lines as the proof in [14], but the control of the dependence on Δ requires additional and rather tedious calculations, whence it is omitted.

3.3 Discussion

The convergence rates for ν_σ can be understood in terms of a deconvolution or statistical inverse problem. The degree of ill-posedness, i.e. the amplification of the noise, is governed by the decay of the characteristic function φ_Δ. For fixed Δ and the exponential decay of φ_Δ in (a) we therefore face a severely ill-posed problem with logarithmic rates of convergence. On the other hand, the risk is smaller for smoother test functions. If we had looked also at analytic test functions, where the Fourier transform decays exponentially fast, then we would also in (a) obtain polynomial rates for fixed Δ. Observe that our estimator does neither rely on the knowledge of the decay behaviour of the unknown characteristic function nor on the test function class considered nor on the asymptotics of the observation distance.

The parametric rate is always attained when the smoothness of the test function sufficiently counterbalances the ill-posedness of the problem. It is remarkable that in all cases a condition on the observation distance of the type $\Delta = O(T^{-p})$ suffices. In the polynomial decay case (b) the ill-posedness is of degree $\Delta\beta$ which is smaller than the smoothness s exactly under the condition $\Delta < s/\beta$ and we need not assume high-frequency observations. Very roughly and intuitively, there is an analogy with estimating the derivative of order $\Delta\beta$ of a regression function and calculating the integral with an s-smooth test function of compact support, which by partial integration equals the integral of the regression function itself with an $(s - \Delta\beta)$-smooth

test function. This L^2-continuous linear functional can be estimated with a parametric rate, see e.g. [10].

Like in [9], we might consider the model that ν possesses a density $g \in C^r$ which we want to estimate. The kernel smoothing argument in [14] then yields in the polynomial decay case (b) a convergence rate for the point-wise risk of order $O(h^r + h^{-\Delta\beta - 1/2}T^{-1/2})$ (modulo a log factor, which is suppressed in the following), where h denotes the kernel bandwidth. An optimal bandwidth choice yields the rate $O(T^{-r/(2r+2\Delta\beta+1)})$. Under this loss we attain the high-frequency rate of convergence $O(T^{-r/(2r+1)})$ under the condition $\Delta \leqslant c(\log T)^{-1}$ with $c > 0$ sufficiently small. This logarithmic decay condition should be compared to [7] and [5] where in the compound Poisson case a polynomial condition is required for the critical observation distance Δ.

4   Proofs

4.1   Proof of Theorem 3.1

We start by recalling some definitions from empirical process theory. Let a probability space $(\mathcal{X}, \mathcal{A}, \mathbb{P})$ be given. For measurable functions $u, l : \mathcal{X} \to \mathbb{R}$, the set

$$[l, u] := \{h : \mathcal{X} \to \mathbb{R} \mid l \leq h \leq u\}$$

is called an ε-bracket, if

$$\int (u - l)^2 d\mathbb{P} < \varepsilon^2.$$

Given some class $\mathbb{F}$ of measurable, real-valued functions on $\mathcal{X}$, we denote by $N_{[]}\left(\varepsilon, \mathbb{F}, L^2(\mathbb{P})\right)$ the minimal number of ε-brackets which are needed to cover $\mathbb{F}$. The entropy integral is defined by

$$J_{[]}\left(\delta, \mathbb{F}, L^2(\mathbb{P})\right) := \int_0^\delta \left(\log N_{[]}\left(\varepsilon, \mathbb{F}, L^2(\mathbb{P})\right)\right)^{\frac{1}{2}} \, d\varepsilon.$$

Finally, a function $F \geq 0$ is called an envelope function for $\mathbb{F}$, if

$$\forall f \in \mathbb{F} : |f| \leq F.$$

Proof of Theorem 3.1 . We decompose $C_{\Delta,T}^{(k)}$ in its real and imaginary part and introduce the set of functions

$$\mathbb{F}_\Delta^k := \left\{ \Delta^{-\frac{1 \wedge k}{2}} \frac{d^k}{du^k} \cos ux : u \in \mathbb{R} \right\} \cup \left\{ \Delta^{-\frac{1 \wedge k}{2}} \frac{d^k}{du^k} \sin ux : u \in \mathbb{R} \right\}.$$

Denote by $\mathbb{P}_\Delta$ the distribution of X_Δ. An application of Corollary 19.35 in [17] gives for any $\Delta > 0$:

$$\sup_T \mathbb{E}\left[\|C_{\Delta,T}^{(k)}\|_{L^\infty(w)} \right] < C J_{[]}\left(\mathbb{E}\left[F^2(X_\Delta) \right], \mathbb{F}_\Delta^{(k)}, L^2(\mathbb{P}_\Delta) \right), \tag{4.1}$$

8

for any envelope function $F = F_\Delta^k$ of $\mathbb{F}_\Delta^k$ and a universal constant C which does not depend on Δ. It is shown in [14] that the right hand side of 4.1 is finite. To make the result uniform in Δ, it remains to consider the behaviour of the entropy integral for $\Delta \in (0,1]$ varying.

To cover $\mathbb{F}_\Delta^k$ with brackets of size ε, we define for grid points $u_{\Delta,j}$ specified later the bracket functions

$$g_{\Delta,j}^\pm(z) = \Delta^{-\frac{1\wedge k}{2}}\left(w(u_{\Delta,j})\frac{\mathrm{d}^k}{\mathrm{d}u^k}\cos(u_{\Delta,j}z) \pm \varepsilon|z|^k\right)\mathbb{I}_{[-M,M]}(z)\pm\Delta^{-\frac{1\wedge k}{2}}|z|^k\mathbb{I}_{[-M,M]^c}(z)$$

and

$$h_{\Delta,j}^\pm(z) = \Delta^{-\frac{1\wedge k}{2}}\left(w(u_{\Delta,j})\frac{\mathrm{d}^k}{\mathrm{d}u^k}\sin(u_{\Delta,j}z) \pm \varepsilon|z|^k\right)\mathbb{I}_{[-M,M]}(z)\pm\Delta^{-\frac{1\wedge k}{2}}|z|^k\mathbb{I}_{[-M,M]^c}(z),$$

with

$$M := M(\varepsilon,\Delta,k) := \inf\left\{m : \Delta^{-(1\wedge k)}\mathbb{E}|X_\Delta|^{2k}\mathbb{I}_{\{|X_\Delta|>m\}} \le \varepsilon^2\right\}.$$

By definition of M, the size of the brackets is

$$\mathbb{E}\left[\left(g_{\Delta,j}^+(X_\Delta) - g_{\Delta,j}^-(X_\Delta)\right)^2\right] \le 4\varepsilon^2\left(\Delta^{-(1\wedge k)}\mathbb{E}X_\Delta^{2k} + 1\right).$$

For $\Delta \le 1$, the expression on the right is uniformly bounded above by $c\varepsilon^2$ for some $c > 0$. This is obvious for $k = 0$. For $k \ge 1$, this is a consequence of the well known fact that $\mathbb{E}\left[X_\Delta^{2k}\right] \le \bar{c}\Delta$ for some $\bar{c} > 0$, which is seen by using the formula

$$\mathbb{E}\left[X_\Delta^{2k}\right] = i^{-2k}\varphi_\Delta^{(2k)}(0) = i^{-2k}\frac{d^{2k}}{du^{2k}}e^{\Delta\Psi(u)}\Big|_{u=0}.$$

An analogous argument gives:

$$\mathbb{E}\left[\left(h_{\Delta,j}^+(X_\Delta) - h_{\Delta,j}^-(X_\Delta)\right)^2\right] \le c\varepsilon^2.$$

For a function $g_u(\bullet) := \Delta^{-\frac{1\wedge k}{2}}w(u)\frac{\partial^k}{\partial u^k}\cos(u\bullet) \in \mathbb{F}_\Delta^k$ to be contained in $[g_{\Delta,j}^-, g_{\Delta,j}^+]$, we have to ensure

$$|w(u)\frac{d^k}{du^k}\cos(uz) - w(u_{\Delta,j})\frac{d^k}{du^k}\cos(u_{\Delta,j}z)| \le \varepsilon|z|^k \quad \forall z \in [-M,M]. \quad (4.2)$$

With the estimate

$$
\begin{aligned}
&|w(u)\cos(uz) - w(u_j)\cos(u_jz)|\mathbb{I}_{[-M,M]}(z)\\
\le\ & (w(u) + w(u_j)) \wedge\\
&\big(|w(u)\cos(uz) - w(u)\cos(u_jz)|\,\mathbb{I}_{[-M,M]}(z) + |w(u)\cos(u_jz) - w(u_j)\cos(u_jz)|\,\mathbb{I}_{[-M,M]}(z)\big)\\
\le\ & (w(u) + w(u_j)) \wedge (M|u - u_j| + Lip(w)|u - u_j|),
\end{aligned}
$$

9

where $Lip(w)$ is the Lipschitz-constant of w, and with the analogous inequality for the sine-function, (4.2) is seen to hold for any $u \in \mathbb{R}$ such that

$$\min\left\{|u - u_{\Delta,j}|(Lip(w) + M), w(u) + w(u_{\Delta,j})\right\} \leq \varepsilon.$$

Hence to cover $\mathbb{F}_\Delta^k$ with brackets of $\mathbb{P}_\Delta$-size $c\varepsilon^2$, we need grid points $u_1, \cdots, u_{J(\varepsilon)}$ such that $w(u_1) \leq \frac{\varepsilon}{2}$, $w(u_{J(\varepsilon)}) \leq \frac{\varepsilon}{2}$ and $|u_j - u_{j+1}| \leq \frac{\varepsilon}{Lip(w) + M(\varepsilon, \Delta, k)}$. For the minimal number $J(\varepsilon)$ of $c\varepsilon$-brackets needed to cover $\mathbb{F}_{\Delta,k}$, this yields the estimate

$$J(\varepsilon) \leq 2U(\varepsilon)(Lip(w) + M(\varepsilon, \Delta, k))/\varepsilon,$$

with

$$U(\varepsilon) := \inf\left\{u \in \mathbb{R} : w(u) \leq \frac{\varepsilon}{2}\right\} \asymp \exp\left(\varepsilon^{-\kappa}\right) \text{ for some } \kappa < 2.$$

The generalized Markov inequality yields for some $c' > 0$:

$$M(\varepsilon, \Delta, k) \leq \left(\mathbb{E}\left[|X_\Delta|^{2k+\gamma}\right]/\Delta^{1 \wedge k}\varepsilon^2\right)^{1/\gamma} < c'\varepsilon^{-2/\gamma}.$$

The second inequality applies the fact that we have the moment bound $\mathbb{E}\left[|X_\Delta|^{2k+\gamma}\right] = O(\Delta)$, which is a consequence of Theorem 1.1 in [8].

The entropy with bracketing is

$$\log N_{[]}(\varepsilon, \mathbb{F}_{\Delta,k}, L^2(\mathbb{P}_\Delta)) \leq \log U(\frac{\varepsilon}{c}) + \log\left(\frac{c(Lip(w) + M(\varepsilon/c, \Delta, k))}{\varepsilon}\right).$$

The upper limit in the entropy integral appearing in (4.1), $\sqrt{\mathbb{E}[F_{\Delta,k}^2(X_\Delta)]}$, is again bounded above uniformly in $\Delta < 1$. We have thus shown that up to some universal constant

$$\sup_{\Delta \leq 1} \sup_T \mathbb{E}\left[\|C_{\Delta,T}^{(k)}\|_{L^\infty(w)}\right]$$
$$\leq \int_0^1 \sqrt{\log(U(\varepsilon))}\, d\varepsilon + \int_0^1 \sqrt{\log\left(Lip(w)/\varepsilon + \varepsilon^{-(2/\gamma+1)}\right)}\, d\varepsilon. \quad (4.3)$$

Now (4.3) is finite since $\log U(\varepsilon) \asymp \varepsilon^{-\kappa}$ for some $\kappa < 2$. This completes the proof. □

4.2 Proof of Theorem 3.2

To prove the upper bounds, we establish a number of technical lemmas giving control on the characteristic exponent and its derivatives. First, we formulate a result which connects the tail behaviour of the characteristic function (which corresponds to the smoothnes of the density) to the jump activity round the origin, extending a result from [14]:

4.1 Lemma. *Let an infinitely divisible law with characteristics $(b, 0, \nu)$ be given such that its characteristic function satisfies*

$$|\varphi(u)| \geq C e^{-c|u|^{\alpha}}$$

for some $0 < \alpha < 2$ and $C, c > 0$. Then for any $\alpha' > \alpha$ the integral

$$\int_{-1}^{1} |x|^{\alpha'} \nu(\,dx)$$

is finite.

Proof. Setting $\kappa := \inf_{1 < x \leq 2} (1 - \cos x) > 0$, we have the series of inequalities

$$
\begin{aligned}
\int_{-1}^{1} |x|^{\alpha'} \nu(\,dx) &= \sum_{m=0}^{\infty} \int_{\{2^{-(m+1)} < |x| \leq 2^{-m}\}} |x|^{\alpha'} \nu(\,dx) \\
&\leq \kappa^{-1} \sum_{m=0}^{\infty} 2^{-\alpha' m} \int \left(1 - \cos(2^{m+1} x)\right) \nu(\,dx) \\
&= \kappa^{-1} \sum_{m=0}^{\infty} 2^{-\alpha' m} \left(-\operatorname{Re}\Psi(2^{m+1})\right) \\
&\leq \kappa^{-1} \left(2c \sum_{m=0}^{\infty} 2^{-(\alpha'-\alpha)m} - \log C \sum_{m=0}^{\infty} 2^{-\alpha' m}\right) < \infty.
\end{aligned}
$$

$\square$

4.2 Lemma. *In the situation of the preceding lemma, let $\alpha \in [1, 2)$ and assume finite moments for the law of order $\alpha' > \alpha$. Then the following bound on the derivative of the characteristic exponent holds for $\alpha' \in (\alpha, 2)$:*

$$\forall u \in \mathbb{R} : \quad |\Psi'(u)| \leq K(1 + |u|^{\alpha'-1}) \tag{4.4}$$

for some $K > 0$.

    For $\alpha < 1$ the derivative of the characteristic exponent is always uniformly bounded:

$$\sup_{u \in \mathbb{R}} |\Psi'(u)| < \infty. \tag{4.5}$$

Proof. Since the diffusion part is zero by assumption, we obtain

$$
\begin{aligned}
|\Psi'(u)| &= \left| ib + i \int \left(e^{iux} - 1\right) x\nu(\,dx) \right| \tag{4.6} \\
&\leq |b| + \int (2 \wedge |ux|) |x| \nu(\,dx) \tag{4.7} \\
&\leq |b| + 2^{2-\alpha'} |u|^{\alpha'-1} \int |x|^{\alpha'} \nu(\,dx). \tag{4.8}
\end{aligned}
$$

and the integral appearing in (4.8) is finite by Lemma 4.1 together with the moment assumption. We have thus shown (4.4). To see (4.5) , we can estimate

$$\left| \Psi'(u) \right| \quad \leq \quad |b| + 2 \int |x| \nu(\,\mathrm{d}x) \tag{4.9}$$

and this expression is finite for $\alpha < 1$ by Lemma 4.1. $\qquad\square$

Next, we focus on the exponential decay behaviour. We first need a result concerning the minimum distance fit of the characteristic function.

4.3 Lemma. *Let $|\varphi_\Delta(u)| \geq Ce^{-\Delta c|u|^\alpha}$. With*

$$I_{\Delta,T} := [-U_{\Delta,T}, U_{\Delta,T}] := \left[-\left(\frac{\log T}{3\Delta} \right)^{\frac{1}{\alpha}}, \left(\frac{\log T}{3\Delta} \right)^{\frac{1}{\alpha}} \right].$$

we find for any observation distance $\Delta = \Delta_T \in (0,1]$

$$\lim_{T \to \infty} \mathbb{P} \left(\forall u \in I_{\Delta,T} : \ |\varphi_{\Delta,T}(u)| \geq \frac{C}{2} e^{-\Delta c u^\alpha} \right) = 1.$$

Proof. ¿From Theorem 3.1 we infer by Markov's inequality

$$\mathbb{P} \left(\exists u \in I_{\Delta,T} : |\varphi_{\Delta,T}(u)| < \frac{C}{2} e^{-\Delta c|u|^\alpha} \right)$$

$$\leq \quad \mathbb{P} \left(\exists u \in I_{\Delta,T} : \ |\varphi_{\Delta,T}(u) - \varphi_\Delta(u)| > \frac{C}{2} e^{-\Delta c|u|^\alpha} \right)$$

$$= \quad \mathbb{P} \left(\sup_{u \in I_{\Delta,T}} |\varphi_{\Delta,T}(u) - \varphi_\Delta(u)| \frac{2}{C} e^{\Delta c|u|^\alpha} > 1 \right)$$

$$\leq \quad w(U_{\Delta,T})^{-1} \frac{2}{C} e^{\Delta |U_{\Delta,T}|^\alpha} \Delta^{\frac{1}{2}} O\left(T^{-\frac{1}{2}} \right).$$

The choice of $U_{\Delta,T}$ ensures that this expression tends to zero for $T \to \infty$, whatever Δ is. $\qquad\square$

Let $\Psi_\Delta := \Delta \Psi(u)$ denote the characteristic exponent of the true characteristic function φ_Δ and $\Psi_{\Delta,T}$ the characteristic exponent of the minimum distance fit $\varphi_{\Delta,T}$. The next two results give control on the deviation of $\Psi_{\Delta,T}$ from Ψ_Δ and of its second derivatives.

4.4 Lemma. *Let $|\varphi_\Delta(u)| \geq Ce^{-\Delta c|u|^\alpha}$. With $K > 0$ from (4.4) the following bound in probability is valid:*

$$\sup_{u \in I_{\Delta,T}} \frac{\left| \Psi'_{\Delta,T}(u) - \Psi'_\Delta(u) \right|}{\Delta w(u)^{-1} e^{\Delta c|u|^\alpha} \left(1 + \Delta^{\frac{1}{2}} K(1 + |u|^{\frac{\alpha}{2}}) \right)} = O_\mathbb{P}(T^{-\frac{1}{2}}). \tag{4.10}$$

Moreover,

$$\sup_{u \in I_{\Delta,T}} \frac{\left| \Psi'_{\Delta,T}(u) \right|}{\Delta w(u)^{-1} e^{\Delta c |u|^\alpha} K(1 + |u|^{\frac{\alpha}{2}})} = O_{\mathbb{P}}(1). \qquad (4.11)$$

Proof. We have, with probability tending to one, for all $u \in U_{\Delta;T}$:

$$\left| \Psi'_{\Delta,T}(u) - \Psi'_\Delta(u) \right| = \left| \frac{\varphi'_{\Delta,T}(u)}{\varphi_{\Delta,T}(u)} - \frac{\varphi'_\Delta(u)}{\varphi_\Delta(u)} \right|$$

$$\leq \frac{|\varphi'_{\Delta,T}(u) - \varphi'_\Delta(u)|}{|\varphi_{\Delta,T}(u)|} + \left| \Psi'_{\Delta,T}(u) \right| \frac{|\varphi_{\Delta,T}(u) - \varphi_\Delta(u)|}{|\varphi_{\Delta,T}(u)|}$$

$$\leq \left(e^{\Delta c |u|^\alpha} w(u)^{-1} \Delta + \Delta K (1 + |u|^{\frac{\alpha}{2}}) e^{\Delta c |u|^\alpha} w(u)^{-1} \Delta^{\frac{1}{2}} \right) \Delta^{-\frac{1}{2}} d\left(\varphi_{\Delta,T}, \varphi_\Delta \right),$$

where the last inequality is a consequence of Lemma 4.2 and Lemma 4.3.
Another application of Theorem 3.1 gives (4.10).
Now (4.11) follows from (4.10), using Lemma 4.4 and the estimate

$$\left| \Psi'_{\Delta,T}(u) \right| \leq \left| \Psi'_\Delta(u) \right| + \left| \Psi'_{\Delta,T}(u) - \Psi'_\Delta(u) \right|.$$

$\square$

4.5 Lemma. *Let $|\varphi_\Delta(u)| \geq C e^{-\Delta c |u|^\alpha}$. For the second derivative of the characteristic exponent we have*

$$\sup_{u \in \mathbb{R}} \left| \Psi''_{\Delta,T}(u) - \Psi''_\Delta(u) \right| = \Delta O_{\mathbb{P}}(1).$$

Moreover, we can give the following bound in probability uniformly on $I_{\Delta,T}$:

$$\sup_{u \in I_{\Delta,T}} \frac{\left| \Psi''_{\Delta,T}(u) - \Psi''_\Delta(u) \right|}{\Delta C e^{\Delta c |u|^\alpha} w(u)^{-1} \left(1 + \Delta^{\frac{1}{2}}(1 + |u|^{\frac{\alpha}{2}}) + \Delta^{\frac{3}{2}}(1 + |u|^\alpha) \right)} = O_{\mathbb{P}} \left(T^{-\frac{1}{2}} \right) (4.12)$$

Proof. To see the first statement of the lemma, recall that the second derivative of the characteristic exponent is always bounded above:

$$\forall u \in \mathbb{R} : \left| \Psi''_\Delta(u) \right| = \Delta \left| -\sigma^2 + \int e^{iux} x^2 \nu(\mathrm{d}x) \right| \leq \Delta \left(\sigma^2 + \int |x|^2 \nu(\mathrm{d}x) \right) < \infty.$$

Then apply the series of inequalities

$$\left| \Psi''_{\Delta,T}(u) - \Psi''_\Delta(u) \right|$$

$$\leq 4 \left(|\Psi''_\Delta(0)| + \left| \Psi''_{\Delta,T}(0) - \Psi''_\Delta(0) \right| \right)$$

$$\leq 4 \left(|\Psi''_\Delta(0)| + \left| \varphi''_{\Delta,T}(0) - \varphi''_\Delta(0) \right| + \left| (\Psi'_{\Delta,T}(0))^2 - (\Psi'_\Delta(0))^2 \right| \right)$$

$$= 4 \left(|\Psi''_\Delta(0)| + \left| \varphi''_{\Delta,T}(0) - \varphi''_\Delta(0) \right| + 2 |\Psi''_\Delta(0)| \left| \varphi'_{\Delta,T}(0) - \varphi'_\Delta(0) \right| + \left| \varphi'_{\Delta,T}(0) - \varphi'_\Delta(0) \right|^2 \right)$$

$$= \Delta O_{\mathbb{P}} \left(1 + T^{-\frac{1}{2}} + \Delta T^{-\frac{1}{2}} + \Delta T^{-1} \right)$$

$$= \Delta O_{\mathbb{P}} (1).$$

13

Next, (4.12) can be seen by estimating

$$
\left| \Psi''_{\Delta,T}(u) - \Psi''_\Delta(u) \right|
$$

$$
= \left| \frac{\varphi''_{\Delta,T}(u)}{\varphi_{\Delta,T}(u)} - \left(\Psi'_{\Delta,T}(u) \right)^2 - \frac{\varphi''_\Delta(u)}{\varphi_\Delta(u)} + \left(\Psi'_\Delta(u) \right)^2 \right|
$$

$$
\leq \frac{\left| \varphi''_{\Delta,T}(u) - \varphi''_\Delta(u) \right|}{|\varphi_\Delta(u)|} + \left| \Psi'_{\Delta,T}(u) \right| \frac{|\varphi_{\Delta,T}(u) - \varphi_\Delta(u)|}{|\varphi_\Delta(u)|}
$$

$$
+ \left| \Psi'_{\Delta,T}(u) + \Psi'_\Delta(u) \right| \left| \Psi'_{\Delta,T}(u) - \Psi'_\Delta(u) \right|.
$$

The desired bound is an immediate consequence of Lemma 4.4. □

For distributions with characteristic functions decaying at most polynomially, we can prove auxiliary results analogous to Lemmas 4.1-4.5. As the proofs run in a completely analogous way, we omit the details and only state the main result:

4.6 Lemma. *Let $|\varphi_\Delta(u)| \geq C(1 + |u|)^{-\Delta\beta}$. Define*

$$
I_{\Delta,T} := \left[-T^{\frac{1}{2\Delta\beta}} \left(\log(e + T) \right)^{-\frac{1/2+2\delta}{\Delta\beta}}, +T^{\frac{1}{2\Delta\beta}} \left(\log T \right)^{-\frac{1/2+2\delta}{\Delta\beta}} \right].
$$

Then we have

$$
\sup_{u \in I_{\Delta,T}} \frac{\left| \Psi''_{\Delta,T}(u) - \Psi''_\Delta(u) \right|}{\Delta C(1 + |u|)^{\Delta\beta} w(u)^{-1}} = O_\mathbb{P} \left(T^{-\frac{1}{2}} \right).
$$

The proof of the upper bound result can now easily be obtained as a consequence of the preceding lemmas.

Proof of Theorem 3.2: The result for $\hat{b}_{\Delta,T}$ is an immediate consequence of Theorem 3.1, using $|\hat{b}_{\Delta,T} - b| = \Delta^{-1}|\varphi'_{\Delta,T}(0) - \varphi'_\Delta(0)|$. For the estimator of ν_σ, applying Parseval's identity, the loss satisfies

$$
\ell_s \left(\hat{\nu}_{\sigma_{\Delta,T}}, \nu_\sigma \right) = \sup_{f \in F_s} \left| \int f(x) \hat{\nu}_{\sigma_{\Delta,T}}(\mathrm{d}x) - \int f(x) \nu_\sigma(\mathrm{d}x) \right|
$$

$$
= \frac{1}{2\pi} \sup_{f \in F_s} \left| \int \mathcal{F}f(u) \overline{\left(\mathcal{F}\hat{\nu}_{\sigma_{\Delta,T}}(u) - \mathcal{F}\nu_\sigma(u) \right)} \, \mathrm{d}u \right|
$$

$$
\leq \frac{1}{2\pi} \sup_{f \in F_s} \int |\mathcal{F}f(u)| \frac{1}{\Delta} \left| \Psi''_\Delta(u) - \Psi''_{\Delta,T}(u) \right| \, \mathrm{d}u
$$

$$
\leq \frac{1}{2\pi} \sup_{u \in \mathbb{R}} (1 + |u|)^{-s} \frac{1}{\Delta} \left| \Psi''_{\Delta,T}(u) - \Psi''_\Delta(u) \right|.
$$

By an application of Lemma 4.5 and Lemma 4.6, we can estimate

a) for $|\varphi_\Delta(u)| \geq Ce^{-\Delta c|u|^\alpha}$:

$$\sup_{u \in \mathbb{R}} (1+|u|)^{-s} \frac{1}{\Delta} \left| \Psi''_{\Delta,T}(u) - \Psi''_\Delta(u) \right|$$

$$\leq \sup_{u \in I_{\Delta,T}} (1+|u|)^{-s} \left(\frac{1 + \Delta^{\frac{1}{2}}(1+|u|^{\frac{\alpha}{2}}) + \Delta^{\frac{3}{2}}(1+|u|^\alpha)}{e^{-\Delta c|u|^\alpha} w(u)} \right) O_{\mathbb{P}}\left(T^{-\frac{1}{2}}\right) \wedge (1+U_{\Delta,T})^{-s}$$

$$= O_{\mathbb{P}}\left(T^{-\frac{1}{2}} \vee \left(\frac{\log T}{\Delta}\right)^{-\frac{s}{\alpha}}\right).$$

b) for $|\varphi_\Delta(u)| \geq C(1+|u|)^{-\Delta\beta}$:

$$\sup_{u \in \mathbb{R}} (1+|u|)^{-s} \frac{1}{\Delta} \left| \Psi''_{\Delta,T}(u) - \Psi''_\Delta(u) \right|$$

$$= O_{\mathbb{P}}\left(T^{-\frac{s}{2\Delta\beta}} (\log(e+T))^{\frac{s(1/2+\delta)}{\Delta\beta}} \vee T^{-\frac{1}{2}}\right).$$

$\square$

4.3 Acknowledgement

We want to thank two anonymous referees and one editor for careful reading and helpful comments.

References

[1] Ole E. Barndorff-Nielsen. Processes of normal inverse Gaussian type. *Finance Stoch.*, 2(1):41–68, 1998.

[2] Denis Belomestny and Markus Reiß. Spectral calibration of exponential Lévy models. *Finance Stoch.*, 10(4):449–474, 2006.

[3] Peter Carr and Dillip Madan. Option valuation using the fast Fourier transform. *Journal of Computational Finance*, 2:61–73, 1998.

[4] Fabienne Comte and Valentine Genon-Catalot. Nonparametric adaptive estimation for pure jump Lévy processes. *Annales de l'I. H. P., Probability and Statistics (to appear)*.

[5] Fabienne Comte and Valentine Genon-Catalot. Nonparametric estimation for pure jump Lévy processes based on high frequency data. *Stochastic processes and their applications (to appear)*.

[6] Rama Cont and Peter Tankov. *Financial modelling with jump processes.* Chapman & Hall/CRC Financial Mathematics Series. Boca Raton, 2004.

[7] José Figueroa-Lopez. Nonparametric estimation for Lévy models based on discrete sampling. *IMS Lecture Notes of the 3rd E.L. Lehmann Symposium*, 57:117–146, 2009.

[8] José E. Figueroa-López. Small-time moment asymptotics for Lévy processes. *Stat. Probab. Lett.*, 78(18):3355–3365, 2008.

[9] José E. Figueroa-López and Christian Houdré. Risk bounds for the non-parametric estimation of Lévy processes. Giné, Evarist (ed.) et al., High dimensional probability. Institute of Mathematical Statistics Lecture Notes - Monograph Series 51, 96-116, 2006.

[10] L. Goldstein and K. Messer. Optimal plug-in estimators for nonparametric functional estimation. *Ann. Statist.*, 20(3):1306–1328, 1992.

[11] Shota Gugushvili. Nonparametric estimation for discretely sampled Lévy processes. *arXiv:0908.3121v2*, 2009.

[12] Shota Gugushvili. Nonparametric estimation of the characteristic triplet of a discretely observed Lévy process. *J. Nonparametric Stat.*, 21(3):321–343, 2009.

[13] Geurt Jongbloed, Frank H. van der Meulen, and Aad.W. van der Vaart. Nonparametric inference for Lévy-driven Ornstein-Uhlenbeck processes. *Bernoulli*, 11(5):759–791, 2005.

[14] Michael Neumann and Markus Reiß. Nonparametric estimation for Lévy processes from low frequency observations. *Bernoulli*, 15(1):223–248, 2009.

[15] Ken-Iti Sato. *Lévy processes and infinisible divisible distributions.* Cambridge University Press, 1999.

[16] Wim Schoutens and Jozef L. Teugels. Lévy processes, polynomials and martingales. *Commun. Stat., Stochastic Models*, 14(1-2):335–349, 1998.

[17] Aad W. Van der Vaart. *Asymptotic statistics.* Cambridge University Press, 1998.