Bernoulli **15**(1), 2009, 223–248
DOI: 10.3150/08-BEJ148

Nonparametric estimation for Lévy processes from low-frequency observations

MICHAEL H. NEUMANN[1] and MARKUS REIß[2]

[1]*Friedrich-Schiller-Universität Jena, Institut für Stochastik, Ernst-Abbe-Platz 2, D–07743 Jena, Germany. E-mail: mneumann@mathematik.uni-jena.de*
[2]*Ruprecht-Karls-Universität Heidelberg, Institut für Angewandte Mathematik, Im Neuenheimer Feld 294, D–69120 Heidelberg, Germany. E-mail: reiss@statlab.uni-heidelberg.de*

We suppose that a Lévy process is observed at discrete time points. A rather general construction of minimum-distance estimators is shown to give consistent estimators of the Lévy–Khinchine characteristics as the number of observations tends to infinity, keeping the observation distance fixed. For a specific C^2-criterion this estimator is rate-optimal. The connection with deconvolution and inverse problems is explained. A key step in the proof is a uniform control on the deviations of the empirical characteristic function on the whole real line.

Keywords: deconvolution; density estimation; Lévy–Khinchine characteristics; minimum distance estimator

1. Introduction

Lévy processes form the fundamental building block for stochastic continuous-time models with jumps. There is an important trend using Lévy models in finance, see Cont and Tankov (2004), but also many recent models in physics or biology rely on Lévy processes. We consider here the problem of estimating the Lévy–Khinchine characteristics from time-discrete observations of a Lévy process. Since these characteristics involve the Lévy measure (or jump measure) and we do not want to impose a parametric model, we face a nonparametric estimation problem.

When the Lévy process $(X_t)_{t\ge 0}$ is observed at high frequency, at times $(t_i)_{i=0,\dots,n}$ with $\max_i(t_i - t_{i-1})$ small, then a large increment $X_{t_i} - X_{t_{i-1}}$ indicates that a jump occurred between time t_{i-1} and t_i. Based on this insight and the continuous-time observation analogue, nonparametric inference for Lévy processes from high-frequency data has been considered by Basawa and Brockwell (1982), Figueroa-López and Houdré (2006) and Nishiyama (2008). For low-frequency observations, however, we cannot be sure to what extent the increment $X_{t_i} - X_{t_{i-1}}$ is due to one or several jumps or just to the Brownian motion part of the Lévy process. The only way to draw inference is to use that the increments form independent realizations of infinitely divisible probability distributions. We shall assume that we dispose of equidistant observations at $t_i = i\Delta$, $i = 0, \dots, n$, and consider the asymptotic behaviour of estimators for $n \to \infty$ and $\Delta > 0$ fixed. This can be cast into the classical framework of i.i.d. observations $(X_{i\Delta} - X_{(i-1)\Delta})_{i=1,\dots,n}$ from an infinitely divisible distribution. A natural question in this framework is to estimate the underlying Lévy–Khinchine characteristics. In this general setting we are only aware of the work by Watteel and Kulperger (2003) who propose and implement an approach for estimating the jump distribution by a fixed spectral cut-off procedure, which is related to the pilot estimator in Section 5 below. In the special case of compound Poisson processes the problem of estimating the jump density is known as decompounding, see van Es, Gugushvili and Spreij (2007) and Gugushvili (2007) and the references therein. For parametric inference under the assumption of a stable law see, for example, Feuerverger and McDunnough (1981b). A related low-frequency problem for the canonical function in Lévy–Ornstein–Uhlenbeck processes has been considered by Jongbloed, van der Meulen and van der Vaart (2005), where a consistent estimator has been constructed.

In Section 2 we recall basic facts about Lévy processes and prepare the idea of minimum-distance estimators based on the empirical characteristic function. Under very general conditions we then show in Section 3 consistency of these estimators for the Lévy–Khinchine characteristics. The only way to achieve this is to merge the diffusion coefficient σ^2 and the Lévy measure ν to a single quantity ν_σ, which is a finite Borel measure, and to consider weak convergence of estimators of ν_σ. In Section 4 we construct a rate-optimal estimator using a minimum-distance fit, based on a C^2-criterion for the empirical characteristic function. A fundamental tool is Theorem 4.1, which gives a uniform control on the deviations of the empirical characteristic function on the whole real line and may be of independent interest. The optimal rates of convergence depend on the decay of the characteristic function as in deconvolution problems. Interestingly, our estimator attains the optimal rates without knowing this decay behaviour and without any further regularization parameter. In Section 5 we briefly discuss the implementation of the estimator, using a two-step procedure, and show a typical numerical example. Most proofs are postponed to Section 6.

2. Basic notions, assumptions and a few simple facts

We assume that we observe a one-dimensional Lévy process $(X_t)_{t\geq 0}$ at equidistant time points $0 = t_0 < t_1 < \cdots < t_n$. Such a process is characterized by its characteristic function

$$\varphi(u, t; \bar{b}, \sigma, \nu) := \mathbb{E}[\exp(\mathrm{i}uX_t)] = \exp(t\Psi(u; \bar{b}, \sigma, \nu)), \qquad u \in \mathbb{R},$$

where

$$\Psi(u) = \Psi(u; \bar{b}, \sigma, \nu) = \mathrm{i}u\bar{b} - \frac{\sigma^2}{2}u^2 + \int_{\mathbb{R}} \left(\mathrm{e}^{\mathrm{i}ux} - 1 - \frac{\mathrm{i}ux}{1+x^2} \right) \nu(\mathrm{d}x).$$

The triplet $(\bar{b}, \sigma, \nu)$ is called Lévy–Khinchine characteristic or characteristic triplet with drift-like part $\bar{b} \in \mathbb{R}$, volatility $\sigma \geq 0$ and jump measure ν, which is a non-negative σ-finite measure on $(\mathbb{R}, \mathcal{B})$ with $\int \frac{x^2}{1+x^2}\nu(\mathrm{d}x) < \infty$. The function Ψ is called characteristic exponent or cumulant function.

For reasons explained below, we introduce a measure $\bar{\nu}_\sigma$ by

$$\bar{\nu}_\sigma(\mathrm{d}x) = \sigma^2\delta_0(\mathrm{d}x) + \frac{x^2}{1+x^2}\nu(\mathrm{d}x),$$

where δ_0 denotes the point measure in zero. This gives another representation of Ψ in terms of $\bar{b} \in \mathbb{R}$ and the finite Borel measure $\bar{\nu}_\sigma$ as

$$\Psi(u) = \Psi(u; \bar{b}, \bar{\nu}_\sigma) = \mathrm{i}u\bar{b} + \int_{\mathbb{R}} \frac{(\mathrm{e}^{\mathrm{i}ux} - 1)(1 + x^2) - \mathrm{i}ux}{x^2} \bar{\nu}_\sigma(\mathrm{d}x).$$

Here we have used the continuous extension of the integrand at $x = 0$, which evaluates to $-u^2/2$. Let $P_{\bar{b},\bar{\nu}_\sigma}$ denote the probability distribution with characteristic function $\varphi(\bullet, t; \bar{b}, \bar{\nu}_\sigma) = \exp(t\Psi(\bullet; \bar{b}, \bar{\nu}_\sigma))$ for some fixed $t > 0$. Writing $\mu_n \Longrightarrow \mu$ for weak convergence of the finite Borel measures μ_n to the finite Borel measure μ on $(\mathbb{R}, \mathcal{B})$, the following well-known result will be essential in the sequel (Theorem VII.2.9 and Remark VII.2.10 in Jacod and Shiryaev (2002) or Theorem 19.1 in Gnedenko and Kolmogorov (1968)).

Proposition 2.1. *The convergence $P_{\bar{b}_n,\bar{\nu}_{\sigma,n}} \Longrightarrow P_{\bar{b},\bar{\nu}_\sigma}$ takes place if and only if $\bar{b}_n \to \bar{b}$ and $\bar{\nu}_{\sigma,n} \Longrightarrow \bar{\nu}_\sigma$.*

By the scaling properties of Lévy processes there is no loss in generality when we suppose $t_k = k$, $k = 0, \dots, n$. We write $\varphi(u; \bar{b}, \bar{\nu}_\sigma)$ short for $\varphi(u, 1; \bar{b}, \bar{\nu}_\sigma)$. Let us introduce the empirical characteristic function of the increments

$$\widehat{\varphi}_n(u) := \frac{1}{n} \sum_{t=1}^{n} \mathrm{e}^{\mathrm{i}u(X_t - X_{t-1})}, \qquad u \in \mathbb{R}.$$

Since these increments are independent and identically distributed it follows from the Glivenko–Cantelli theorem that

$$P_{\bar{b},\bar{\nu}_\sigma}\Big(\widehat{\varphi}_n(u) \underset{n\to\infty}{\longrightarrow} \varphi(u; \bar{b}, \bar{\nu}_\sigma)\ \forall u \in \mathbb{R}\Big) = 1. \tag{2.1}$$

We will consider minimum distance fits, that is, we intend to choose $\widehat{\bar{b}}_n$ and $\widehat{\bar{\nu}}_{\sigma,n}$ such that, for an appropriate metric d,

$$d(\widehat{\varphi}_n, \varphi(\bullet; \widehat{\bar{b}}_n, \widehat{\bar{\nu}}_{\sigma,n})) = \inf_{\tilde{b}\in\mathbb{R}, \tilde{\nu}_\sigma \in \mathcal{M}(\mathbb{R})} d(\widehat{\varphi}_n, \varphi(\bullet; \tilde{b}, \tilde{\nu}_\sigma)). \tag{2.2}$$

Here $\mathcal{M}(\mathbb{R})$ denotes the space of all finite Borel measures on $(\mathbb{R}, \mathcal{B})$. Our basic motivation for this estimation procedure arises from the fact that an exact maximum likelihood estimator is not feasible since there is in general no closed form expression for the probability density of the observations available. Moreover, it is well-known that methods based on the empirical characteristic function can be asymptotically efficient; see Feuerverger and McDunnough (1981a, 1981b). Since we are not sure that the infimum in (2.2) is always obtained, we take a sequence of positive reals $(\delta_n)_{n\in\mathbb{N}}$ with $\delta_n \to 0$ as $n \to \infty$ and choose $\widehat{\bar{b}}_n$ and $\widehat{\bar{\nu}}_{\sigma,n}$ such that

$$d(\widehat{\varphi}_n, \varphi(\bullet; \widehat{\bar{b}}_n, \widehat{\bar{\nu}}_{\sigma,n})) \le \inf_{\tilde{b}\in\mathbb{R}, \tilde{\nu}_\sigma \in \mathcal{M}(\mathbb{R})} d(\widehat{\varphi}_n, \varphi(\bullet; \tilde{b}, \tilde{\nu}_\sigma)) + \delta_n. \tag{2.3}$$

For the metric d, we assume that

$$\lim_{n\to\infty} d(\widehat{\varphi}_n, \varphi(\bullet; \bar{b}, \bar{\nu}_\sigma)) = 0 \qquad P_{\bar{b},\bar{\nu}_\sigma}\text{-almost surely} \tag{2.4}$$

and that the following implication holds:

$$\left\{ \begin{aligned} & \lim_{n\to\infty} d(\varphi(\bullet; \bar{b}_n, \bar{\nu}_{\sigma,n}), \varphi(\bullet; \bar{b}, \bar{\nu}_\sigma)) = 0 \\ \Longrightarrow \quad & \\ & \lim_{n\to\infty} \int_s^t \varphi(u; \bar{b}_n, \bar{\nu}_{\sigma,n})\,\mathrm{d}u = \int_s^t \varphi(u; \bar{b}, \bar{\nu}_\sigma)\,\mathrm{d}u \qquad \forall s,t\in\mathbb{R}. \end{aligned} \right. \tag{2.5}$$

A simple example of such a distance is given by the weighted L^p-norms,

$$d(\varphi_1, \varphi_2) = \left(\int_{-\infty}^{\infty} |\varphi_1(u) - \varphi_2(u)|^p w(u)\,\mathrm{d}u\right)^{1/p},$$

where $p \ge 1$ and $w:\mathbb{R}\to(0,\infty)$ is a continuous weight function with $\int_{-\infty}^{\infty} w(u)\,\mathrm{d}u < \infty$. Then assumption (2.4) follows by dominated convergence from the convergence result (2.1), while assumption (2.5) is immediate.

3. Consistency

We derive from the triangle inequality the definition of the minimum-distance estimator and assumption (2.4) that

$$\begin{aligned} d(\varphi(\bullet; \widehat{\bar{b}}_n, \widehat{\bar{\nu}}_{\sigma,n}), \varphi(\bullet; \bar{b}, \bar{\nu}_\sigma)) &\le d(\varphi(\bullet; \widehat{\bar{b}}_n, \widehat{\bar{\nu}}_{\sigma,n}), \widehat{\varphi}_n) + d(\widehat{\varphi}_n, \varphi(\bullet; \bar{b}, \bar{\nu}_\sigma)) \\ &\le 2d(\widehat{\varphi}_n, \varphi(\bullet; \bar{b}, \bar{\nu}_\sigma)) + \delta_n \\ &\longrightarrow 0 \qquad P_{\bar{b},\bar{\nu}_\sigma}\text{-a.s.} \end{aligned} \tag{3.1}$$

By assumption (2.5) this implies for the integrated characteristic function that

$$P_{\bar{b},\bar{\nu}_\sigma}\left(\int_s^t \varphi(u; \widehat{\bar{b}}_n, \widehat{\bar{\nu}}_{\sigma,n})\,\mathrm{d}u \underset{n\to\infty}{\longrightarrow} \int_s^t \varphi(u; \bar{b}, \bar{\nu}_\sigma)\,\mathrm{d}u \ \forall s,t\in\mathbb{R}\right) = 1. \tag{3.2}$$

By Theorem 6.3.3 in Chung (1974), page 163, we obtain from (3.2) that

$$P_{\widehat{\bar{b}}_n, \widehat{\bar{\nu}}_{\sigma,n}} \longrightarrow_v P_{\bar{b},\bar{\nu}_\sigma} \qquad P_{\bar{b},\bar{\nu}_\sigma}\text{-a.s.},$$

where '$\longrightarrow_v$' denotes vague convergence to a possibly defective (i.e., with a mass less than 1) measure. However, since this vague limit is a probability measure, it turns out that the mode of convergence is actually the weak one, that is,

$$P_{\widehat{\bar{b}}_n, \widehat{\bar{\nu}}_{\sigma,n}} \Longrightarrow P_{\bar{b},\bar{\nu}_\sigma} \qquad P_{\bar{b},\bar{\nu}_\sigma}\text{-a.s.} \tag{3.3}$$

As an immediate consequence of equation (3.3) and Proposition 2.1 above we obtain the following consistency result for the parameters of the Lévy process:

Theorem 3.1. *If the distance d satisfies properties* (2.4) *and* (2.5), *then the minimum distance fit* $(\widehat{\bar{b}}_n, \widehat{\bar{\nu}}_{\sigma,n})$ *is a strongly consistent estimator, that is, with probability one we have for $n \to \infty$*

$$\widehat{\bar{b}}_n \to \bar{b} \quad \textit{and} \quad \widehat{\bar{\nu}}_{\sigma,n} \Longrightarrow \bar{\nu}_\sigma .$$

Remark 3.2. Without further assumptions we cannot estimate the diffusion parameter σ in a uniformly consistent way. We have for example that the stable law with characteristic function $\varphi_\alpha(u) = \mathrm{e}^{-|u|^\alpha/2}$ converges for $\alpha \uparrow 2$ to the standard normal law ($\alpha = 2$) in total variation norm: by Scheffé's lemma it suffices to show pointwise convergence of the density functions, which follows from the L^1-convergence of the characteristic functions. Hence, for n observations no test can separate the hypotheses $H_0 : \alpha = 2$ and $H_1 : \alpha < 2$. Since we have $\sigma = 1$ for $\alpha = 2$ and $\sigma = 0$ for $\alpha < 2$, this implies for the estimation problem uniform inconsistency in the following sense:

$$\limsup_{n\to\infty} \inf_{\hat{\sigma}_n} \sup_{\bar{b},\bar{\nu}_\sigma} P_{\bar{b},\bar{\nu}_\sigma}(|\widehat{\sigma}_n - \sigma| \geq 1/2) > 0,$$

where the infimum is taken over all estimators based on n observations. Thus, from a statistical perspective the estimation of the volatility σ makes no sense, unless we restrict the class of Lévy processes under consideration, for example, to the finite intensity case as in Belomestny and Reiß (2006).

The practical implementation of the minimum distance method raises naturally the question of computational feasibility. It is certainly not possible to compute $\widehat{\bar{\nu}}_{\sigma,n}$ by an optimization over the full set $\mathcal{M}(\mathbb{R})$. In our simulations, for example, we approximate the measure $\bar{\nu}_\sigma$ by measures with step-wise constant densities. To assess the effect of such an approximation, consider a sequence of subsets $\mathcal{M}^{(n)} \subseteq \mathcal{M}(\mathbb{R})$ with the density property that there exist measures $\widetilde{\nu}^{(n)} \in \mathcal{M}^{(n)}$ with $\widetilde{\nu}^{(n)} \Longrightarrow \bar{\nu}_\sigma$, as $n \to \infty$. The definition from (2.3) is now replaced by

$$d(\widehat{\varphi}_n, \varphi(\bullet; \widehat{\bar{b}}_n, \widehat{\bar{\nu}}_{\sigma,n})) \leq \inf_{\tilde{b}\in\mathbb{R},\, \tilde{\nu}_\sigma \in \mathcal{M}^{(n)}} d(\widehat{\varphi}_n, \varphi(\bullet; \widetilde{b}, \widetilde{\nu}_\sigma)) + \delta_n .$$

We obtain instead of (3.1) that

$$\begin{aligned} d(\varphi(\bullet; \widehat{\bar{b}}_n, \widehat{\bar{\nu}}_{\sigma,n}), \varphi(\bullet; \bar{b}, \bar{\nu}_\sigma)) \;&\leq\; 2d(\widehat{\varphi}_n, \varphi(\bullet; \bar{b}, \bar{\nu}_\sigma)) + \delta_n + d\big(\varphi(\bullet; \bar{b}, \bar{\nu}_\sigma), \varphi\big(\bullet; \bar{b}, \widetilde{\nu}^{(n)}\big)\big) \\ &\longrightarrow 0 \qquad P_{\bar{b},\bar{\nu}_\sigma}\text{-a.s.} \end{aligned}$$

Hence, we obtain in complete analogy to Theorem 3.1 that with probability one for $n \to \infty$

$$\widehat{\bar{b}}_n \to \bar{b} \quad \text{and} \quad \widehat{\bar{\nu}}_{\sigma,n} \Longrightarrow \bar{\nu}_\sigma .$$

Given the existence of certain moments for $P_{\bar{b},\bar{\nu}_\sigma}$, we could also search our minimum-distance estimator in the class of those parameter values that fit the empirical moments. Using a similar error decomposition and the consistency of the empirical moments, this approach will also yield consistent estimators under mild conditions on the distance d.

4. A rate-optimal estimator

4.1. The construction

In this section we intend to devise estimators which attain optimal rates of convergence. We henceforth restrict the class of Lévy processes to those with finite second moments. This is equivalent to requiring that the Lévy measure satisfies $\int x^2 \nu(\mathrm{d}x) < \infty$. In this case the following reparametrization of the characteristic exponent is much more convenient:

$$\Psi(u; b, \sigma, \nu) = \mathrm{i}ub - \frac{\sigma^2}{2}u^2 + \int_{\mathbb{R}} (\mathrm{e}^{\mathrm{i}ux} - 1 - \mathrm{i}ux)\nu(\mathrm{d}x),$$

where the parameter $b = \bar{b} + \int_{\mathbb{R}} (x - \frac{x}{1+x^2})\nu(\mathrm{d}x)$ denotes now indeed the mean trend because of $\mathbb{E}[X_1] = -\mathrm{i}\varphi'(0) = b$. Let us mention that this is the original Kolmogorov canonical representation of a Lévy process (Kolmogorov (1932)), the historical background of which is nicely exposed by Mainardi and Rogosin (2006). Instead of $\bar{\nu}_\sigma$, we consider the finite measure ν_σ defined by

$$\nu_\sigma(\mathrm{d}x) = \sigma^2 \delta_0(\mathrm{d}x) + x^2 \nu(\mathrm{d}x),$$

which allows the nice identity $\mathrm{Var}(X_1) = -\varphi''(0) + \varphi'(0)^2 = \nu_\sigma(\mathbb{R})$. From now on, we shall express the characteristic exponent in terms of (b, ν_σ):

$$\Psi(u) = \Psi(u; b, \nu_\sigma) = \mathrm{i}ub + \int_{\mathbb{R}} \frac{\mathrm{e}^{\mathrm{i}ux} - 1 - \mathrm{i}ux}{x^2} \nu_\sigma(\mathrm{d}x).$$

While b can be easily estimated by $\frac{1}{n}\sum_{t=1}^n (X_t - X_{t-1}) = X_n/n$, the construction of an optimal nonparametric estimator of ν_σ requires more work. Before we start with our search for optimal rates of convergence for estimators of ν_σ, we have to decide about an appropriate metric to measure the deviation of any potential estimator $\widetilde{\nu}_{\sigma,n}$ from its target ν_σ.

The parameter ν_σ lies in the space of finite Borel measures, which is naturally equipped with the total variation norm. As we have seen above in the consistent estimation problem for σ, this topology is too strong here. Moreover, we are usually not interested in the problem of estimating ν_σ itself, but rather in estimating integrals $\int f \, \mathrm{d}\nu_\sigma$ for certain integrands f. In mathematical finance for example, the so-called Δ in the quadratic hedging approach requires calculating $\int \frac{C(t,S(1+z)) - C(t,S)}{Sz} \nu_\sigma(\mathrm{d}z)$, where $C(t, S)$ denotes the option price at time t and S the corresponding stock price, see Proposition 10.5 in Cont and Tankov (2004). This is why we choose to measure the performance of our estimator by metrizing weak convergence with certain classes F of continuous test functions f:

$$l(\widetilde{\nu}_{\sigma,n}, \nu_\sigma) = \sup\left\{\left|\int f \, \mathrm{d}\widetilde{\nu}_{\sigma,n} - \int f \, \mathrm{d}\nu_\sigma\right| : f \in F\right\}.$$

Note that for any class F of uniformly bounded, equicontinuous functions consistency with respect to weak convergence implies $l(\widetilde{\nu}_{\sigma,n}, \nu_\sigma) \to 0$ (Dudley (1989), Corollary 11.3.4). For instance, the bounded Lipschitz metric is generated by the test functions of Lipschitz norm less than one.

Let us introduce the Fourier transform for functions $f \in L^1(\mathbb{R})$ or measures $\mu \in \mathcal{M}(\mathbb{R})$ by

$$\mathcal{F}f(u) = \int f(x)\mathrm{e}^{\mathrm{i}ux}\,\mathrm{d}x, \qquad \mathcal{F}\mu(u) = \int \mathrm{e}^{\mathrm{i}ux}\mu(\mathrm{d}x), \qquad u \in \mathbb{R}.$$

Note that we have by Parseval's equality

$$\int f\,\mathrm{d}\nu_\sigma = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathcal{F}f(u)\overline{\mathcal{F}\nu_\sigma(u)}\,\mathrm{d}u,$$

provided $\mathcal{F}f \in L^1(\mathbb{R})$ (Katznelson (1976), Theorem VI.2.2). Estimation of ν_σ turns out to be particularly transparent when we employ the fact that

$$\Psi''(u) = \frac{\mathrm{d}^2}{\mathrm{d}u^2}\int \frac{\mathrm{e}^{\mathrm{i}ux} - 1 - \mathrm{i}ux}{x^2}\nu_\sigma(\mathrm{d}x) = -\mathcal{F}\nu_\sigma(u),$$

and consequently

$$\mathcal{F}\nu_\sigma(u) = -\frac{\mathrm{d}^2}{\mathrm{d}u^2}\log(\varphi(u)) = \frac{\varphi'(u)^2}{\varphi(u)^2} - \frac{\varphi''(u)}{\varphi(u)}. \tag{4.1}$$

Recall that $\int x^2\nu(\mathrm{d}x) < \infty$ implies $\mathbb{E}[X_t^2] < \infty$ and hence $\varphi \in C^2$. Moreover, in order to recover Ψ from φ we use the distinguished logarithm of the complex-valued function $u \mapsto \varphi(u)$, which is required to ensure $\log(\varphi(0)) = 0$ and continuity of $u \mapsto \log(\varphi(u))$, see Cont and Tankov (2004). This formula indicates that estimating ν_σ is strongly related to estimating φ in a C^2-sense. Before we study rates of convergence, we need to investigate uniform rates of convergence of the empirical characteristic function $\widehat{\varphi}_n$ and its derivatives.

4.2. Estimating the characteristic function

For i.i.d. random variables $(Z_t)_{t\in\mathbb{N}}$, denote by

$$C_n(u) := n^{-1/2}\sum_{t=1}^{n}(\mathrm{e}^{\mathrm{i}uZ_t} - \mathbb{E}[\mathrm{e}^{\mathrm{i}uZ_1}])$$

the normalized characteristic function process. Furthermore, denote by $C_n^{(k)}$ its kth derivative which exists if $\mathbb{E}|Z_1|^k < \infty$. For an appropriate weight function $w : \mathbb{R} \longrightarrow [0,\infty)$, we consider

$$\mathbb{E}\|C_n^{(k)}\|_{L_\infty(w)} := \mathbb{E}\sup_{u\in\mathbb{R}}\{|C_n^{(k)}(u)|w(u)\}.$$

For every $k \geq 0$ we have the following general result.

Theorem 4.1. *Suppose that $(Z_t)_{t\in\mathbb{N}}$ are i.i.d. random variables with $\mathbb{E}|Z_1|^{2k+\gamma} < \infty$ for some $\gamma > 0$ and let the weight function be defined as $w(u) = (\log(\mathrm{e} + |u|))^{-1/2-\delta}$ for some $\delta > 0$. Then*

$$\sup_{n\geq 1}\mathbb{E}\|C_n^{(k)}\|_{L_\infty(w)} < \infty.$$

Its proof is given in Section 6.1. Let us mention that the logarithmic decay of the weight function w is in accordance with the well-known result that $\widehat{\varphi}_n \to \varphi$ a.s. holds uniformly on intervals $[-T_n, T_n]$ whenever $\log(T_n)/n \to 0$, see Csörgő and Totik (1983).

4.3. Upper risk bounds

In view of (4.1) and Theorem 4.1, we define our estimators of b and ν_σ by a minimum distance fit based on a weighted C^2-norm. Defining

$$d^{(2)}(\varphi_1, \varphi_2) := \sum_{k=0}^{2} \|\varphi_1^{(k)} - \varphi_2^{(k)}\|_{L^\infty(w)},$$

we choose the estimators $\widehat{b}_n \in \mathbb{R}$ and $\widehat{\nu}_{\sigma,n} \in \mathcal{M}(\mathbb{R})$ such that

$$d^{(2)}(\varphi(\bullet; \widehat{b}_n, \widehat{\nu}_{\sigma,n}), \widehat{\varphi}_n) \le \inf_{\tilde{b}\in\mathbb{R}, \tilde{\nu}_\sigma \in \mathcal{M}(\mathbb{R})} d^{(2)}(\varphi(\bullet; \widetilde{b}, \widetilde{\nu}_\sigma), \widehat{\varphi}_n) + \delta_n, \tag{4.2}$$

where $\delta_n \to 0$ as $n \to \infty$. We verify by Theorem 4.1 that $d^{(2)}$ satisfies assumptions (2.4) and (2.5), hence Theorem 3.1 gives immediately a consistency result. Moreover, with the choice $\delta_n = \mathrm{O}(n^{-1/2})$ these estimators will turn out to be rate-optimal.

While b can always be estimated at rate $n^{-1/2}$, rates of convergence of $\int f \, \mathrm{d}\widehat{\nu}_{\sigma,n}$ as an estimator of $\int f \, \mathrm{d}\nu_\sigma$ depend both on the smoothness of f and on the decay of $|\varphi(u)|$ as $|u| \to \infty$. For the function f, we will assume that it belongs to the class

$$F_s := \left\{ f : \int (1 + |u|)^s |\mathcal{F} f(u)| \, \mathrm{d}u \le 1 \right\},$$

for some $s \ge 0$. Note that $\int |\mathcal{F} f(u)| \, \mathrm{d}u \le 1$ implies by the Riemann–Lebesgue lemma that f is continuous with $\|f\|_\infty \le 1$. By the Fourier theory the condition $f \in F_s$ is slightly stronger than requiring $f \in C^s$ with $\|f\|_{C^s} \le 1$ for a suitable norming of C^s. We therefore introduce a loss function for an estimator $\widehat{\mu}$ of the finite measure μ by

$$\ell_s(\widehat{\mu}, \mu) := \sup_{f \in F_s} \left| \int f \, \mathrm{d}\widehat{\mu} - \int f \, \mathrm{d}\mu \right|.$$

Note that by duality the loss ℓ_s can be interpreted as a negative smoothness norm of order $-s$.

The faster $|\varphi(u)|$ decays, the more difficult it will be to estimate ν_σ. We consider in particular the following three cases:

(a) Gaussian part. If $\sigma^2 > 0$, then the characteristic function φ has Gaussian tails, that is,

$$\log|\varphi(u)| = \mathrm{Re}(\log \varphi(u)) = -\sigma^2 u^2/2\big(1 + \mathrm{o}(u)\big) \qquad \text{as } |u| \to \infty.$$

(To see this, note that $F(x, u) := (\mathrm{e}^{\mathrm{i}ux} - 1 - \mathrm{i}ux)/(ux)^2$ is uniformly bounded with $\lim_{|u|\to\infty} F(x, u) = 0$ for $x \ne 0$ such that by dominated convergence $\lim_{|u|\to\infty} \int_{\mathbb{R}} F(x, u)x^2 \nu(\mathrm{d}x) = 0$ and thus $\log \varphi(u) = -\sigma^2 u^2/2 + \mathrm{o}(u^2)$.)

(b) Exponential decay. Here the characteristic function φ decays at most exponentially, that is, for some $\alpha > 0$, $C > 0$,

$$|\varphi(u)| \geq C\mathrm{e}^{-\alpha|u|} \qquad \text{for all } u \in \mathbb{R}.$$

Examples of distributions with this property include normal inverse Gaussian (Cont and Tankov (2004), page 117) and generalized tempered stable distributions (Cont and Tankov (2004), page 122).

(c) Polynomial decay. In this case the characteristic function satisfies, for some $\beta \geq 0$, $C > 0$,

$$|\varphi(u)| \geq C(1 + |u|)^{-\beta} \qquad \text{for all } u \in \mathbb{R}.$$

Typical examples for this are the compound Poisson distribution, the gamma distribution, the variance gamma distribution and the generalized hyperbolic distribution (Cont and Tankov (2004), pages 75, 116, 117, 127).

The proof of the following main theorem is postponed to Section 6.2.

Theorem 4.2. *Suppose that* $\mathbb{E}_{b,\nu_\sigma}|X_1|^{4+\gamma} < \infty$ *for some* $\gamma > 0$. *We choose the weight function* w *as* $w(u) = (\log(\mathrm{e} + |u|))^{-1/2-\delta}$, *where* δ *is any positive number. The estimators* $\widehat{b}_n$ *and* $\widehat{\nu}_{\sigma,n}$ *of* b *and* ν_σ, *respectively, are chosen according to* (4.2) *with* $\delta_n = \mathrm{O}(n^{-1/2})$. *Then*

$$\mathbb{E}_{b,\nu_\sigma}|\widehat{b}_n - b| = \mathrm{O}(n^{-1/2})$$

and for any $s > 0$

$$\ell_s(\widehat{\nu}_{\sigma,n}, \nu_\sigma) = O_{P_{b,\nu_\sigma}}\left(n^{-1/2} \bullet \sup_{u\in[0,U_n]}\left\{\frac{(1+u)^{2-s}}{w(u)|\varphi(u; b, \nu_\sigma)|}\right\}\right),$$

where

$$U_n := \inf\left\{u > 0 : \frac{(1+u)^2 n^{-1/2}}{w(u)|\varphi(u; b, \nu_\sigma)|} \geq 1\right\}.$$

The constants in the risk bounds depend continuously on $|b|$ *and* $\nu_\sigma(\mathbb{R})$. *In the specific cases we obtain the following rates of convergence for* $\ell_s(\widehat{\nu}_{\sigma,n}, \nu_\sigma)$ *in* P_{b,ν_σ}*-probability*:

(a) Gaussian part: $(\log n)^{-s/2}$,
(b) Exponential decay: $(\log n)^{-s}$,
(c) Polynomial decay of order $\beta \geq 0$: $[(\log n)^{1/2+2\delta} n^{-1/2}]^{s/\beta} \vee n^{-1/2}$.

Remark 4.3. The results are presented for convergence in probability, but the proof immediately yields convergence of moments of order $1/2$ of the loss in cases (a) and (b), see equation (6.4). Higher moments are achieved whenever the order of the moment bound in Theorem 4.1 can be increased.

4.4. Lower risk bounds

We prove that the rates of convergence obtained in Theorem 4.2 for cases (a)–(c) are optimal, at least up to a logarithmic factor in the latter case. The proof in Section 6.3 can be naturally generalized to cover further decay scenarios of the characteristic function.

Theorem 4.4. *For $C, \bar{C} > 0$ large enough and for any $\alpha > 0$, $\beta \geq 0$ introduce the following nonparametric classes of ν_σ:*

$$\mathcal{A}(C,\sigma) := \{\nu_\sigma \in \mathcal{M}(\mathbb{R}) | \nu_\sigma(\mathbb{R}) \leq C\} \qquad (\sigma > 0),$$
$$\mathcal{B}(C,\alpha) := \left\{\nu_\sigma \in \mathcal{M}(\mathbb{R}) | \nu_\sigma(\mathbb{R}) \leq C, |\varphi(u)| \geq \bar{C}\mathrm{e}^{-\alpha|u|}\right\} \qquad (\sigma = 0),$$
$$\mathcal{C}(C,\bar{C},\beta) := \{\nu_\sigma \in \mathcal{M}(\mathbb{R}) | \nu_\sigma(\mathbb{R}) \leq C, |\varphi(u)| \geq \bar{C}^{-1}(1+|u|)^{-\beta}\} \qquad (\sigma = 0).$$

Then we obtain for some fixed $b \in \mathbb{R}$ and for any $s > 0$ the following minimax lower bounds, where $\widetilde{\nu}_{\sigma,n}$ denotes any estimator of ν_σ based on n observations:

$$\textbf{(a)}\ \exists \varepsilon > 0: \qquad \liminf_{n\to\infty} \inf_{\tilde{\nu}_{\sigma,n}} \sup_{\nu_\sigma \in \mathcal{A}(C,\sigma)} P_{b,\nu_\sigma}\big((\log n)^{s/2} \ell_s(\widetilde{\nu}_{\sigma,n}, \nu_\sigma) > \varepsilon\big) > 0,$$

$$\textbf{(b)}\ \exists \varepsilon > 0: \qquad \liminf_{n\to\infty} \inf_{\tilde{\nu}_{\sigma,n}} \sup_{\nu_\sigma \in \mathcal{B}(C,\alpha)} P_{b,\nu_\sigma}\big((\log n)^{s} \ell_s(\widetilde{\nu}_{\sigma,n}, \nu_\sigma) > \varepsilon\big) > 0,$$

$$\textbf{(c)}\ \exists \varepsilon > 0: \qquad \liminf_{n\to\infty} \inf_{\tilde{\nu}_{\sigma,n}} \sup_{\nu_\sigma \in \mathcal{C}(C,\bar{C},\beta)} P_{b,\nu_\sigma}\big(n^{(s/2\beta)\wedge(1/2)} \ell_s(\widetilde{\nu}_{\sigma,n}, \nu_\sigma) > \varepsilon\big) > 0.$$

4.5. Discussion

The convergence rates for $\widehat{\nu}_{\sigma,n}$ can be understood in analogy with a deconvolution problem where the Fourier transform of the error density decays like the characteristic function φ in our case, see, for example, Fan (1991). The interesting point here is that this decay property is not assumed to be known and depends on the parameters to be estimated. At first sight, it is rather surprising that our minimum distance estimator adapts automatically to the decay of φ, even for the whole range of loss functions ℓ_s, $s > 0$. This is due to the fact that the noise level in the empirical characteristic function $\widehat{\varphi}_n$ is of the same size for different frequencies and this is where we fit our estimator. In contrast, when fitting the characteristic exponent Ψ, which is more attractive from a computational point of view and for example advocated in Jongbloed, van der Meulen and van der Vaart (2005), we face a highly heteroskedastic noise level in $\log(\widehat{\varphi}_n(u))$ governed by $|\varphi(u)|^{-1}$ because of $\log(\widehat{\varphi}_n(u)) - \Psi(u) \approx (\widehat{\varphi}_n(u) - \varphi(u))/\varphi(u)$.

Another point of view on our estimation problem is that we want to estimate the linear functional $\int f\, \mathrm{d}\nu_\sigma$ based on an inverse problem setting for estimating ν_σ. In an abstract Hilbert scale context, adaptive estimation for this has been considered by Goldenshluger and Pereverzev (2003) and their rate for the polynomially ill-posed case reads in our notation $(n/\log(n))^{-(r+s)/(2r+2\beta)} \vee n^{-1/2}$, with r the regularity of ν_σ, s the regularity of f and β the degree of ill-posedness. In our case, we measure the regularity s of f in the Fourier domain by an L^1-criterion such that a dual L^∞-criterion for the regularity of ν_σ yields $r = 0$ because $\|\mathcal{F}\nu_\sigma\|_\infty$ is finite. Hence, the rate $(n/\log(n))^{-s/2\beta} \vee n^{-1/2}$, up to the logarithmic factor of power δ, obtained in case (c) of Theorem 4.2, confirms this analogy. We suspect that the gap by a logarithmic factor in the polynomial case between our upper and lower bound is mainly due to a suboptimal lower bound, because ℓ_s can be expressed in the Fourier domain via

$$\ell_s(\widehat{\mu}, \mu) = \sup_{f \in F_s} \left| \int \mathcal{F} f(u) \overline{\mathcal{F}(\widehat{\mu} - \mu)(u)} \, \mathrm{d}u \right| = \sup_{u \in \mathbb{R}} (1 + |u|)^{-s} |\mathcal{F}(\widehat{\mu} - \mu)(u)|,$$

giving a supremum-type norm.

It is certainly remarkable that no regularization parameter is involved in our estimation procedure, which becomes more intuitive by noticing that the results of Section 3 imply consistency already for $s = 0$. On the other hand, better rates of convergence can be obtained when we restrict the model to measures ν_σ which have a regular Lebesgue density g_σ. A natural plug-in approach yields the kernel-type estimator $\widehat{g}_{\sigma,n,h}(x) := K_h * \widehat{\nu}_{\sigma,n}(x)$, convolving the minimum-distance estimator with a smooth kernel K_h of bandwidth $h > 0$. Noting that $\int f \widehat{g}_{\sigma,n,h} = \int (f * K_h) \, \mathrm{d}\widehat{\nu}_{\sigma,n}$, we infer that the bound on the stochastic error

$$\left| \int f(\widehat{g}_{\sigma,n,h} - K_h * g_\sigma) \right| = \left| \int (f * K_h) \, \mathrm{d}(\widehat{\nu}_{\sigma,n} - \nu_\sigma) \right|$$

is controlled by the regularity of $f * K_h$. To be more specific, consider a function f with $|\mathcal{F} f(u)| \asymp (1 + |u|)^{-s-1}$ (e.g., $f(x) = \mathrm{e}^{-|x|}$ with $s = 1$), suppose $\sup_u (1 + |u|)^r |\mathcal{F} g_\sigma(u)| < \infty$ for $r > 0$ and assume polynomial decay of order $\beta \geq s$ of the characteristic function. Then $\int (1 + |u|)^\beta |\mathcal{F} f(u) \mathcal{F} K_h(u)| \, \mathrm{d}u \asymp h^{s-\beta}$ holds such that $c h^{-s+\beta} f * K_h$ lies in F_β, $c > 0$ some small constant, and Theorem 4.2 implies that

$$\left| \int f(\widehat{g}_{\sigma,n,h} - K_h * g_\sigma) \right| = \mathrm{O}_P(h^{s-\beta} n^{-1/2} (\log n)^{1/2+2\delta}).$$

Together with an easy bias estimate of order h^{s+r} this yields for the estimation error $|\int f(\widehat{g}_{\sigma,n,h} - g_\sigma)|$ up to logarithmic factors the rate $n^{-(r+s)/(2r+2\beta)}$, provided the bandwidth is chosen in an optimal way. We conclude that our results also allow us to obtain risk bounds under smoothness restrictions, which are coherent with the abstract results in Goldenshluger and Pereverzev (2003). The rates should also be compared with the case of continuous-time observations on $[0, T]$, where Figueroa-López and Houdré (2006) obtained the classical nonparametric rate $T^{-r/(2r+1)}$ for estimating g_σ on a bounded interval.

5. Implementation

Although the main focus of our work is theoretical, we point out how the minimum distance estimator can be implemented and show a numerical example. The main computational problem is that the procedure requires that we minimize a nonlinear functional over the space of all finite measures. One possibility is to use a global optimization procedure, for example, based on simulated annealing, see Hall and Yao (2003) for an application to minimum-distance fits based on characteristic functions. Here we shall look for a good preliminary estimator and minimize the $d^{(2)}$-criterion locally around this pilot estimator, which turns out to be more stable in simulations than global optimization routines.

We use the identification formula (4.1) to build a first-stage plug-in estimator $(\widetilde{b}_n, \widetilde{\nu}_{\sigma,n})$. While the mean b will be easily estimated by

$$\widetilde{b}_n := \frac{1}{n}\sum_{t=1}^{n}(X_t - X_{t-1}) = X_n/n,$$

we have to be more careful with an estimator of ν_σ. Since $\mathcal{F}\nu_\sigma(u) = -\varphi''(u)/\varphi(u) + (\varphi'(u)/\varphi(u))^2$ one might be tempted to estimate its Fourier transform just by plugging in the empirical characteristic function $\widehat{\varphi}_n$ for φ. It turns out, however, that the occurrence of $\widehat{\varphi}_n(u)$ in the denominator might have unfavorable effects, particularly if $|\varphi(u)|$ is small. To get some intuition for a possible remedy, consider the problem of estimating $1/\varphi(u)$. $1/\widehat{\varphi}_n(u)$ is certainly a good estimator as long as $|\widehat{\varphi}_n(u)|$ is not too small. On the other hand, since the noise level of $\widehat{\varphi}_n(u)$ is $\mathrm{O}(n^{-1/2})$ we should no longer rely on $1/\widehat{\varphi}_n(u)$ if $\widehat{\varphi}_n(u) = \mathrm{O}(n^{-1/2})$. To take this into account, one can use $\mathbb{1}_{\{|\hat{\varphi}_n(u)|\geq\kappa n^{-1/2}\}}/\widehat{\varphi}_n(u)$ as an estimator for $1/\varphi(u)$ which can be proven to satisfy

$$\mathbb{E}_{b,\nu_\sigma}\left|\frac{\mathbb{1}_{\{|\hat{\varphi}_n(u)|\geq\kappa n^{-1/2}\}}}{\widehat{\varphi}_n(u)} - \frac{1}{\varphi(u)}\right|^p = \mathrm{O}\left(\left(\frac{n^{-1/2}}{|\varphi(u)|^2} \wedge \frac{1}{|\varphi(u)|}\right)^p\right),$$

for any positive threshold value κ and all $p \in \mathbb{N}$. This is what we can at best expect from an estimator of $1/\varphi(u)$. Using this idea we define our preliminary estimator of $\mathcal{F}\nu_\sigma(u)$ by

$$\mathcal{F}\widetilde{\nu}_{\sigma,n}(u) := -\left(\frac{\widehat{\varphi}_n''(u)}{\widehat{\varphi}_n(u)} - \left(\frac{\widehat{\varphi}_n'(u)}{\widehat{\varphi}_n(u)}\right)^2\right)\mathbb{1}_{\{|\hat{\varphi}_n(u)|\geq\kappa n^{-1/2}\}}, \tag{5.1}$$

where κ is a positive constant. In Section 6.4 below we shall prove the following result.

Proposition 5.1. *We have* $\mathbb{E}_{b,\nu_\sigma}(\tilde{b}_n - b)^2 = \mathrm{O}(n^{-1})$ *and for* $u \in \mathbb{R}$

$$\mathbb{E}_{b,\nu_\sigma}|\mathcal{F}\widetilde{\nu}_{\sigma,n}(u) - \mathcal{F}\nu_\sigma(u)| = \mathrm{O}\left(\left(\frac{n^{-1/2}}{|\varphi(u)|} \wedge 1\right)\left(1 + |\Psi'(u)|^2\right)\right).$$

This will give pointwise rates of convergence in a similar fashion as before and serves well as a starting point of a local optimization routine. Note that this pilot estimator is very easy and fast to implement. Yet it has certain drawbacks, most important $\mathcal{F}\widetilde{\nu}_{\sigma,n}$ is usually not positive semidefinite so that $\widetilde{\nu}_{\sigma,n}$ is not necessarily a non-negative measure.

In practice, our two-stage procedure works reasonably well. For a numerical example we simulate a Lévy process $(X_t)_{t\geq 0}$ with $\sigma = 1$, $b = 1$ and $\nu(\mathrm{d}x) = x^{-1}\mathrm{e}^{-x}\mathbb{1}_{\{x>0\}}\,\mathrm{d}x$. The process X is a superposition of an infinite-intensity gamma process and a standard Brownian motion. The law of its increments $X_t - X_{t-1}$ is the convolution of an $N(0,1)$- and an Exp(1)-distribution. We have $n = 1000$ observations, see Figure 1 (left) for a histogram of the increments. The sample is rather disperse with some increments close to 10 and a sample mean of $\tilde{b}_n = 0.936$ (true $b = 1$). The true characteristic function has Gaussian decay and its absolute value is shown together with that of the empirical characteristic function in Figure 1 (right).

We discretize the pilot estimate $\widetilde{\nu}_{\sigma,n}$ of the jump measure by using a Haar wavelet basis on the interval $[-10, 10]$ with 15 basis functions. Moreover, we allow for a point measure in zero to have

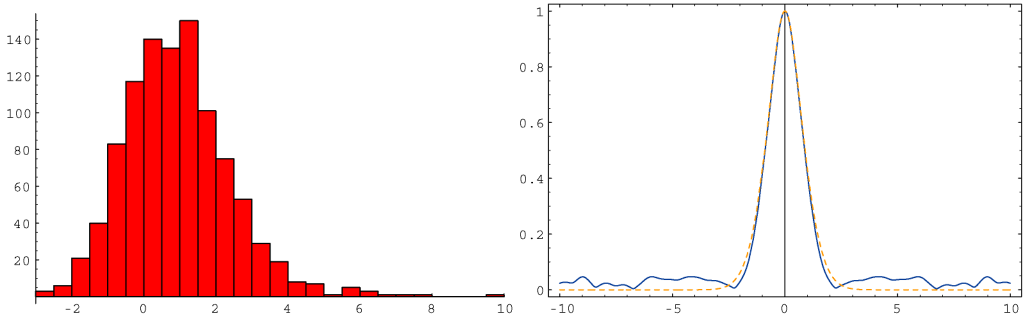

Figure 1. Left: Histogram of the data. Right: modulus of the empirical (solid) and true (dashed) characteristic function.

a better resolution there. Its pilot mass is set to zero. Using the `FindMinimum` local optimization procedure in Mathematica, we minimize the $d^{(2)}$-criterion locally around the discretized pilot estimator, constraining to non-negative Lévy measures. In Figure 2 (left) we display for the given data the imaginary part of the empirical characteristic function together with the imaginary parts of the other characteristic functions of interest (true, pilot, final estimator). The errors in fitting the real part are less pronounced because there are fewer oscillations around zero (note $(\operatorname{Re}\varphi)'(0) = 0$). Typically, the pilot estimator gives already a reasonably good fit and the final estimator has a characteristic function which is closer to the empirical characteristic function than the true one.

Figure 2 (right) finally shows the densities of the rescaled Lévy measures ν_σ, but suppresses the point masses in zero. Note that the original Lévy density and also its plug-in estimators have a singularity at zero because of $\nu(\mathrm{d}x) = x^{-2}\nu_\sigma(\mathrm{d}x)$ for $x \neq 0$. The parameters are estimated as $\widehat{b}_n = 0.922$ (true $b = 1$) and $\widehat{\nu}_{\sigma,n}(\{0\})^{1/2} = 1.092$ (true $\sigma = 1$). The pilot estimator has no point mass in zero and its density is therefore large around zero. It is seen that the final estimator improves upon the pilot estimator, in particular by excluding negative values and catching the point mass in zero. Given 1000 observations and a Gaussian deconvolution problem, the estimation problem is quite hard. The rough, step-wise form of the final estimator is not so pleasant for the human eye, but we only want to use this estimator as an integrator of smooth functions and, as discussed above, we could apply a kernel to obtain a smooth density function. As an example for a functional to be estimated, we calculated $\int_1^\infty x^{-2}\widehat{\nu}_\sigma(\mathrm{d}x)$ which estimates $\nu([1,\infty))$, the probability of jumps larger than one. In this sample, the true value 0.22 was estimated by 0.16. Let us remark that the high-frequency estimator, using the relative frequency of increments $X_t - X_{t-1}$ that are larger than one, yields the estimate 0.46. The large error of the latter confirms a strong violation of the underlying high-frequency assumption that between two observations very rarely more than one larger jump occurs and that the diffusion part is negligible. Hence, the frequency of the observations must indeed be considered as low for the construction of the estimator.

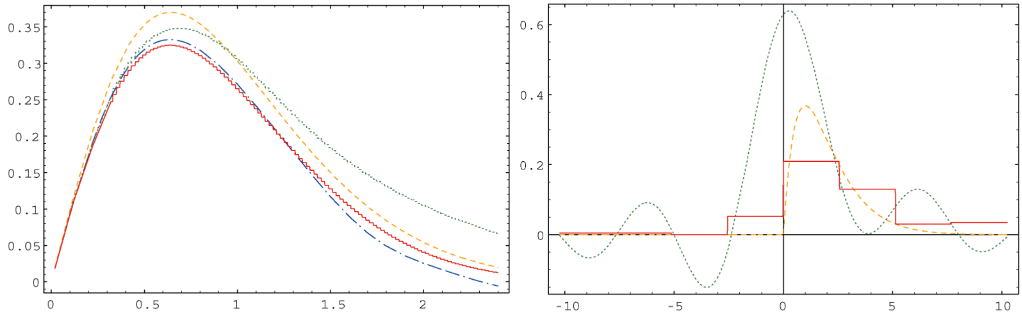

Figure 2. Left: Imaginary part of the empirical (dot-dashed), true (dashed), pilot (dotted) and final estimated (solid) characteristic function. Right: pilot (dotted), final (solid) estimator and true (dashed) density of ν_σ; the pilot estimator does not have a point mass in zero.

6. Proofs

6.1. Proof of Theorem 4.1

We begin the proof with a few definitions. Given two functions $l, u : \mathbb{R} \longrightarrow \mathbb{R}$ the bracket $[l, u]$ denotes the set of functions f with $l \le f \le u$. For a set G of functions the L^2-bracketing number $N_{[\cdot]}(\varepsilon, G)$ is the minimum number of brackets $[l_i, u_i]$, satisfying $\mathbb{E}[(u_i(Z_1) - l_i(Z_1))^2] \le \varepsilon^2$, that are needed to cover G. The associated bracketing integral is defined as

$$J_{[\cdot]}(\delta, G) = \int_0^\delta \sqrt{\log\bigl(N_{[\cdot]}(\varepsilon, G)\bigr)}\, \mathrm{d}\varepsilon.$$

Furthermore, a function $\overline{f}$ is called an envelope function for G if $|f| \le \overline{f}$ holds for all $f \in G$.

To apply Corollary 19.35 from van der Vaart (1998), we decompose C_n in its real and imaginary parts,

$$\mathrm{Re}(C_n(u)) = n^{-1/2} \sum_{t=1}^{n} \bigl(\cos(uZ_t) - \mathbb{E}\cos(uZ_1)\bigr),$$

$$\mathrm{Im}(C_n(u)) = n^{-1/2} \sum_{t=1}^{n} \bigl(\sin(uZ_t) - \mathbb{E}\sin(uZ_1)\bigr).$$

Accordingly, we consider the following class of functions:

$$G_k = \left\{ z \mapsto w(u) \frac{\partial^k}{\partial u^k} \cos(uz) \middle| u \in \mathbb{R} \right\} \cup \left\{ z \mapsto w(u) \frac{\partial^k}{\partial u^k} \sin(uz) \middle| u \in \mathbb{R} \right\}.$$

An envelope function $\overline{f}_k$ for G_k is given by $\overline{f}_k = |x|^k$. Now we obtain from Corollary 19.35 in van der Vaart (1998) that

$$\mathbb{E}\big\|C_n^{(k)}\big\|_{L_\infty(w)} \le C\big\{\mathbb{E}(\overline{f}_k(Z_1))^2 + J_{[\cdot]}\big(\sqrt{\mathbb{E}Z_1^{2k}}, G_k\big)\big\}. \tag{6.1}$$

Since $\mathbb{E}Z_1^{2k} < \infty$ it remains to bound the bracketing integral on the right-hand side of (6.1). Inspired by Yukich (1985), we proceed by setting, for every $\varepsilon > 0$,

$$M := M(\varepsilon, k) := \inf\big\{m > 0 \mid \mathbb{E}\big[Z_1^{2k}\mathbb{1}_{\{|Z_1|>m\}}\big] \le \varepsilon^2\big\}.$$

Furthermore, we set, for grid points $u_j \in \mathbb{R}$ to be specified below,

$$g_j^{\pm}(z) = \left(w(u_j)\frac{\partial^k}{\partial u^k}\cos(u_j z) \pm \varepsilon|z|^k\right)\mathbb{1}_{[-M,M]}(z) \pm \|w\|_\infty |z|^k \mathbb{1}_{[-M,M]^c}(z),$$

$$h_j^{\pm}(z) = \left(w(u_j)\frac{\partial^k}{\partial u^k}\sin(u_j z) \pm \varepsilon|z|^k\right)\mathbb{1}_{[-M,M]}(z) \pm \|w\|_\infty |z|^k \mathbb{1}_{[-M,M]^c}(z).$$

We obtain for the width of the brackets that

$$\begin{aligned}\mathbb{E}\big[\big(g_j^+(Z_1) - g_j^-(Z_1)\big)^2\big] &\le \mathbb{E}\big[4\varepsilon^2 Z_1^{2k}\mathbb{1}_{[-M,M]}(Z_1) + 4\|w\|_\infty^2 Z_1^{2k}\mathbb{1}_{[-M,M]^c}(Z_1)\big] \\ &\le 4\varepsilon^2(\mathbb{E}Z_1^{2k} + \|w\|_\infty^2),\end{aligned}$$

and, analogously,

$$\mathbb{E}\big[\big(h_j^+(Z_1) - h_j^-(Z_1)\big)^2\big] \le 4\varepsilon^2(\mathbb{E}Z_1^{2k} + \|w\|_\infty^2).$$

It remains to choose the grid points u_j in such a way that the brackets cover the set G_k. We consider an arbitrary $u \in \mathbb{R}$ and any grid point u_j. Then with the Lipschitz constant $\mathrm{Lip}(w)$ of the weight function w

$$\begin{aligned}&\left|w(u)\frac{\partial^k}{\partial u^k}\cos(uz) - w(u_j)\frac{\partial^k}{\partial u^k}\cos(u_j z)\right| \\ &\quad \le |z|^k \min\big\{|u-u_j|\big(\mathrm{Lip}(w) + \|w\|_\infty|z|\big), w(u) + w(u_j)\big\}.\end{aligned}$$

Therefore, the function $z \mapsto w(u)\frac{\partial^k}{\partial u^k}\cos(uz)$ is contained in the bracket $[g_j^-, g_j^+]$ if

$$\min\big\{|u-u_j|\big(\mathrm{Lip}(w) + \|w\|_\infty M\big), w(u) + w(u_j)\big\} \le \varepsilon.$$

Consequently, we choose the grid points as

$$u_j = j\varepsilon/\big(\mathrm{Lip}(w) + \|w\|_\infty M(\varepsilon, k)\big),$$

for $|j| \le J(\varepsilon)$, where $J(\varepsilon)$ is the smallest integer such that $u_{J(\varepsilon)}$ is greater than or equal to

$$U(\varepsilon) = \inf\left\{u > 0 \,\middle|\, \sup_{v:\,|v|\ge u} w(v) \le \varepsilon/2\right\}.$$

This yields the estimate $N_{[\cdot]}(\varepsilon, G_k) \le 2(2J(\varepsilon) + 1)$. It follows from the generalized Markov inequality that

$$M(\varepsilon, k) \le (\mathbb{E}[|Z_1|^{2k+\gamma}]/\varepsilon^2)^{1/\gamma}.$$

Now we obtain from the inequality

$$J(\varepsilon) \le 2U(\varepsilon)\big(\mathrm{Lip}(w) + \|w\|_\infty M(\varepsilon, k)\big)/\varepsilon + 1$$

that $\log(N_{[\cdot]}(\varepsilon, G_k)) = \mathrm{O}(\log(J(\varepsilon))) = \mathrm{O}(\varepsilon^{-(\delta+1/2)^{-1}} + \log(\varepsilon^{-1-2/\gamma})) = \mathrm{O}(\varepsilon^{-\kappa})$ for $\kappa = (\delta + 1/2)^{-1} < 2$. This implies

$$\int_0^\delta \sqrt{\log\big(N_{[\cdot]}(\varepsilon, G_k)\big)}\,\mathrm{d}\varepsilon < \infty,$$

as required.

6.2. Proof of Theorem 4.2

To simplify the notation, we use the abbreviations $\Psi_n(u) = \Psi(u; \widehat{b}_n, \widehat{\nu}_{\sigma,n})$ and $\varphi_n(u) = \exp(\Psi_n(u))$.

First of all, we obtain from the triangle inequality that

$$d^{(2)}(\varphi_n, \varphi) \le d^{(2)}(\widehat{\varphi}_n, \varphi) + d^{(2)}(\widehat{\varphi}_n, \varphi_n) \le 2d^{(2)}(\widehat{\varphi}_n, \varphi) + \delta_n. \tag{6.2}$$

Proof for $\widehat{b}_n$. We have that $\varphi'(0) = \mathrm{i}b$ and $\varphi_n'(0) = \mathrm{i}\widehat{b}_n$. Therefore, we obtain from (6.2) and Theorem 4.1 that

$$\begin{aligned}\mathbb{E}_{b,\nu_\sigma}|\widehat{b}_n - b| &= \mathbb{E}_{b,\nu_\sigma}|\varphi_n'(0) - \varphi'(0)| \le \mathbb{E}_{b,\nu_\sigma} d^{(2)}(\varphi_n, \varphi)\\ &\le 2\mathbb{E}_{b,\nu_\sigma} d^{(2)}(\widehat{\varphi}_n, \varphi) + \delta_n = \mathrm{O}(n^{-1/2}).\end{aligned}$$

Proof for $\widehat{\nu}_{\sigma,n}$. We consider the following set of "unfavorable" events:

$$A_n := \{\widehat{\nu}_{\sigma,n}(\mathbb{R}) > \nu_\sigma(\mathbb{R}) + 1\} \cup \{|\widehat{b}_n| > |b| + 1\}.$$

From $\varphi'(0)^2 - \varphi''(0) = \nu_\sigma(\mathbb{R})$ and the analogous formula for φ_n it follows that

$$\begin{aligned}|\widehat{\nu}_{\sigma,n}(\mathbb{R}) - \nu_\sigma(\mathbb{R})| &= \big|\big(\varphi'(0)^2 - \varphi_n'(0)^2\big) - \big(\varphi''(0) - \varphi_n''(0)\big)\big|\\ &\le \big(2|\varphi'(0)| + d^{(2)}(\varphi, \varphi_n) + 1\big)d^{(2)}(\varphi, \varphi_n).\end{aligned} \tag{6.3}$$

Consequently, the (generalized) Markov inequality yields

$$\begin{aligned}P_{b,\nu_\sigma}(A_n) &\le \mathbb{E}_{b,\nu_\sigma}[|\widehat{\nu}_{\sigma,n}(\mathbb{R}) - \nu_\sigma(\mathbb{R})| \wedge 1] + \mathbb{E}_{b,\nu_\sigma}[|\widehat{b}_n - b|]\\ &\le \mathbb{E}_{b,\nu_\sigma}\big[\big(\big(2|b| + d^{(2)}(\varphi, \varphi_n) + 1\big)d^{(2)}(\varphi, \varphi_n)\big) \wedge 1\big] + \mathbb{E}_{b,\nu_\sigma}\big[d^{(2)}(\varphi, \varphi_n)\big]\\ &\le \mathbb{E}_{b,\nu_\sigma}\big[(4|b| + 2)d^{(2)}(\varphi, \varphi_n)\big] + \mathbb{E}_{b,\nu_\sigma}\big[d^{(2)}(\varphi, \varphi_n)\big]\\ &\le (4|b| + 3)\big(\mathbb{E}_{b,\nu_\sigma}\big[d^{(2)}(\widehat{\varphi}_n, \varphi)\big] + \delta_n\big) = \mathrm{O}(n^{-1/2}),\end{aligned}$$

which implies that

$$
\begin{aligned}
&\mathbb{1}_{A_n} \bullet \sup_{f \in F_s} \left| \int f \, \mathrm{d}\widehat{\nu}_{\sigma,n} - \int f \, \mathrm{d}\nu_\sigma \right| \\
&\quad \le \mathbb{1}_{A_n} \bullet \sup_{f \in F_s} \|f\|_\infty \bullet \big(\widehat{\nu}_{\sigma,n}(\mathbb{R}) + \nu_\sigma(\mathbb{R})\big) \\
&\quad \le 2\nu_\sigma(\mathbb{R})\mathbb{1}_{A_n} + \big(2|\varphi'(0)| + d^{(2)}(\varphi, \varphi_n) + 1\big) d^{(2)}(\varphi, \varphi_n) \\
&\quad = \mathrm{O}_{P_{b,\nu_\sigma}}(n^{-1/2}).
\end{aligned}
\tag{6.4}
$$

It remains to analyze the loss under A_n^c. It follows from Parseval's identity that

$$
\begin{aligned}
&\left| \int f \, \mathrm{d}\widehat{\nu}_{\sigma,n} - \int f \, \mathrm{d}\nu_\sigma \right| \\
&\quad = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathcal{F} f(u) \overline{\big(\mathcal{F}\widehat{\nu}_{\sigma,n}(u) - \mathcal{F}\nu_\sigma(u)\big)} \, \mathrm{d}u \\
&\quad = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathcal{F} f(u) \overline{\left\{ \left(\left(\frac{\varphi_n'(u)}{\varphi_n(u)} \right)^2 - \left(\frac{\varphi'(u)}{\varphi(u)} \right)^2 \right) - \left(\frac{\varphi_n''(u)}{\varphi_n(u)} - \frac{\varphi''(u)}{\varphi(u)} \right) \right\}} \, \mathrm{d}u.
\end{aligned}
\tag{6.5}
$$

The differences occurring in the integrand on the right-hand side of (6.5) can be estimated using $\varphi'/\varphi = \Psi'$, $\varphi_n'/\varphi_n = \Psi_n'$:

$$
\begin{aligned}
&\left| \left(\frac{\varphi_n'(u)}{\varphi_n(u)} \right)^2 - \left(\frac{\varphi'(u)}{\varphi(u)} \right)^2 \right| \\
&\quad = \left| \frac{\varphi_n'(u)}{\varphi_n(u)} - \frac{\varphi'(u)}{\varphi(u)} \right| |\Psi_n'(u) + \Psi'(u)| \\
&\quad \le \left\{ \left| \frac{\varphi_n(u) - \varphi(u)}{\varphi(u)} \right| |\Psi_n'(u)| + \left| \frac{\varphi_n'(u) - \varphi'(u)}{\varphi(u)} \right| \right\} |\Psi_n'(u) + \Psi'(u)|
\end{aligned}
\tag{6.6}
$$

and

$$
\begin{aligned}
\left| \frac{\varphi_n''(u)}{\varphi_n(u)} - \frac{\varphi''(u)}{\varphi(u)} \right| &\le \left| \frac{\varphi_n(u) - \varphi(u)}{\varphi(u)} \right| \left| \frac{\varphi_n''(u)}{\varphi_n(u)} \right| + \left| \frac{\varphi_n''(u) - \varphi''(u)}{\varphi(u)} \right| \\
&= \left| \frac{\varphi_n(u) - \varphi(u)}{\varphi(u)} \right| |\Psi_n''(u) + (\Psi_n'(u))^2| + \left| \frac{\varphi_n''(u) - \varphi''(u)}{\varphi(u)} \right|.
\end{aligned}
\tag{6.7}
$$

Note that the following estimates hold true under A_n^c:

$$
|\Psi_n'(u)| \le |\widehat{b}_n| + |u|\widehat{\nu}_{\sigma,n}(\mathbb{R}) \le |b| + 1 + |u|\big(\nu_\sigma(\mathbb{R}) + 1\big), \tag{6.8}
$$

$$
|\Psi_n''(u)| \le |\mathcal{F}\widehat{\nu}_{\sigma,n}(u)| \le \widehat{\nu}_{\sigma,n}(\mathbb{R}) \le \nu_\sigma(\mathbb{R}) + 1. \tag{6.9}
$$

Hence, we obtain from (6.5) to (6.9) and the trivial estimate $|\mathcal{F}\widehat{\nu}_{\sigma,n}(u)-\mathcal{F}\nu_\sigma(u)|\le\widehat{\nu}_{\sigma,n}(\mathbb{R})+\nu_\sigma(\mathbb{R})$ that under A_n^c, with some constant $C>0$,

$$\begin{aligned}
&\left|\int f\,\mathrm{d}\widehat{\nu}_{\sigma,n}-\int f\,\mathrm{d}\nu_\sigma\right|\\
&\quad\le C\int_{-\infty}^{\infty}|\mathcal{F}f(u)|\left(\frac{|\varphi_n(u)-\varphi(u)|+|\varphi_n'(u)-\varphi'(u)|+|\varphi_n''(u)-\varphi''(u)|}{|\varphi(u)|}(1+|u|)^2\wedge 1\right)\mathrm{d}u\\
&\quad\le C\int_{-\infty}^{\infty}(1+|u|)^s|\mathcal{F}f(u)|\,\mathrm{d}u\bullet\sup_{u\in\mathbb{R}}\left\{(1+|u|)^{-s}\left(\frac{(1+|u|)^2d^{(2)}(\varphi_n,\varphi)}{w(u)|\varphi(u)|}\wedge 1\right)\right\}\\
&\quad\le C\sup_{u\ge 0}\left\{(1+u)^{-s}\left(\frac{(1+u)^2n^{-1/2}}{w(u)|\varphi(u)|}\wedge 1\right)\right\}\left(n^{1/2}d^{(2)}(\varphi_n,\varphi)+1\right).
\end{aligned}$$

By monotonicity of $(1+u)^{-s}$ we can replace the supremum over $[0,\infty)$ by the supremum over $[0,U_n]$ and we arrive at

$$\begin{aligned}
&\mathbb{E}_{b,\nu_\sigma}\left[\mathbb{1}_{A_n^c}\bullet\sup_{f\in F_s}\left|\int f\,\mathrm{d}\widehat{\nu}_{\sigma,n}-\int f\,\mathrm{d}\nu_\sigma\right|\right]\\
&\quad=\mathrm{O}\left(\sup_{u\in[0,U_n]}\left\{(1+u)^{-s}\left(\frac{(1+u)^2n^{-1/2}}{w(u)|\varphi(u)|}\wedge 1\right)\right\}\right).
\end{aligned}\tag{6.10}$$

Together with the bound (6.4) on the set A_n this yields the asserted general estimate. Tracing back the constants, we see that they depend continuously on $|b|$ and $\nu_\sigma(\mathbb{R})$.

Proof of the rate results (a), (b).

- **(a)** Under the condition $\log|\varphi(u)|=-\sigma^2u^2/2(1+\mathrm{o}(u))$ we have $U_n\asymp\sqrt{\log n}$ and we obtain the rate $U_n^{-s}=(\log n)^{-s/2}$.
- **(b)** If $|\varphi(u)|\ge C\mathrm{e}^{-\alpha u}$, then we have $U_n\asymp\log n$ and we obtain the rate $U_n^{-s}=(\log n)^{-s}$.

Proof of the rate result (c). The same reasoning as for cases (a) and (b) would only yield the rate $((\log n)^{1/2+\delta}n^{1/2})^{-s/(\beta+2)}$ for $s\in(0,\beta+2]$ and the parametric rate for $s>\beta+2$. In the polynomial case (c), though, better estimates for $|\Psi_n'(u)|$ hold, that is, we can improve upon (6.8). First, we formulate and prove a lemma for $|\Psi'(u)|$.

Lemma 6.1. *If a Lévy process with a finite first moment has a characteristic function* (*at time* $t=1$) *satisfying* $|\varphi(u)|\ge C(1+|u|)^{-\beta}$ *for some* $\beta\ge 0$, $C>0$ *and all* $u\in\mathbb{R}$, *then* $\int_{[-1,+1]}|x|^\alpha\nu(\mathrm{d}x)$ *is finite for all* $\alpha>0$ *and the derivative of its characteristic exponent is uniformly bounded*:

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

Proof. Since we have necessarily $\sigma^2=0$ in the Lévy–Khinchine characteristic as well as $\int_{[-1,1]^c}|x|\nu(\mathrm{d}x)<\infty$ from the first moment condition, the additional property $\int_{[-1,+1]}|x|\times$

$\nu(\mathrm{d}x) < \infty$ implies

$$\sup_{u\in\mathbb{R}} |\Psi'(u)| = \sup_{u\in\mathbb{R}} \left| \mathrm{i}b + \int (\mathrm{e}^{\mathrm{i}ux} - 1)\mathrm{i}x\nu(\mathrm{d}x) \right| \le |b| + 2\int |x|\nu(\mathrm{d}x) < \infty.$$

It therefore remains to prove the first result for any $\alpha > 0$. We obtain with $c := \min_{u\in[1,2]}(1 - \cos(u)) > 0$:

$$\begin{aligned}
\int_{[-1,+1]} |x|^\alpha \nu(\mathrm{d}x) &\le \sum_{n=1}^{\infty} \int_{\{x:2^{-n}\le|x|\le 2^{-n+1}\}} |x|^\alpha \nu(\mathrm{d}x) \\
&\le \sum_{n=1}^{\infty} 2^{-\alpha(n-1)} \int_{\{x:2^{-n}\le|x|\le 2^{-n+1}\}} c^{-1}\big(1 - \cos(2^n x)\big)\nu(\mathrm{d}x) \\
&\le c^{-1} \sum_{n=1}^{\infty} 2^{-\alpha(n-1)} \operatorname{Re}(-\Psi(2^n)) \\
&\le c^{-1} \sum_{n=1}^{\infty} 2^{-\alpha(n-1)} \big(\log(C^{-1}) + \beta \log(1 + 2^n)\big).
\end{aligned}$$

This latter series is obviously finite. □

Resuming the proof for case (c), we remark that $|\varphi(u)| \ge C(1 + |u|)^{-\beta}$ implies for any $U > 0$

$$\begin{aligned}
&P_{b,\nu_\sigma}\left(\exists u \in [-U, U] : |\varphi_n(u)| < \frac{C}{2}(1 + |u|)^{-\beta}\right) \\
&\le P_{b,\nu_\sigma}\left(\sup_{|u|\le U} |\varphi_n(u) - \varphi(u)|(1 + |u|)^\beta \ge C/2\right) \\
&\le \frac{2}{C}\mathbb{E}[\|\varphi_n - \varphi\|_{L^\infty(w)}] w(U)^{-1}(1 + U)^\beta = \mathrm{O}\big(n^{-1/2} w(U)^{-1}(1 + U)^\beta\big).
\end{aligned}$$

Consequently, for $U_n \to \infty$ with $w(U_n)^{-1}U_n^\beta = \mathrm{o}(n^{1/2})$ we have

$$\lim_{n\to\infty} P_{b,\nu_\sigma}\left(\forall u \in [-U_n, U_n] : |\varphi_n(u)| \ge \frac{C}{2}(1 + |u|)^{-\beta}\right) = 1; \tag{6.11}$$

in the sequel we shall work with $U_n = n^{1/(2\beta)}(\log n)^{-(1/2+2\delta)/\beta}$. Theorem 4.1, Lemma 6.1 and equation (6.11) then yield

$$\begin{aligned}
\sup_{|u|\le U_n} |\Psi_n'(u) - \Psi'(u)| &\le \sup_{|u|\le U_n} \left\{ \frac{|\varphi_n'(u) - \varphi'(u)|}{|\varphi_n(u)|} + |\Psi'(u)| \frac{|\varphi(u) - \varphi_n(u)|}{|\varphi_n(u)|} \right\} \\
&= \mathrm{O}_P(n^{-1/2}) w(U_n)^{-1} \frac{2}{C}(1 + |U_n|)^\beta.
\end{aligned} \tag{6.12}$$

Together with estimate (6.12) and again Lemma 6.1 we have thus established for $n \to \infty$

$$\sup_{|u|\le U_n} |\Psi_n'(u)| = \mathrm{O}_P\big(1 + n^{-1/2} w(U_n)^{-1} |U_n|^{\beta}\big) \\ = \mathrm{O}_P(1). \tag{6.13}$$

We therefore get instead of (6.10) the estimate

$$\begin{aligned}
&\sup_{f\in F_s}\left|\int f\,\mathrm{d}\widehat{\nu}_{\sigma,n} - \int f\,\mathrm{d}\nu_\sigma\right| \\
&\quad= \sup_{u\in\mathbb{R}}\left\{(1+|u|)^{-s}\left(\frac{(\mathrm{O}_P(1) + u^2 \mathbb{1}_{\{|u|\ge U_n\}})n^{-1/2}}{w(u)|\varphi(u)|} \wedge 1\right)\right\}\mathrm{O}_P\big(n^{1/2} d^{(2)}(\varphi_n,\varphi) + 1\big) \\
&\quad= \sup_{u\in\mathbb{R}}\left\{(1+|u|)^{-s}\left(\frac{(\mathrm{O}_P(1) + u^2 \mathbb{1}_{\{|u|\ge U_n\}})n^{-1/2}}{(\log(\mathrm{e}+|u|))^{-1/2-\delta}(1+|u|)^{-\beta}} \wedge 1\right)\right\}\mathrm{O}_P(1).
\end{aligned}$$

For $s \le \beta$ the right-hand side is of order $\mathrm{O}_P(U_n^{-s})$ and we obtain

$$\sup_{f\in F_s}\left|\int f\,\mathrm{d}\widehat{\nu}_{\sigma,n} - \int f\,\mathrm{d}\nu_\sigma\right| = \mathrm{O}_P\big(n^{-s/2\beta}(\log n)^{s(1/2+2\delta)/\beta}\big),$$

while for $s > \beta$ the parametric rate $\mathrm{O}_P(n^{-1/2})$ follows.

6.3. Proof of Theorem 4.4

The lower bound will be established by looking at a decision problem between two local alternatives, see, for example, Korostelev and Tsybakov (1993) for the general idea. For $\gamma > 0$ and $\beta > 0$ consider the bilateral Gamma distribution which is obtained as the law of $X - Y$ where X and Y are independent and both $\Gamma(\gamma, \beta/2)$-distributed. This bilateral Gamma distribution is infinitely divisible with the following characteristic function and Lévy triplet:

$$\varphi_\Gamma(u) := (1+\gamma^{-2}u^2)^{-\beta/2}, \qquad b_\Gamma = 0, \sigma_\Gamma = 0, \nu_\Gamma(\mathrm{d}x) := \beta|x|^{-1}\mathrm{e}^{-\gamma|x|}\,\mathrm{d}x.$$

Its density f_Γ satisfies $f_\Gamma(x) \ge c\mathrm{e}^{-\gamma|x|}$ for some $c > 0$ (Küchler and Tappe (2008)). For $\sigma \ge 0$ consider the infinitely divisible distribution with characteristic function

$$\varphi_0(u) := \varphi_\Gamma(u)\mathrm{e}^{\mathrm{i}ub - \sigma^2u^2/2}, \tag{6.14}$$

which has a density f_0 that is a convolution of f_Γ with a normal density and therefore still satisfies $f_0(x) \ge c\mathrm{e}^{-\gamma|x|}$ with some $c > 0$. The corresponding Lévy density satisfies $\nu_0 = \nu_\Gamma$.

Let us further introduce for $K > 0$ and $\rho > 0$

$$\mu_K(x) := \mathrm{e}^{-x^2/(2\rho^2)}\sin(Kx).$$

For any $\beta > 0$ and $\gamma > 0$ we can choose ρ sufficiently small such that $\nu_0(x) + \mu_K(x) \geq 0$ holds for all $K > 0$. In this case the following characteristic function also generates an infinitely divisible distribution:

$$\varphi_K(u) := \varphi_0(u)\exp\biggl(\int_{\mathbb{R}} (\mathrm{e}^{\mathrm{i}ux} - 1)\mu_K(\mathrm{d}x)\biggr) = \varphi_0(u)\exp(\mathcal{F}\mu_K(u)).$$

Using the fact that $\sin(Kx)\sin(ux) = (\cos((K-u)x) - \cos((K+u)x))/2$ we obtain the following explicit calculation of the Fourier transform of μ_K:

$$\begin{aligned}\mathcal{F}\mu_K(u) &= \mathrm{i}\int_{-\infty}^{\infty} \mathrm{e}^{-x^2/(2\rho^2)}\sin(Kx)\sin(ux)\,\mathrm{d}x\\ &= \mathrm{i}\int_0^{\infty} \mathrm{e}^{-x^2/(2\rho^2)}\cos\bigl((K-u)x\bigr)\,\mathrm{d}x - \mathrm{i}\int_0^{\infty} \mathrm{e}^{-x^2/(2\rho^2)}\cos\bigl((K+u)x\bigr)\,\mathrm{d}x\\ &= \mathrm{i}\rho\sqrt{\pi/2}\bigl(\mathrm{e}^{-\rho^2(K-u)^2/2} - \mathrm{e}^{-\rho^2(K+u)^2/2}\bigr).\end{aligned}$$

Note that φ_K has the same decay behaviour as φ_0 due to $\lim_{|u|\to\infty}\mathcal{F}\mu_K(u) = 0$. Therefore $\nu_{0,\sigma}$ and $\nu_{K,\sigma}$ lie in the class $\mathcal{A}(C,\sigma)$ $(\sigma > 0)$ or $\mathcal{C}(C,\bar{C},\beta)$ $(\sigma = 0)$, respectively, provided $C, \bar{C}$ are large enough.

Let us now estimate the χ^2-distance between the distributions with characteristic functions φ_K and φ_0:

$$\begin{aligned}\chi^2(f_K, f_0) &:= \int_{-\infty}^{\infty} \frac{(f_K(x) - f_0(x))^2}{f_0(x)}\,\mathrm{d}x\\ &\leq c^{-1}\int_{-\infty}^{\infty} \bigl(\mathrm{e}^{\gamma|x|/2}f_K(x) - \mathrm{e}^{\gamma|x|/2}f_0(x)\bigr)^2\,\mathrm{d}x\\ &\leq c^{-1}\biggl\{\int_{-\infty}^{\infty} \bigl(\mathrm{e}^{\gamma x/2}f_K(x) - \mathrm{e}^{\gamma x/2}f_0(x)\bigr)^2\,\mathrm{d}x\\ &\qquad + \int_{-\infty}^{\infty} \bigl(\mathrm{e}^{-\gamma x/2}f_K(x) - \mathrm{e}^{-\gamma x/2}f_0(x)\bigr)^2\,\mathrm{d}x\biggr\}.\end{aligned} \tag{6.15}$$

For functions g whose Fourier transform can be extended holomorphically to complex values z with $|\operatorname{Im}(z)| < \gamma$ we have:

$$\mathcal{F}(\mathrm{e}^{\pm\gamma x/2}g(x))(u) = \int g(x)\mathrm{e}^{(\mathrm{i}u\pm\gamma/2)x}\,\mathrm{d}x = \mathcal{F}g\bigl(u \pm (-\mathrm{i})\gamma/2\bigr).$$

Using this identity in Plancherel's formula and then the estimate $|\mathrm{e}^z - 1| \leq |z|\mathrm{e}^{|\operatorname{Re}(z)|}$, $z \in \mathbb{C}$, together with $|\mathcal{F}\mu_K(u)| \leq \|\mu_K\|_{L^1}$, we continue from (6.15):

$$\begin{aligned}&\chi^2(f_K, f_0)\\ &\quad\leq \frac{c^{-1}}{2\pi}\int_{-\infty}^{\infty} \bigl(|\varphi_K(u - \mathrm{i}\gamma/2) - \varphi_0(u - \mathrm{i}\gamma/2)|^2 + |\varphi_K(u + \mathrm{i}\gamma/2) - \varphi_0(u + \mathrm{i}\gamma/2)|^2\bigr)\,\mathrm{d}u\end{aligned}$$

$$= \frac{c^{-1}}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{-\sigma^2 u^2} \left| \frac{3}{4} + \frac{u^2}{\gamma^2} + \frac{\mathrm{i}u}{\gamma} \right|^{-\beta} \left(\left| \mathrm{e}^{\mathcal{F}\mu_K(u-\mathrm{i}\gamma/2)} - 1 \right|^2 + \left| \mathrm{e}^{\mathcal{F}\mu_K(u+\mathrm{i}\gamma/2)} - 1 \right|^2 \right) \mathrm{d}u$$

$$\le \frac{\mathrm{e}^{2\|\mu_K\|_{L^1}}}{2c\pi} \int_{-\infty}^{\infty} \mathrm{e}^{-\sigma^2 u^2} \left(\frac{3}{4} + \frac{u^2}{\gamma^2} \right)^{-\beta} \left(|\mathcal{F}\mu_K(u-\mathrm{i}\gamma/2)|^2 + |\mathcal{F}\mu_K(u+\mathrm{i}\gamma/2)|^2 \right) \mathrm{d}u$$

$$= \frac{\mathrm{e}^{2\|\mu_K\|_{L^1}} \rho^2}{4c\pi} \int_{-\infty}^{\infty} \mathrm{e}^{-\sigma^2 u^2} \left(\frac{3}{4} + \frac{u^2}{\gamma^2} \right)^{-\beta} \left(\mathrm{e}^{-\rho^2(u-K)^2/2} - \mathrm{e}^{-\rho^2(u+K)^2/2} \right)^2 \mathrm{d}u.$$

The last line is for $K \to \infty$ of order $\int_{-\infty}^{\infty} \mathrm{e}^{-\sigma^2 u^2}(1+u^2)^{-\beta}(\mathrm{e}^{-\rho^2(u-K)^2} + \mathrm{e}^{-\rho^2(u+K)^2})\,\mathrm{d}u$. In the case $\sigma = 0$ (polynomial decay) this gives the order $K^{-2\beta}$, whereas for $\sigma > 0$ (Gaussian part) the order is $\mathrm{e}^{-\sigma^2 K^2(1+\mathrm{o}(1))}$.

For n observations the distributions do not separate provided $K^{-2\beta} \asymp n^{-1}$ ($\sigma = 0$) and $\mathrm{e}^{-\sigma^2 K^2(1+\mathrm{o}(1))} \asymp n^{-1}$ ($\sigma > 0$), respectively. Consequently, when choosing $K_n \asymp n^{1/2\beta}$ ($\sigma = 0$), respectively $K_n = c\sqrt{\log(n)}$ with $c > 0$ sufficiently large ($\sigma > 0$), this closeness of the distributions implies (Korostelev and Tsybakov (1993)) that for any sequence of estimators $(\widehat{\nu}_{\sigma,n})_n$ we have

$$\liminf_{n\to\infty} \left\{ P_0\left(\ell_s(\widehat{\nu}_{\sigma,n}, \nu_{0,\sigma}) \ge \ell_s(\nu_{K_n,\sigma}, \nu_{0,n})/2\right) + P_{K_n}\left(\ell_s(\widehat{\nu}_{\sigma,n}, \nu_{0,\sigma}) \ge \ell_s(\nu_{K_n,\sigma}, \nu_{0,n})/2\right) \right\}$$
$$> 0.$$

It remains to consider the loss ℓ_s between the alternatives. Using the formula $\mathcal{F}(x^2 \times \mathrm{e}^{-x^2/(2\rho^2)})(u) = \rho^3(1-\rho^2u^2)\mathrm{e}^{-\rho^2u^2/2}$, we calculate:

$$\begin{aligned}
\ell_s(\nu_{K,\sigma}, \nu_{0,\sigma}) &= \sup_{f\in F_s} \left| \int_{-\infty}^{\infty} f(x) x^2 \mathrm{e}^{-x^2/2\rho^2} \sin(Kx)\,\mathrm{d}x \right| \\
&= \frac{1}{2\pi} \sup_{f\in F_s} \left| \mathrm{Im}\left(\left(\mathcal{F}f * \mathcal{F}(x^2 \mathrm{e}^{-x^2/2\rho^2}) \right)(K) \right) \right| \\
&= \frac{1}{2\pi} \sup_{f\in F_s} \left| \int_{-\infty}^{\infty} \mathrm{Im}(\mathcal{F}f(x)) \rho^3 \left(1 - \rho^2(K-u)^2\right) \mathrm{e}^{-\rho^2(K-u)^2/2}\,\mathrm{d}u \right| \\
&= \frac{1}{2\pi} \rho^3 \sup_{u\in\mathbb{R}} \left\{ (1+|u|)^{-s} |1 - \rho^2(K-u)^2| \mathrm{e}^{-\rho^2(K-u)^2/2} \right\} \\
&\asymp K^{-s}.
\end{aligned}$$

Setting $\varepsilon := \liminf_{n\to\infty} K_n^s \ell_s(\nu_{K,\sigma}, \nu_{0,\sigma})/2 > 0$, we have thus shown

$$\liminf_{n\to\infty} \sup_{\nu_\sigma} P_{b,\nu_\sigma}\left(K_n^s \ell_s(\widehat{\nu}_{\sigma,n}, \nu_\sigma) \ge \varepsilon \right) > 0.$$

For $\sigma = 0$ (polynomial decay) this gives the desired lower bound $K_n^{-s} = n^{-s/(2\beta)}$ for any $\beta > 0$ and for $s \le \beta$. For $s > \beta$ a standard parametric argument shows that the minimax rate is never faster than $n^{-1/2}$. For $\sigma > 0$ (Gaussian part) we obtain the lower bound $K_n^{-s} = (\log n)^{-s/2}$, which matches exactly the upper bound.

In the case (b), that is, where $|\varphi(u)| \geq C\mathrm{e}^{-\alpha|u|}$, we consider instead of (6.14)

$$\varphi_0(u) = \varphi_\Gamma(u)\varphi_\alpha(u),$$

where φ_α is an infinitely divisible characteristic function with $|\varphi(u)| \asymp \mathrm{e}^{-\alpha|u|}$ such that the corresponding density function f_α has faster exponential decay than f_0. For example, a tempered stable law (Cont and Tankov (2004), Proposition 4.2) with $\nu(\mathrm{d}x) = \alpha|x|^{-2}\mathrm{e}^{-|\lambda|x}\,\mathrm{d}x$ and $\lambda > 0$ sufficiently large meets these requirements. The remaining steps of the proof are exactly the same, just replace $\mathrm{e}^{-\sigma^2u^2/2}$ by $\mathrm{e}^{-\alpha|u|}$.

6.4. Proof of Proposition 5.1

Note first that $\mathbb{E}_{b,\nu_\sigma}|\widetilde{b}_n - b|^2 = \mathrm{O}(n^{-1})$ follows directly from $\mathbb{E}X_1^2 < \infty$.

To prove the result for the jump measure, we distinguish between two cases. We set $\widehat{\Psi}'_n(u) = \widehat{\varphi}'_n(u)/\widehat{\varphi}_n(u)$ and $\widehat{\Psi}''_n(u) = \widehat{\varphi}''_n(u)/\widehat{\varphi}_n(u) - (\widehat{\varphi}'_n(u)/\widehat{\varphi}_n(u))^2$.

Case 1: $|\varphi(u)| \geq 2\kappa n^{-1/2}$

It follows from (6.6) and (6.7) that

$$\begin{aligned}
&|\mathcal{F}\widetilde{\nu}_{\sigma,n}(u) - \mathcal{F}\nu_\sigma(u)| \\
&\quad\leq \left\{\left|\frac{\widehat{\varphi}_n(u) - \varphi(u)}{\varphi(u)}\right| |\widehat{\Psi}'_n(u)| + \left|\frac{\widehat{\varphi}'_n(u) - \varphi'(u)}{\varphi(u)}\right| |\widehat{\Psi}'_n(u) + \Psi'(u)|\right\} \mathbb{1}_{\{|\hat{\varphi}_n(u)| \geq \kappa n^{-1/2}\}} \\
&\qquad + \left\{\left|\frac{\widehat{\varphi}_n(u) - \varphi(u)}{\varphi(u)}\right| |\widehat{\Psi}''_n(u) + (\widehat{\Psi}'_n(u))^2| + \left|\frac{\widehat{\varphi}''_n(u) - \varphi''(u)}{\varphi(u)}\right|\right\} \mathbb{1}_{\{|\hat{\varphi}_n(u)| \geq \kappa n^{-1/2}\}} \qquad (6.16) \\
&\qquad + |\mathcal{F}\nu_\sigma(u)| \mathbb{1}_{\{|\hat{\varphi}_n(u)| < \kappa n^{-1/2}\}} \\
&\quad = T_{n,1} + T_{n,2} + T_{n,3},
\end{aligned}$$

say.

We obtain from the inequality

$$\frac{\mathbb{1}_{\{|\hat{\varphi}_n(u)| \geq \kappa n^{-1/2}\}}}{\widehat{\varphi}_n(u)} \leq \frac{1}{|\varphi(u)|} + \frac{|\widehat{\varphi}_n(u) - \varphi(u)|}{\kappa n^{-1/2}|\varphi(u)|}$$

that

$$\mathbb{E}\big[|\widehat{\varphi}_n(u)|^{-p}\mathbb{1}_{\{|\hat{\varphi}_n(u)| \geq \kappa n^{-1/2}\}}\big] = \mathrm{O}(|\varphi(u)|^{-p}) \qquad (6.17)$$

holds for all $p \in \mathbb{N}$. This implies, by $\widehat{\Psi}'_n(u) = (\widehat{\varphi}'_n(u) - \varphi'(u))/\widehat{\varphi}_n(u) + \Psi'(u)\varphi(u)/\widehat{\varphi}_n(u)$, that

$$\mathbb{E}\big[|\widehat{\Psi}'_n(u)|^p\mathbb{1}_{\{|\hat{\varphi}_n(u)| \geq \kappa n^{-1/2}\}}\big] = \mathrm{O}\big(\big(1 + |\Psi'(u)|\big)^p\big). \qquad (6.18)$$

Therefore, we obtain that

$$\mathbb{E}T_{n,1} = \mathrm{O}\left(\frac{n^{-1/2}}{|\varphi(u)|}\big(1+|\Psi'(u)|\big)^2\right). \tag{6.19}$$

Since

$$\begin{aligned}\widehat{\Psi}_n''(u) &= \frac{\widehat{\varphi}_n''(u)}{\widehat{\varphi}_n(u)} - (\widehat{\Psi}_n'(u))^2\\ &= \frac{\widehat{\varphi}_n''(u)-\varphi''(u)}{\widehat{\varphi}_n(u)} + \big(\Psi''(u)+(\Psi'(u))^2\big)\frac{\varphi(u)}{\widehat{\varphi}_n(u)} - (\widehat{\Psi}_n'(u))^2,\end{aligned}$$

we obtain, in conjunction with (6.17) and (6.18), that

$$\mathbb{E}\big[|\widehat{\Psi}_n''(u)|^2 \mathbb{1}_{\{|\hat{\varphi}_n(u)|\ge \kappa n^{-1/2}\}}\big] = \mathrm{O}\big(\big(1+|\Psi'(u)|\big)^2\big).$$

We conclude that

$$\mathbb{E}T_{n,2} = \mathrm{O}\left(\frac{n^{-1/2}}{|\varphi(u)|}\big(1+|\Psi'(u)|\big)^2\right). \tag{6.20}$$

Finally, it follows from Hoeffding's inequality for bounded random variables that

$$\begin{aligned}P\big(|\widehat{\varphi}_n(u)| < \kappa n^{-1/2}\big) &\le P\big(|\widehat{\varphi}_n(u)-\varphi(u)| > |\varphi(u)| - \kappa n^{-1/2}\big)\\ &\le P\big(|\widehat{\varphi}_n(u)-\varphi(u)| > |\varphi(u)|/2\big)\\ &\le \exp(-cn|\varphi(u)|^2),\end{aligned}$$

for some $c>0$. This yields that $P(|\widehat{\varphi}_n(u)| < \kappa n^{-1/2}) = \mathrm{O}(n^{-1/2}|\varphi(u)|^{-1})$, and therefore

$$\mathbb{E}T_{n,3} = \mathrm{O}\left(\frac{n^{-1/2}}{|\varphi(u)|}\right). \tag{6.21}$$

Equations (6.16), (6.19), (6.20) and (6.21) yield the desired bound in the case $|\varphi(u)| \ge 2\kappa n^{-1/2}$.

Case 2: $|\varphi(u)| < 2\kappa n^{-1/2}$

In contrast to Case 1, this time we use the following decomposition:

$$\begin{aligned}&|\mathcal{F}\widetilde{\nu}_{\sigma,n}(u) - \mathcal{F}\nu_\sigma(u)|\\ &\quad\le \left\{\left|\frac{\widehat{\varphi}_n(u)-\varphi(u)}{\widehat{\varphi}_n(u)}\right||\Psi'(u)| + \left|\frac{\widehat{\varphi}_n'(u)-\varphi'(u)}{\widehat{\varphi}_n(u)}\right||\Psi'(u)+\widehat{\Psi}_n'(u)|\right\}\mathbb{1}_{\{|\hat{\varphi}_n(u)|\ge\kappa n^{-1/2}\}}\\ &\qquad + \left\{\left|\frac{\widehat{\varphi}_n(u)-\varphi(u)}{\widehat{\varphi}_n(u)}\right||\Psi''(u)+(\Psi'(u))^2| + \left|\frac{\widehat{\varphi}_n''(u)-\varphi''(u)}{\widehat{\varphi}_n(u)}\right|\right\}\mathbb{1}_{\{|\hat{\varphi}_n(u)|\ge\kappa n^{-1/2}\}}\\ &\qquad + |\mathcal{F}\nu_\sigma(u)|\mathbb{1}_{\{|\hat{\varphi}_n(u)|<\kappa n^{-1/2}\}}.\end{aligned} \tag{6.22}$$

Taking into account that Ψ'' is bounded and using again (6.18) as well as the trivial estimate $|\mathcal{F}\nu_\sigma(u)| \le \nu_\sigma(\mathbb{R}) < \infty$ we obtain that

$$\mathbb{E}|\mathcal{F}\widetilde{\nu}_{\sigma,n}(u) - \mathcal{F}\nu_\sigma(u)| = \mathrm{O}\big(\big(1+|\Psi'(u)|\big)^2\big),$$

as required.

Acknowledgements

We thank Peter Tankov for the idea of how to prove Lemma 6.1 and Shota Gugushvili for useful discussions and hints.

References

Basawa, I.V. and Brockwell, P.J. (1982). Non-parametric estimation for non-decreasing Lévy processes. *J. Roy. Statist. Soc. Ser. B* **44** 262–269. MR0676217

Belomestny, D. and Reiß, M. (2006). Spectral calibration of exponential Lévy models. *Finance Stoch.* **10** 449–474. MR2276314

Cont, R. and Tankov, P. (2004). *Financial Modelling with Jump Processes*. Boca Raton: Chapman and Hall. MR2042661

Chung, K.L. (1974). *A Course in Probability Theory*, 2nd ed. San Diego: Academic Press. MR1796326

Csörgő, S. and Totik, V. (1983). On how long interval is the empirical characteristic function uniformly consistent. *Acta Sci. Math. (Szeged)* **45** 141–149. MR0708779

Dudley, R.M. (1989). *Real Analysis and Probability*. Belmont: Wadsworth. MR0982264

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272. MR1126324

Feuerverger, A. and McDunnough, P. (1981a). On the efficiency of empirical characteristic function procedures. *J. Roy. Statist. Soc. Ser. B* **43** 20–27. MR0610372

Feuerverger, A. and McDunnough, P. (1981b). On some Fourier methods for inference. *J. Amer. Statist. Assoc.* **76** 379–387. MR0624339

Figueroa-López, J. and Houdré, C. (2006). Risk bounds for the nonparametric estimation of Lévy processes. In *High Dimensional Probability*. *IMS Lecture Notes* **51** 96–116. Beachwood, OH: IMS. MR2387763

Gnedenko, B.V. and Kolmogorov, A.N. (1968). *Limit Distributions for Sums of Independent Random Variables*, 2nd ed. Reading, MA: Addison-Wesley. MR0233400

Goldenshluger, A. and Pereverzev, S.V. (2003). On adaptive inverse estimation of linear functionals in Hilbert scales. *Bernoulli* **9** 783–807. MR2047686

Gugushvili, S. (2007). Decompounding under Gaussian noise. Available at arXiv:0711.0719v1.

Hall, P. and Yao, Q. (2003). Inference in components of variance models with low replication. *Ann. Statist.* **31** 414–441. MR1983536

Jacod, J. and Shiryaev, A. (2002). *Limit Theorems for Stochastic Processes*, 2nd ed. *Grundlehren* **288**. Berlin: Springer. MR0959133

Jongbloed, G., van der Meulen, F.H. and van der Vaart, A.W. (2005). Nonparametric inference for Lévy-driven Ornstein–Uhlenbeck processes. *Bernoulli* **11** 759–791. MR2172840

Katznelson, Y. (1976). *An Introduction to Harmonic Analysis*, 2nd ed. New York: Dover. MR0422992

Kolmogorov, A.N. (1932). Sulla formula generale di un processo stochastico omogeneo (Un problema di Bruno de Finetti). *Rendiconti della R. Accademia Nazionale dei Lincei (Ser. VI)* **15** 866–869. (In Italian.)

Korostelev, A.P. and Tsybakov, A.B. (1993). *Minimax Theory of Image Reconstruction. Lecture Notes in Statistics* **82**. New York: Springer. MR1226450
Küchler, U. and Tappe, S. (2008). On the shapes of bilateral Gamma densities. *Statist. Probab. Lett.* **78** 2478–2484.
Mainardi, F. and Rogosin, S. (2006). The origin of infinitely-divisible distributions: from de Finetti's problem to Lévy–Khinchine formula. *Math. Methods Econ. Finance* **1** 37–55. MR2317505
Nishiyama, Y. (2008). Nonparametric estimation and testing time-homogeneity for processes with independent increments. *Stochastic. Process. Appl.* **118** 1043–1055. MR2418257
van Es, B., Gugushvili, S. and Spreij P. (2007). A kernel-type nonparametric density estimator for decompounding. *Bernoulli* **13** 672–694. MR2348746
van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge: Cambridge Univ. Press. MR1652247
Watteel, R.N. and Kulperger, R.J. (2003). Nonparametric estimation of the canonical measure for infinitely divisible distributions. *J. Statist. Comput. Simul.* **73** 525–542. MR1986343
Yukich, J.E. (1985). Weak convergence of the empirical characteristic function. *Proc. Amer. Math. Soc.* **95** 470–473. MR0806089

Received July 2007 and revised March 2008