Nonparametric inference for discretely sampled Lévy processes

(Submitted on 21 Aug 2009 (v1), last revised 25 May 2011 (this version, v3))

Shot Gugushvili

Nonparametric inference for discretely sampled Lévy processes

Shota Gugushvili
Department of Mathematics
Vrije Universiteit Amsterdam
De Boelelaan 1081a
1081 HV Amsterdam
The Netherlands
s.gugushvili@vu.nl

May 26, 2011

Abstract

Given a sample from a discretely observed Lévy process $X = (X_t)_{t \geq 0}$ of the finite jump activity, the problem of nonparametric estimation of the Lévy density ρ corresponding to the process X is studied. An estimator of ρ is proposed that is based on a suitable inversion of the Lévy-Khintchine formula and a plug-in device. The main results of the paper deal with upper risk bounds for estimation of ρ over suitable classes of Lévy triplets. The corresponding lower bounds are also discussed.

Keywords: Empirical characteristic function; empirical process; Fourier inversion; Lévy density; Lévy process; maximal inequality; mean square error.
AMS subject classification: 62G07, 62G20

1 Introduction

Recent years have witnessed a great revival of interest in Lévy processes, which is primarily due to the fact that they have found numerous applications in various fields. The main interest has been in mathematical finance, see e.g. [28] for a detailed treatment and many references, however Lévy processes obtained due attention also in queueing, telecommunications, extreme value theory, quantum theory and many others. A thorough exposition of the fundamental properties of Lévy processes can be found e.g. in [8], [44] and [52].

It is well-known that Lévy processes have a close link with infinitely divisible distributions: if $X = (X_t)_{t \geq 0}$ is a Lévy process, then its marginal distributions are all infinitely divisible and are determined by the distribution of X_Δ, where $\Delta > 0$ is an arbitrary fixed number. Conversely, given an infinitely divisible distribution μ, one can construct a Lévy process $X = (X_t)_{t \geq 0}$, such that $P_{X_\Delta} = \mu$, cf. Theorem 7.10 in [52]. Hence the law of the process X can be uniquely characterised by the characteristic function of X_Δ, where $\Delta > 0$ is some fixed number. By the Lévy-Khintchine formula for infinitely divisible distributions, the characteristic function ϕ_{X_Δ} of X_Δ can be written as

$$\phi_{X_\Delta}(t) = e^{\psi_\Delta(t)},$$

where the exponent ψ_Δ, called the characteristic or Lévy exponent, is given by

$$\psi_\Delta(t) = \Delta i \gamma_0 t - \Delta \frac{1}{2} \sigma^2 t^2 + \Delta \int_{\mathbb{R} \setminus \{0\}} (e^{itx} - 1 - itx 1_{[|x| \leq 1]}) \nu(dx), \qquad (1)$$

see Theorem 8.1 of [52]. Here $\gamma_0 \in \mathbb{R}$, $\sigma \geq 0$, and ν is a measure concentrated on $\mathbb{R} \setminus \{0\}$, such that $\int_{\mathbb{R} \setminus \{0\}} (1 \wedge x^2) \nu(dx) < \infty$. This measure is called the Lévy measure, while the triple $(\gamma_0, \sigma^2, \nu)$ is referred to as the characteristic or Lévy triplet of X. The parameter γ_0 is called a drift parameter and a constant σ^2 is a diffusion parameter. The representation in (1) in terms of the Lévy triplet is unique. It then follows that the Lévy triplet determines uniquely the law of any Lévy process. Therefore, many statistical inference problems for Lévy processes can be reduced to inference on the corresponding characteristic triplets.

Until quite recently most of the existing literature dealt with parametric inference procedures for Lévy processes, see e.g. [2]–[5], [9]–[11], [20], [41], [49], [51] and [59]. However, a nonparametric approach is also possible and arises if one does not impose parametric assumptions on the Lévy measure, or its density, in case the latter exists. A nonparametric approach can give e.g. valuable indications about the shape of the Lévy density. Furthermore, parametric inference for Lévy processes is complicated by the fact that for many Lévy processes their marginal densities are often intractable or not

available in closed form. This makes the implementation of such a standard parameter estimation method as the maximum likelihood method difficult. We refer e.g. to [1], [13]–[15], [21], [23]–[26], [29], [34], [38], [42]–[43], [48], [58], as well as the proceedings [40] and references therein for a nonparametric approach to inference for Lévy processes.

In the present work we will assume that the Lévy measure ν has a finite total mass, i.e. $\nu(\mathbb{R}) < \infty$, and that it has a density ρ. In essence this means that the Lévy process that we sample from is a sum of a linear drift, a rescaled Brownian motion and a compound Poisson process. Thus this model is related to Merton's model of an asset price, see [46]. Nonparametric inference for a similar model was considered in [6], [21] and [38].

Since in our case $\nu(\mathbb{R}) < \infty$, the Lévy-Khintchine exponent can be rewritten as

$$\psi_\Delta(t) = \Delta i\gamma t - \Delta\frac{1}{2}\sigma^2 t^2 + \Delta\int_{-\infty}^{\infty}(e^{itx} - 1)\rho(x)dx. \qquad (2)$$

The triple (γ, σ^2, ρ) is again referred to as a Lévy triplet. Note that γ in (2) differs from γ_0 in (1).

Suppose that the Lévy process $X = (X_t)_{t\geq 0}$ is observed at discrete time instances $\Delta, 2\Delta, \ldots, n\Delta$, with Δ kept fixed. This sampling case is usually referred to as the low frequency data case. For the case when Δ is allowed to depend on n and $\Delta \to 0, n\Delta \to \infty$ as $n \to \infty$ see e.g. [25], [26] or [37]. In this case it is customary to talk about high frequency data case. Returning to the case with a fixed Δ, by a rescaling argument, without loss of generality, we can take $\Delta = 1$. Based on observations $X_1, \ldots, X_n$, our goal in this paper is to estimate nonparametrically the Lévy density ρ. Notice that this is an inverse problem in that ρ is associated with jump sizes of a Lévy process and their intensity, the jumps themselves are not directly observable under the present sampling scheme, and consequently ρ has to be estimated from indirect observations $X_1, \ldots, X_n$.

We will base our estimator of ρ on a suitable inversion of ϕ_{X_1}. The idea of expressing the Lévy measure or the Lévy density in terms of ϕ_{X_1} and then replacing ϕ_{X_1} by its natural nonparametric estimator, the empirical characteristic function, to obtain a plug-in type estimator for the Lévy measure or the Lévy density has been successfully applied e.g. in [21], [24], [34], [38], [48] and [58]. The logic behind this approach is that except of some particular cases, e.g. that of the compound Poisson process, see [14] and [15], finding an explicit relationship expressing the Lévy measure or its density directly in terms of the distribution of X_1 without referring to the Fourier transforms is difficult. This hampers the use of a plug-in device, which is one of the most popular and useful methods for obtaining estimators in statistics. On the other hand the Fourier approach allows one to cover a large class of examples, as shown in the above-mentioned papers.

Observe that the model we consider in the present work shares many features characteristic of a convolution model with partially or totally unknown

error distribution, see [17], [27], [45] and [47]. For instance, the Gaussian components in $X_1, \ldots, X_n$ in our case will play a role similar to the measurement error in those papers, in case the latter has a normal distribution.

We proceed to the construction of an estimator of ρ. First by differentiating the Lévy-Khintchine formula we will derive a suitable inversion formula for ρ. Suppose that $\int_{\mathbb{R}} x^2 \rho(x) dx < \infty$. Since ρ has a finite second moment, so does X_1 by Corollary 25.8 in [52]. Also $\mathrm{E}\left[|X_1|\right]$ is finite by the Cauchy-Schwarz inequality. Hence we can differentiate ϕ_{X_1} with respect to t to obtain

$$\phi'_{X_1}(t) = \phi_{X_1}(t) \left(i\gamma - \sigma^2 t + i \int_{-\infty}^{\infty} e^{itx} x \rho(x) dx \right). \tag{3}$$

Notice that differentiation of $\int_{\mathbb{R}} (e^{itx} - 1) \rho(x) dx$ under the integral sign is justified by the dominated convergence theorem, applicable because of our assumptions on ρ. Next rewrite (3) as

$$\frac{\phi'_{X_1}(t)}{\phi_{X_1}(t)} = i\gamma - \sigma^2 t + i \int_{\mathbb{R}} e^{itx} x \rho(x) dx, \tag{4}$$

which is possible, because $\phi_{X_1}(t) \neq 0$ for all $t \in \mathbb{R}$, see e.g. Theorem 7.6.1 in [22]. Differentiating both sides of this identity with respect to t, we get

$$\frac{\phi''_{X_1}(t)\phi_{X_1}(t) - (\phi'_{X_1}(t))^2}{(\phi_{X_1}(t))^2} = -\sigma^2 - \int_{\infty}^{\infty} e^{itx} x^2 \rho(x) dx, \tag{5}$$

where again we interchanged the differentiation and integration order in the righthand side of (4) to obtain the righthand side of (5). Thus by rearranging the terms we have

$$\int_{-\infty}^{\infty} e^{itx} x^2 \rho(x) dx = \frac{(\phi'_{X_1}(t))^2 - \phi''_{X_1}(t)\phi_{X_1}(t)}{(\phi_{X_1}(t))^2} - \sigma^2. \tag{6}$$

Suppose that the righthand side is integrable, which is implied by the assumption that ϕ''_ρ is integrable. Here ϕ_ρ denotes the Fourier transform of ρ. Then by the Fourier inversion argument the relationship

$$x^2 \rho(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \left(\frac{(\phi'_{X_1}(t))^2 - \phi''_{X_1}(t)\phi_{X_1}(t)}{(\phi_{X_1}(t))^2} - \sigma^2 \right) dt$$

holds. If $x \neq 0$, this yields

$$\rho(x) = \frac{1}{2\pi x^2} \int_{-\infty}^{\infty} e^{-itx} \left(\frac{(\phi'_{X_1}(t))^2 - \phi''_{X_1}(t)\phi_{X_1}(t)}{(\phi_{X_1}(t))^2} - \sigma^2 \right) dt, \tag{7}$$

4

and we obtain a desired inversion formula. This formula coincides with the one given in [16][1]. The formula has to be compared to related inversion formulae given in [24], [26], [48] and [58]. Notice that under stronger moment conditions on X_1 one can perform the differentiation step in the above derivation not twice, but three times, thereby eliminating σ^2 from (5), and one can obtain an inversion formula of the same type as in (7), but not involving σ^2 explicitly, see e.g. [26]. We do not pursue this path, as a study of asymptotic properties of an estimator of ρ of the same type as we propose below based on this different inversion formula would require stronger moment conditions on X_1, cf. the discussion in the next section. It would also involve longer and more technical proofs of the asymptotic results. Finally, under certain smoothness assumptions on the Lévy density it would lead to an estimator with worse convergence rate than the one that we propose below. See Section 2 for an additional discussion.

Denote $Z_j = X_j - X_{j-1}$ and observe that $Z_1, \ldots, Z_n$ are i.i.d., which follows from the stationary independent increments property of a Lévy process. Let $\hat{\phi}(t) = n^{-1} \sum_{j=1}^{n} e^{itZ_j}$. By the strong law of large numbers, for every fixed t, the empirical characteristic function $\hat{\phi}(t)$ and its derivatives with respect to t, $\hat{\phi}'(t)$ and $\hat{\phi}''(t)$, converge a.s. to $\phi_{X_1}(t), \phi'_{X_1}(t)$ and $\phi''_{X_1}(t)$, respectively. Using a plug-in device, a possible estimator of $\rho(x)$ could then be

$$\frac{1}{2\pi x^2} \int_{-\infty}^{\infty} e^{-itx} \left(\frac{(\hat{\phi}'(t))^2}{(\hat{\phi}(t))^2} - \frac{\hat{\phi}''(t)}{\hat{\phi}(t)} - \hat{\sigma}^2 \right) dt, \tag{8}$$

where $\hat{\sigma}^2$ is some estimator of σ^2. The problem with this 'estimator' of ρ is that in general the integrand in (8) is not integrable. Furthermore, small values of $\hat{\phi}(t)$ might render the estimator numerically unstable, since $\hat{\phi}(t)$ appears in the denominator in (8). Therefore, as an estimator of ρ we propose the following modification of (8),

$$\hat{\rho}(x) = \frac{1}{2\pi x^2} \int_{-\infty}^{\infty} e^{-itx} \left(\frac{(\hat{\phi}'(t))^2}{(\hat{\phi}(t))^2} 1_{G_t} - \frac{\hat{\phi}''(t)}{\hat{\phi}(t)} 1_{G_t} - \hat{\sigma}^2 \right) \phi_w(ht) dt. \tag{9}$$

Here ϕ_w denotes the Fourier transform of a kernel function w, while a number $h > 0$ denotes a bandwidth. This terminology is borrowed from the kernel estimation theory, see e.g. [54]. The integral in (9) is finite under the assumption that ϕ_w has a compact support, for instance on $[-1, 1]$. We define the set G_t in (9) by

$$G_t = \left\{ |\hat{\phi}(t)| \geq \kappa_n e^{-\Sigma^2/(2h^2)} \right\}. \tag{10}$$

---

[1][16] contains a more general result valid also for Lévy densities with infinite total mass. However, the statement of the theorem in [16] mistakenly claims that the Lévy density ρ is bounded under the assumptions given in [16]. In reality this can in general be ascertained only for $x^2 \rho(x)$. Examples (e) and (f) considered in [16] illustrate our point.

Hence G_t depends on h, as well as a constant Σ and a sequence $\kappa_n \to 0$ of real numbers to be specified in the next section, where we also give some additional heuristics for the definition of G_t. A general reason for using truncation with 1_{G_t} is a desire of numerical stability, but truncation in (9) will also help in proving the asymptotic results from Section 2. At this point notice that we could have also used a "diagonal-out" estimator

$$\frac{2}{n(n-1)} \sum_{1 \leq j < k \leq n} e^{itZ_j} e^{itZ_k}$$

to estimate $(\phi_{X_1}(t))^2$ in the denominator of (7) and a similar "diagonal-out" estimator to estimate $(\phi'_{X_1}(t))^2$. An advantage of these two estimators is that they are unbiased estimators of $(\phi_{X_1}(t))^2$ and $(\phi'_{X_1}(t))^2$, respectively, while $(\hat{\phi}(t))^2$ and $(\hat{\phi}'(t))^2$ are not. On the theoretical side study of the resulting modification of $\hat{\rho}$ would require the use of the theory of U-statistics, see e.g. Chapter 12 in [55]. However, since in the present paper we are mainly concerned with rates of convergence for estimation of ρ, we refrain from studying this possible modification of $\hat{\rho}$.

It remains to propose an estimator of σ^2. To this end we use an estimator from [38] defined via

$$\hat{\sigma}^2 = \int_{\mathbb{R}} \max\{\min\{M_n, \log(|\hat{\phi}(t)|)\}, -M_n\} v_h(t) dt. \tag{11}$$

Here v_h is a kernel function depending on h, while M_n denotes a sequence of positive numbers diverging to infinity at a suitable rate. Appropriate conditions on all three will be given in the next section. The estimator $\hat{\sigma}^2$ is again based on the Lévy-Khintchine formula and we refer to [38] for the heuristics of its introduction. There does not seem to exist an 'easy' way to define an estimator of σ^2. 'Nonparametric' estimators of finite-dimensional parameters in semiparametric deconvolution problems (these are related to the problem we are considering in the present paper) have already been proposed in the literature, see e.g. [17] and [39]. In the context of Lévy processes 'nonparametric' estimators of finite-dimensional parameters have been used e.g. in [6] and [38]. These estimators can often be proven to be rate-optimal.

If ϕ_w is symmetric and real-valued, then by taking a complex conjugate one can see that $\hat{\rho}$ is real-valued, because this amounts to changing the integration variable from t into $-t$ in (9). On the other hand, positivity of $\hat{\rho}$ is not guaranteed, which is a slight drawback often shared by estimators based on Fourier inversion and kernel smoothing. However, one can always consider $\hat{\rho}^+(x) = \max(\hat{\rho}(x), 0)$ instead of $\hat{\rho}(x)$. For this modified estimator we have $E[(\hat{\rho}^+(x) - \rho(x))^2] \leq E[(\hat{\rho}(x) - \rho(x))^2]$ and hence its performance is at least as good as that of $\hat{\rho}$, if the mean square error is used as the

performance criterion. We restrict our attention to studying the estimator $\hat{\rho}$ only.

The structure of the paper is as follows: in the next section we will study the asymptotic behaviour of the mean square error of the proposed estimator of ρ. In particular we will derive convergence rates of our estimator over appropriate classes of Lévy triplets and discuss the corresponding lower bounds for estimation of ρ. The section is concluded with a discussion on the obtained results and possible extensions. The proofs of results from Section 2 are collected in Section 3.

2   Results

We first formulate conditions that will be used to establish asymptotic properties of the estimator $\hat{\rho}$. We also supply some comments on these conditions. Introduce a jump size density $f(x) := \rho(x)/\nu(\mathbb{R})$.

Condition 2.1. *Let the unknown Lévy density ρ belong to the class*

$$
W(\beta, L, L', L'', K, \Lambda) = \Big\{ \rho : \rho(x) = \nu(\mathbb{R}) f(x), f \text{ is a probability density,}
$$
$$
\int_{-\infty}^{\infty} |t|^{\beta} |\phi_f(t)| dt \leq L,
$$
$$
|\phi_f(t)| \leq \frac{L'}{|t|^{\beta}},
$$
$$
|\phi'_f(t)| \leq \frac{L''}{|t|^{\beta}},
$$
$$
\int_{-\infty}^{\infty} x^{12} f(x) dx \leq K,
$$
$$
\phi''_f \text{ is integrable,}
$$
$$
\nu(\mathbb{R}) \in (0, \Lambda] \Big\},
$$

where β, L, L', L'', K and Λ are strictly positive numbers.

This condition is similar to the one given in [38] and we refer to the latter for additional discussion. When β is an integer, the integrability condition on ϕ_f in Condition 2.1 is roughly equivalent to f having a derivative of order β. The moment condition on f, and consequently on ρ, is admittedly strong, but on the other hand in mathematical finance it is customary to assume that ρ has a finite exponential moment. The moment condition in Condition 2.1 is used to prove an appropriate maximal inequality for $\hat{\phi}$ and its derivatives, see Theorem 2.2, which constitutes one of the important working tools of the paper.

Condition 2.2. *Let σ be such that $\sigma \in [0, \Sigma]$, where Σ is a strictly positive number.*

For the case when $\Sigma = 0$, that is to say when $\sigma = 0$ is known beforehand, we refer to [24] and [34]. Observe that in general σ determines how fast the characteristic function ϕ_{X_1} decays at plus and minus infinity, because as it is easy to see, one has

$$|\phi_{X_1}(t)| \geq e^{-2\Lambda - \Sigma^2 t^2/2}. \tag{12}$$

The knowledge of Σ, which we will assume, gives us a lower bound on the rate of decay of ϕ_{X_1} at plus and minus infinity (uniformly in $\sigma \in [0, \Sigma]$).

Condition 2.3. *Let γ be such that $|\gamma| \leq \Gamma$, where Γ is a positive number.*

This condition is the same as the one in [38], cf. also [6].

Condition 2.4. *Let the bandwidth $h = h_n$ depend on n and be such that $h_n = (\eta \log n)^{-1/2}$ with $0 < \eta < 1/(2\Sigma^2)$.*

This condition is similar to the one given in [38]. Notice that in order to keep our notation compact, we will suppress the dependence of h_n on n in the notation. The fact that the bandwidth h depends on Σ has a parallel in the condition on the smoothing parameter in [24], see Remark 4.2 there, and also arises in deconvolution problems with unknown error distribution, see [17]. As usual in kernel estimation, see e.g. p. 7 in [54], a choice of h establishes a trade-off between the bias and the variance of the estimator: too small an h will result in an estimator with small bias but large variance, while too large an h results in the estimator with large bias but small variance. From Theorems 2.3 and 2.4 it will follow that the choice of ρ as in Condition 2.4 is optimal in one particular situation in a sense that it asymptotically minimises the order of the mean square error of the estimator $\hat{\rho}$ at a fixed point x.

Condition 2.5. *Let the kernel w be the sinc kernel: $w(x) = \sin x/(\pi x)$.*

The sinc kernel has also been used in [38] when estimating the Lévy density. Its use is frequent in deconvolution problems, see e.g. [17]. The Fourier transform of the sinc kernel is given by $\phi_w(t) = 1_{[-1,1]}(t)$.

Condition 2.6. *Let the sequence κ_n be such that $\kappa_n = \kappa|\log h|^{-1}$ for a constant $\kappa > 0$.*

This is a technical condition used in the proofs. Other sufficiently slowly vanishing sequences $\{\kappa_n\}$ can also be used, ours is just one concrete example. The intuition behind Condition 2.6 is that up to a constant $e^{-2\Lambda}$, the term $e^{-\Sigma^2/(2h^2)}$ gives a lower bound for the absolute value of the characteristic function $\phi_{X_1}(t)$ on the interval $[-h^{-1}, h^{-1}]$, cf. (12). For n large enough, with an indicator 1_{G_t} in the definition of $\hat{\rho}$ we thus cut-off those frequencies t for which $|\hat{\phi}(t)|$ becomes smaller than the lower bound for $|\phi_{X_1}(t)|$ over

$t \in [-h^{-1}, h^{-1}]$. Of course different truncation methods are also possible and we refer e.g. to [24] for an alternative truncation method in the definition of an estimator of a Lévy density in a problem similar to ours. We think that it is natural to incorporate the knowledge of Σ in the selection of the threshold in (9), since the knowledge of Σ is required anyway when selecting the bandwidth h. With our choice of h the set G_t can also be characterised in terms of the sample size n, because h is a function of n, see Condition 2.4. Thus our truncation method is not dissimilar from the one in the deconvolution problem studied in [47].

Next we recall two conditions from [38], which were used to study the asymptotics of the estimator $\hat{\sigma}^2$. For the convenience of a reader we also state a result on the asymptotic behaviour of its mean square error. The latter is used in the proof of Theorem 2.3 below.

Condition 2.7. *Let the kernel $v_h(t) = h^3 v(ht)$, where the function v is continuous and real-valued, has a support on $[-1, 1]$ and is such that*

$$\int_{-1}^{1} v(t)dt = 0, \quad \int_{-1}^{1} \left(-\frac{t^2}{2}\right) v(t)dt = 1, \quad v(t) = O(t^{\beta}) \text{ as } t \to 0.$$

Here β is the same as in Condition 2.1.

It is for simplicity of the proofs that we assume that the smoothing parameter h in the definition of $\hat{\sigma}^2$ is the same as in Condition 2.2. In practice the two need not be equal, although they have to be of the same order.

Condition 2.8. *Let the truncating sequence $M = (M_n)_{n \geq 1}$ be such that $M_n = m_n h^{-2}$, where $m_n = |\log h|^{-1}$.*

Here we implicitly assume that n is large enough, so that m_n is real and $m_n > 0$. Other conditions are also possible, ours is just one concrete example. The use of the truncation in the definition of $\hat{\sigma}^2$ in (11) is that it prevents the estimator from exploding: $|\hat{\phi}(t)|$ can in general take arbitrarily small values and $\log(|\hat{\phi}(t)|)$ consequently can become arbitrarily large.

In the remainder of the paper we will often use the symbols $\lesssim$ and $\gtrsim$ when comparing two sequences a_n and b_n, respectively meaning a_n is less or equal than b_n, or a_n is greater or equal than b_n up to a constant that does not depend on n. The symbol $\asymp$ will be used to denote the fact that two sequences of real numbers are asymptotically of the same order.

Theorem 2.1. *Denote by $\mathcal{T}$ the collection of all Lévy triplets satisfying Conditions 2.1–2.3 and assume Conditions 2.4, 2.7 and 2.8. Let the estimator $\hat{\sigma}^2$ be defined by (11). Then*

$$\sup_{\mathcal{T}} \mathrm{E}\left[(\hat{\sigma}^2 - \sigma^2)^2\right] \lesssim (\log n)^{-\beta - 3}$$

holds.

Even though Condition 2.1 differs slightly from its counterpart in [38], this does not affect the proof of Theorem 2.1. Although the convergence rate of the estimator $\hat{\sigma}^2$ is logarithmic, the contribution of $\hat{\sigma}^2$ to an upper bound on the mean square error of $\hat{\rho}(x)$ is asymptotically negligible compared to other terms, as can be seen from the proof of Theorem 2.3. By techniques similar to those used in [39] in a related deconvolution problem, it is expected that under the same conditions on the class of Lévy triplets as in Theorem 2.1 one can prove that $\hat{\sigma}^2$ is rate-optimal, but since our emphasis in the present work is on estimation of a Lévy density, we refrain from studying this question. For additional discussion on the estimator $\hat{\sigma}^2$ see [38].

Notice that had we not assumed $\nu(\mathbb{R}) \leq \Lambda < \infty$, there would not exist a uniformly consistent estimator of σ^2, see Remark 3.2 in [48]. In fact even the existence of a consistent estimator of σ^2 is not clear in that general setting.

Together with the above theorem, an important tool in studying the estimator $\hat{\rho}$ is the following maximal inequality for the empirical characteristic function $\hat{\phi}(t)$ and its derivatives. Set $\hat{\phi}^{(0)}(t) = \hat{\phi}(t)$ and likewise $\phi_{X_1}^{(0)}(t) = \phi_{X_1}(t)$.

Theorem 2.2. *Let $k \geq 0$ and $r \geq 1$ be integers. Then we have*

$$
\mathrm{E}\left[\left(\sup_{t \in [-h^{-1}, h^{-1}]} |\hat{\phi}^{(k)}(t) - \phi_{X_1}^{(k)}(t)|\right)^r\right]
$$
$$
\lesssim \left(\||x|^{k+1}\|_{\mathbb{L}_{2 \vee r}(\mathrm{P})}^r + \||x|^k\|_{\mathbb{L}_{2 \vee r}(\mathrm{P})}^r\right) \frac{1}{h^r n^{r/2}}, \quad (13)
$$

provided $\||x|^{k+1}\|_{\mathbb{L}_{2 \vee r}(\mathrm{P})}$ is finite. Here the probability P on the righthand side refers to the law of X_1, which is uniquely characterised by the triplet (γ, σ^2, ρ).

The theorem constitutes a generalisation of the corresponding result for $\hat{\phi}$ and $r = 2$ given in [38]. The theorem is of possible general interest as well. For related results on the empirical characteristic function see Theorem 1 in [31] and Theorem 4.1 in [48].

Equipped with the above two theorems, we are now ready to formulate the first main result of the paper, which concerns the mean square error of the estimator $\hat{\rho}$ at a fixed point $x \neq 0$. Notice that we prefer to work with asymptotics uniform in Lévy triplets, since existence of the superefficiency phenomenon in nonparametric estimation makes it difficult to interpret fixed parameter asymptotics, see e.g. [12] for a discussion. This also explains why we imposed certain smoothness assumptions on the class of Lévy densities: too large a class of densities, e.g. of all continuous densities, usually cannot be handled when dealing with uniform asymptotics, see e.g. Theorem 1 on p. 36 in [32] for an example from probability density estimation.

Theorem 2.3. *Denote by $\mathcal{T}$ the collection of all Lévy triplets satisfying Conditions 2.1–2.3 and assume Conditions 2.4–2.8. Let the estimator $\hat{\rho}$ be defined by (9). Then we have*

$$\sup_{\mathcal{T}} \mathrm{E}[(\hat{\rho}(x) - \rho(x))^2] \lesssim (\log n)^{-\beta}$$

for every fixed $x \neq 0$.

Thus the convergence rate of our estimator turns out to be logarithmic, just as for the estimator of ρ proposed in [38]. This result can be easily understood on an intuitive level by comparison to a nonparametric deconvolution problem: if the distribution of the measurement error in a deconvolution model is normal, and if the class of the target densities is massive enough, e.g. some Hölder or Sobolev class (see Definitions 1.2 and 1.11 in [54]), the minimax convergence rate for estimation of an unknown density will be logarithmic for both the mean squared error and mean integrated squared error as measures of risk, see [35] and [36]. Of course the same holds true also for deconvolution models with unknown error variance, see [17] and [45]. Exactly as kernel-type estimators in semiparametric deconvolution problems, our estimator $\hat{\rho}$ also involves division by an estimator of a characteristic function (or to be more precise by its square), a slight difference being that in semiparametric deconvolution problems we divide by an estimator of the characteristic function of the measurement error variable, while in the definition of $\hat{\rho}$ we divide by $\hat{\phi}$, an estimator of ϕ_{X_1}. For large enough n the empirical characteristic function $\hat{\phi}$ should be close to the true characteristic function ϕ_{X_1} on the interval $[-h^{-1}, h^{-1}]$. Since up to a constant term, ϕ_{X_1} behaves at plus and minus infinity as a normal characteristic function, the logarithmic convergence rate of the estimator ρ is then no surprise. Exactly as in normal deconvolution problem over a Hölder or Sobolev class of densities, cf. [35] and [36], it is due to the dominating squared bias of $\hat{\rho}$, i.e. roughly speaking the term T_1 in the proof of Theorem 2.3. More formally, in the theorem given below we actually prove that our estimator $\hat{\rho}$ attains the minimax convergence rate for estimation of the Lévy density ρ at a fixed point x over a suitable class of Lévy triplets when the risk is measured by the mean square error.

Theorem 2.4. *Let T be a Lévy triplet (γ, σ^2, ρ), such that $|\gamma| \leq \Gamma$, $\sigma \in [0, \Sigma]$, $\nu(\mathbb{R}) \in (0, \Lambda)$, where Γ, Σ and Λ are strictly positive constants. Assume furthermore that*

$$\int_{-\infty}^{\infty} |t|^\beta |\phi_f(t)| dt \leq L; \quad |\phi_f(t)| \leq \frac{L'}{|t|^\beta}; \quad |\phi_f'(t)| \leq \frac{L'}{|t|^\beta} \tag{14}$$

for strictly positive constants β, L, L' and L''. Let $\mathcal{T}$ be a collection of all Lévy triplets satisfying these conditions. Then for every fixed $x \neq 0$ we have

$$\inf_{\widetilde{\rho}_n} \sup_{\mathcal{T}} \mathrm{E}[(\widetilde{\rho}_n(x) - \rho(x))^2] \gtrsim (\log n)^{-\beta}, \tag{15}$$

11

where the infimum on the lefthand side is taken over all estimators $\widetilde{\rho}_n$ based on observations $X_1, \ldots, X_n$.

The proof of the theorem is such that it also works for the case when $\sigma > 0$ is assumed known and is fixed. Therefore the knowledge of σ does not lead to some estimator of ρ with a better rate of convergence. This is unlike the semiparametric deconvolution problem with unknown error variance, see [17], where the fact that the measurement error variance is unknown slows down even further the convergence rate. Disregarding the moment condition in Condition 2.1, an easy consequence of Theorems 2.3 and 2.4 is that $\hat{\rho}$ is rate-optimal.

A slow, logarithmic convergence rate of $\hat{\rho}$ seems to indicate that samples of very large size are needed to accurately estimate ρ. However, it is known that in deconvolution problems kernel-type density estimators perform well for reasonable sample sizes, provided the noise term variance is not too large, see e.g. [30], [33] or [57]. Likewise, a spectral cut-off method of [6] and [7] produces good results for small values of σ in the problem of calibration of exponential Lévy models. Since in the financial setting it is perhaps unnatural to assume that σ is known and $\sigma \to 0$ as $n \to \infty$, which constitutes the mathematical formalisation of the statement that in the asymptotic setting the noise level is low, and since in the present work we are mainly concerned with asymptotics, we will explore a different possibility, namely that the Lévy density is much smoother than the Hölder or Sobolev class Lévy densities. Our results will parallel those from [18], where it is shown in the deconvolution context that better than logarithmic convergence rates can be obtained in case when the target density is supersmooth itself, i.e. essentially has a characteristic function that decays exponentially fast at plus and minus infinity.

We first give a condition on the class of Lévy densities.

Condition 2.9. *Let the unknown Lévy density ρ belong to the class*

$$A(\alpha, s, L, L', L'', K, \Lambda) = \Big\{ \rho : \rho(x) = \nu(\mathbb{R}) f(x), f \text{ is a probability density},$$

$$\int_{-\infty}^{\infty} |\phi_f(t)|^2 \exp(2\alpha |t|^s) dt \leq L,$$

$$\frac{|\phi_f(t)|}{|t|^{(1-s)/2} e^{-\alpha |t|^s}} \leq L',$$

$$\frac{|\phi_f'(t)|}{|t|^{(1-s)/2} e^{-\alpha |t|^s}} \leq L'',$$

$$\int_{-\infty}^{\infty} x^{12} f(x) dx \leq K,$$

$$\phi_f'' \text{ is integrable},$$

$$\nu(\mathbb{R}) \in (0, \Lambda] \Big\},$$

where α, s, L, K and Λ are strictly positive numbers.

The 'size' of the class $A(\alpha, s, L, L', L'', K, \Lambda)$ is much smaller than the 'size' of the class $W(\beta, L, L', L'', K, \Lambda)$, and it is intuitively clear that better convergence rates can be expected for estimation of ρ over the former class than over the latter class. We will refer to the class $A(\alpha, s, L, L', L'', K, \Lambda)$ as the class of supersmooth Lévy densities.

Since the estimator $\hat{\rho}$ depends on the estimator $\hat{\sigma}^2$, we first need to study the asymptotics of the latter. With a different class of Lévy densities than in Theorem 2.1, the conditions on the bandwidth h and kernel v_h have to be modified accordingly. These are supplied below.

Condition 2.10. *Let the bandwidth h depend on n and be such that h is a positive solution of the equation*

$$\frac{2\alpha}{h^s} + \frac{2\Sigma^2}{h^2} = \log n - (\log\log n)^2. \tag{16}$$

Here we thus suppose that s is known. We also assume that n is large enough, so that equation (16) indeed has a positive root. Condition 2.10 is motivated by a similar condition on the bandwidth in the deconvolution problem studied in [18]. An optimal bandwidth, i.e. a bandwidth that asymptotically minimises the risk of the estimator (or an upper bound on it), is typically computed in kernel estimation by differentiating an upper bound on the risk of the estimator with respect to h, setting the derivative to zero and solving h from the obtained equation. However, in our case an optimal h can also be computed from (16), cf. Section 3 in [18], and we give the corresponding argument in the proof of Theorem 2.6. The two methods of course yield the same asymptotic results.

Condition 2.11. *Let the kernel $v_h(t) = h^3 v(ht)$, where the function v is continuous and real-valued, has a support on $[-\sqrt{2}, -1] \bigcup [1, \sqrt{2}]$ and is such that*

$$\int_{\mathbb{R}} v(t) dt = 0, \quad \int_{\mathbb{R}} \left(-\frac{t^2}{2} \right) v(t) dt = 1.$$

Instead of defining the support of v by $[-\sqrt{2}, -1] \bigcup [1, \sqrt{2}]$, we could have defined it as $[-a, -1] \bigcup [1, a]$ for $1 < a \le \sqrt{2}$, which would result in a better convergence rate for $\hat{\sigma}^2$. However, $a = \sqrt{2}$ actually suffices for the purpose of estimation of ρ, as a contribution of $\hat{\sigma}^2$ to an upper bound on the risk of $\hat{\rho}$ will still be asymptotically of at most the same order as that of other terms, cf. the proof of Theorem 2.6. We do not address the problem of constructing a rate-optimal estimator of σ^2 in the present paper.

The following result holds true.

Theorem 2.5. *Denote by $\mathcal{T}$ the collection of all Lévy triplets satisfying Conditions 2.2, 2.3 and 2.9 and assume that Conditions 2.8, 2.10 and 2.11 hold. Let $s < 2$ and let the estimator $\hat{\sigma}^2$ be defined by (11). Then*

$$\sup_{\mathcal{T}} \mathrm{E}\left[(\hat{\sigma}^2 - \sigma^2)^2\right] \lesssim h^{s+5} \exp\left(-\frac{2\alpha}{h^s}\right)$$

holds, where h is defined in Condition 2.10.

The asymptotics of the estimator $\hat{\sigma}^2$ (and also those of $\hat{\rho}$) change qualitatively when $s > 2$. In particular, the convergence rate of $\hat{\sigma}^2$ becomes polynomial. Although supersmooth densities with $s > 2$ are in principle conceivable, they do not include well-known representatives of the class of supersmooth densities, cf. a relevant discussion in [18]. Therefore without much loss of generality we assume that $s < 2$.

With the above result we can finally study the asymptotics of $\hat{\rho}$ over the class of supersmooth Lévy densities.

Theorem 2.6. *Suppose that conditions of Theorem 2.5 are satisfied and let in addition Condition 2.6 hold. Then we have*

$$\sup_{\mathcal{T}} \mathrm{E}[(\hat{\rho}(x) - \rho(x))^2] \lesssim h^{s-1} \exp\left(-\frac{2\alpha}{h^s}\right)$$

for every fixed $x \neq 0$. In particular, for $s = 1$ an upper bound

$$\sup_{\mathcal{T}} \mathrm{E}[(\hat{\rho}(x) - \rho(x))^2] \lesssim \exp\left(-2\alpha\left(\frac{\log n}{2\Sigma^2}\right)^{1/2}\right) \qquad (17)$$

is valid.

Since $h \asymp (\log n)^{-1/2}$, which can be shown as formula (27) of [18], it is easy to see that the convergence rate of $\hat{\rho}$ is faster than any power of $\log n$ and hence much better than that in Theorem 2.3. The case $s = 1$ is particularly interesting, as it corresponds to the class of Lévy densities that admit an analytic continuation into a strip of the complex plane.

A natural question is whether $\hat{\rho}$ is rate-optimal over a class of supersmooth Lévy densities. We will not provide a formal statement and its proof, but instead will restrict ourselves to an intuitive discussion, which we hope is more enlightening. To answer the question of rate-optimality of $\hat{\rho}$, one has first to establish a lower bound for estimation of $\rho(x)$ over a class of supersmooth Lévy densities. Disregarding the moment condition in Condition 2.9, this can be done by following a general scheme of the proof of Theorem 2.4 combined with some of the techniques from [17], [19] or [39]. This lower bound will be similar to the one given in Theorem 4 in [19] and in fact for $s = 1$ one will have

$$\inf_{\widetilde{\rho}_n} \sup_{\mathcal{T}} \mathrm{E}[(\widetilde{\rho}_n(x) - \rho(x))^2] \gtrsim \exp\left(-2\alpha\left(\frac{\log n}{\Sigma^2}\right)^{1/2}\right), \qquad (18)$$

where the infimum is taken over the class of all estimators $\widetilde{\rho}$ based on a sample $X_1, \ldots, X_n$ from the process X. Unfortunately, the lower bound in (18) is too small in comparison to the upper bound in (17). Although we are not completely sure, we still think that the lower and upper risk bounds that we give in Theorem (17) and (18) are sharp as far as their rates of decay are concerned: we think that it is the estimator $\hat{\rho}$ that cannot attain the minimax convergence rate. Given that this is true, an intuitive explanation of the suboptimality of $\hat{\rho}$ in the present setting might be the following: the construction of $\hat{\rho}$ in (9) involves division by $(\hat{\phi}(t))^2$, which is close to $(\phi_X(t))^2$ on $[-h^{-1}, h^{-1}]$ for n large enough. Hence in essence we are dealing with a kernel-type deconvolution density estimator which involves division by $(\phi_X(t))^2$, whereas in conventional deconvolution problems the kernel estimator involves division by the characteristic function of the measurement error variable and not its square, see e.g. [35]. By a rough analogy, assuming that the Gaussian component in the Lévy process plays a role similar to the measurement error in the deconvolution problems, one can see that the variance of our estimator $\hat{\rho}$ of a Lévy density is larger than the variance of a kernel-type deconvolution density estimator, compare p. 1266 in [35] and an upper bound on the term T_2 in the proof of Theorem 2.6. In order to render the variance asymptotically negligible, a somewhat larger bandwidth would thus be required in the former case than in the latter case. However, unlike the case when the Lévy density satisfies Condition 2.1, this has a dramatic effect on the bias of the estimator (as far as its order is concerned) for the class of supersmooth Lévy densities and the suboptimality of $\hat{\rho}$ results: it is the squared bias, or roughly speaking the term T_1 in the proof of Theorem 2.6, that dominates the asymptotics of $\hat{\rho}$. No such problem seems to arise in [24], where unlike our setting it is a priori assumed that $\gamma = 0, \sigma = 0$, and as a consequence one can derive a different inversion formula than (7), cf. formula (19) below, which involves only division by ϕ_{X_1} and not by its square.

In light of the above observations another natural question that arises in this context is whether one has to use (4) instead of (7) as a basis of construction of an estimator of ρ: under appropriate conditions with the former formula one can express the Lévy density ρ as

$$\rho(x) = -\frac{1}{2\pi x} \int_{\mathbb{R}} e^{-itx} \left(i \frac{\phi'_{X_1}(t)}{\phi_{X_1}(t)} + \gamma + i\sigma^2 t \right) dx, \qquad (19)$$

which involves division by the first power of ϕ_{X_1} only. By replacing ϕ_{X_1} by the empirical characteristic function $\hat{\phi}$ and σ^2 and γ by their estimators and by application of an appropriate amount of regularisation we would thus get an estimator of ρ that in its form is closer to a conventional kernel-type deconvolution density estimator in that under the integral sign it involves division by the first power of the (estimated) characteristic function only. It is nevertheless unclear whether this approach can lead to an estimator

of ρ with a better (optimal in the best case) convergence rate than the one we are considering in the present work: one has to find estimators of γ and σ^2 that converge at an optimal rate in the present context, which does not seem to be an easy task.

Another interesting question that arises in the present context is that of adaptation: construction of our estimator of ρ does rely on knowledge of the smoothness degree of a Lévy density, see in particular Conditions 2.7, 2.10 and 2.11. In practice it might happen that this smoothness degree is unknown and it is desirable to have an estimator of ρ that automatically achieves the optimal rate of convergence without knowledge of the smoothness degree of a Lévy density. We view this as a separate problem and do not address it in the present work. Relevant results are available in the context of pure jump Lévy processes and we refer e.g. to [24] for additional details. Note that the proofs of the adaptation results in that paper require nontrivial amount of technical work. In any case, in our setting an adaptive estimator and σ^2 would be required.

We conclude this section by a brief comparison of $\hat{\rho}$ to the estimator ρ_n of ρ proposed in [38]. Up to some additional truncation, the latter estimator is given by

$$\rho_n(x) = \frac{1}{2\pi} \int_{-1/h}^{1/h} e^{-itx} \operatorname{Log}\left(\frac{\hat{\phi}(t)}{e^{i\hat{\gamma}t}e^{-\hat{\nu}(\mathbb{R})}e^{-\hat{\sigma}^2 t^2/2}}\right) dt, \qquad (20)$$

where Log denotes the so-called distinguished logarithm, i.e. a 'logarithm' that is a continuous and single-valued function of t, see Theorem 7.6.2 of [22] for its construction. Furthermore, $\hat{\gamma}$, $\hat{\nu}(\mathbb{R})$ and $\hat{\sigma}^2$ are estimators of the parameters $\gamma, \nu(\mathbb{R})$ and σ^2, respectively. Notice that in general the distinguished logarithm $\operatorname{Log}(g(t))$ of some function g is not a composition of a fixed branch of an ordinary logarithm with g. The estimator ρ_n seems to be given by a more complicated expression than $\hat{\rho}$, because it depends explicitly on estimators of γ and $\nu(\mathbb{R})$ in addition to the estimator of σ^2. The matter is furthermore complicated by the need to use the distinguished logarithm. The latter in (20) can be defined only for those ω's from the sample space Ω for which $\hat{\phi}$ as a function of t does not hit zero on $[-h^{-1}, h^{-1}]$. For those ω's for which this is not satisfied, ρ_n has to be assigned an arbitrary value, e.g. one can assume that ρ_n is a standard normal density. It is shown in [38] that as $n \to \infty$, the probability of the event that $\hat{\phi}$ hits zero for t in $[-h^{-1}, h^{-1}]$ vanishes under appropriate conditions. However, an almost sure result of a similar type remains to be unknown (it has been established only in the context of [6] in [53]). Also in practice the fact that $\hat{\phi}$ does not vanish can be checked for a discrete grid of points t only and it could happen that one misses the fact that $\hat{\phi}(t)$ is zero for some $t \in [-h^{-1}, h^{-1}]$. All this seems to be a disadvantage of the estimator ρ_n. On the other hand the estimator $\hat{\rho}$ is undefined for $x = 0$ and a study of its asymptotic properties requires

16

stronger moment conditions on the Lévy density ρ. Also, a division by x^2 in the vicinity of the origin might render it numerically unstable. In conclusion, both estimators are rate-optimal over an appropriate class of Lévy triplets, but each of them seems to have its own advantages over another.

3 Proofs

Proof of Theorem 2.2. The proof is similar in spirit to the one in [38], pp. 334–335, which in turn mimicks the one in [17], pp. 326–327. Since both of the proofs are deficient, here we also seize an opportunity to rectify them.

We have

$$
\mathrm{E}\left[\left(\sup_{t\in[-h^{-1},h^{-1}]}|\hat{\phi}^{(k)}(t)-\phi_{X_1}^{(k)}(t)|\right)^r\right]=\frac{1}{n^{r/2}}\mathrm{E}\left[\left(\sup_{t\in[-h^{-1},h^{-1}]}|G_nv_{t,k}|\right)^r\right],
$$

where $G_n v_{t,k}$ denotes an empirical process

$$
G_n v_{t,k}=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(v_{t,k}(Z_j)-\mathrm{E}[v_{t,k}(Z_j)])
$$

and the function $v_{t,k}$ is defined as $v_{t,k}:x\mapsto(ix)^ke^{itx}$. Introduce the functions $v_{t,k,1}:x\mapsto x^k\sin(tx)$ and $v_{t,k,2}:x\mapsto x^k\cos(tx)$. Since $|i^k|=1$ and $e^{itx}=\cos(tx)+i\sin(tx)$, the c_r-inequality gives

$$
\mathrm{E}\left[\left(\sup_{t\in[-h^{-1},h^{-1}]}|G_nv_{t,k}|\right)^r\right]\lesssim\mathrm{E}\left[\left(\sup_{t\in[-h^{-1},h^{-1}]}|G_nv_{t,k,1}|\right)^r\right]
$$
$$
+\mathrm{E}\left[\left(\sup_{t\in[-h^{-1},h^{-1}]}|G_nv_{t,k,2}|\right)^r\right].
$$

Furthermore, by differentiability of $v_{t,k,j}$ with respect to t and the mean-value theorem we have

$$
|v_{t,k,j}(x)-v_{s,k,j}(x)|\leq|x|^{k+1}|t-s| \tag{21}
$$

for $j=1,2$. Consequently, for a fixed x the function $v_{t,k,j}$ is Lipschitz in t with a Lipschitz constant $|x|^{k+1}$.

In what follows we will need some results from the theory of empirical processes. For all the unexplained terminology and notation we refer e.g. to Section 19.2 of [55] or Section 2.1.1 of [56]. First of all, by the inequality (21) and by Theorem 2.7.11 of [56] the bracketing number $N_{[]}$ of the class of functions $\mathbb{F}_{n,j}$ (for $j=1,2$ this refers to the collection of functions $v_{t,k,j}$ for $t\in[-h^{-1},h^{-1}]$) can be bounded by the covering number N of the interval $I_n=[-h^{-1},h^{-1}]$ as follows

$$
N_{[]}(2\epsilon\||x|^{k+1}\|_{\mathbb{L}_2(Q)};\mathbb{F}_{n,j};\mathbb{L}_2(Q))\leq N(\epsilon;I_n;|\cdot|).
$$

17

Here Q is any probability measure. Since it is easily seen that for the covering and bracketing numbers of the classes $\mathbb{F}_{n,j}$, $j = 1, 2$, we have the inequality

$$N(\epsilon \||x|^{k+1}\|_{\mathbb{L}_2(Q)}; \mathbb{F}_{n,j}; \mathbb{L}_2(Q)) \le N_{[]}(2\epsilon \||x|^{k+1}\|_{\mathbb{L}_2(Q)}; \mathbb{F}_{n,j}; \mathbb{L}_2(Q)),$$

cf. p. 84 in [56], and since

$$N(\epsilon; I_n; |\cdot|) \le \frac{1}{\epsilon}\frac{2}{h} + 1,$$

we obtain that

$$N(\epsilon \||x|^{k+1}\|_{\mathbb{L}_2(Q)}; \mathbb{F}_{n,j}; \mathbb{L}_2(Q)) \le \frac{1}{\epsilon}\frac{2}{h} + 1. \tag{22}$$

By taking $s = 0$, it follows from the definition of $v_{t,k,j}$ and (21) that the function $F_{h,1}(x) = |x|^{k+1}h^{-1}$ can be used as an envelope for the class $\mathbb{F}_{n,1}$, while $F_{h,2}(x) = |x|^{k+1}h^{-1} + |x|^k$ can serve as an envelope for $\mathbb{F}_{n,2}$. Next define $J(1, \mathbb{F}_{n,j})$, the entropy of the class $\mathbb{F}_{n,j}$, as

$$J(1, \mathbb{F}_{n,j}) = \sup_Q \int_0^1 \{1 + \log(N(\epsilon \|F_{h,j}(x)\|_{\mathbb{L}_2(Q)}; \mathbb{F}_{n,j}; \mathbb{L}_2(Q)))\}^{1/2} d\epsilon,$$

where $j = 1, 2$, and the supremum is taken over all discrete probability measures Q, such that $\|F_{h,j}(x)\|_{\mathbb{L}_2(Q)} > 0$. Notice that $\mathbb{F}_{n,j}$'s are measurable classes of functions with measurable envelopes. Theorem 2.14.1 in [56] then implies that

$$\mathrm{E}\left[\left(\sup_{t \in [-h^{-1}, h^{-1}]} |G_n v_{t,k,j}|\right)^r\right] \lesssim \|F_{h,j}(x)\|_{\mathbb{L}_{2 \vee r}(\mathrm{P})}^r (J(1, \mathbb{F}_{n,j}))^r.$$

Here the probability measure P on the righthand side is associated with the distribution of X_1. We next need to work out the quantities on the righthand side of the above display. Observe that

$$\|F_{h,1}(x)\|_{\mathbb{L}_{2 \vee r}(\mathrm{P})}^r = \frac{1}{h^r} \||x|^{k+1}\|_{\mathbb{L}_{2 \vee r}(\mathrm{P})}^r.$$

Moreover, we have

$$\|F_{h,2}(x)\|_{\mathbb{L}_{2 \vee r}(\mathrm{P})}^r \lesssim \frac{1}{h^r}(\||x|^{k+1}\|_{\mathbb{L}_{2 \vee r}(\mathrm{P})}^r + \||x|^k\|_{\mathbb{L}_{2 \vee r}(\mathrm{P})}^r),$$

provided $h \le 1$. Here we also used the $c_{2 \vee r}$-inequality. It thus remains to bound the entropy $J(1, \mathbb{F}_{n,j})$. By the fact that

$$\|F_{h,1}(x)\|_{\mathbb{L}_2(Q)} = h^{-1}\||x|^{k+1}\|_{\mathbb{L}_2(Q)}$$

and by taking ϵ/h instead of ϵ in (22) we get

$$N(\epsilon \|F_{h,1}(x)\|_{\mathbb{L}_2(Q)}; \mathbb{F}_{n,j}; \mathbb{L}_2(Q)) \le \frac{2}{\epsilon} + 1. \tag{23}$$

Furthermore, since $\|F_{h,2}(x)\|_{\mathbb{L}_2(Q)} \geq \||x|^{k+1}h^{-1}\|_{\mathbb{L}_2(Q)}$, by monotonicity of the covering number N in the size of the covering balls combined with (23) we obtain that

$$N(\epsilon\|F_{h,2}(x)\|_{\mathbb{L}_2(Q)}; \mathbb{F}_{n,j}; \mathbb{L}_2(Q)) \leq \frac{2}{\epsilon} + 1. \qquad (24)$$

Inserting the bounds from (23) and (24) into the definition of $J(1, \mathbb{F}_{n,j})$, we see that

$$J(1, \mathbb{F}_{n,j}) \leq \int_0^1 \left\{ 1 + \log\left(\frac{2}{\epsilon} + 1\right) \right\}^{1/2} d\epsilon < \infty.$$

This yields the statement of the theorem. $\qquad\qquad\qquad\qquad\qquad\square$

Proof of Theorem 2.3. By the c_2-inequality we have

$$\mathrm{E}[(\hat{\rho}(x) - \rho(x))^2] \lesssim |\rho(x) - \widetilde{\rho}(x)|^2 + \mathrm{E}[|\hat{\rho}(x) - \widetilde{\rho}(x)|^2] = T_1 + T_2,$$

where

$$\widetilde{\rho}(x) = \frac{1}{2\pi x^2} \int_{-1/h}^{1/h} e^{-itx} \left(\frac{(\phi'_{X_1}(t))^2 - \phi''_{X_1}(t)\phi_{X_1}(t)}{(\phi_{X_1}(t))^2} - \sigma^2 \right) dt.$$

We will first work out the term T_1. By (6) we have

$$-\phi''_\rho(t) = \frac{(\phi'_{X_1}(t))^2 - \phi''_{X_1}(t)\phi_{X_1}(t)}{(\phi_{X_1}(t))^2} - \sigma^2.$$

Then by the Fourier inversion argument we can write

$$\rho(x) - \widetilde{\rho}(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \phi_\rho(t) dt + \frac{1}{2\pi x^2} \int_{-1/h}^{1/h} e^{-itx} \phi''_\rho(t) dt.$$

Integrating by parts twice the second term on the righthand side of the above display and using Condition 2.1, we obtain

$$\frac{1}{2\pi x^2} \int_{-1/h}^{1/h} e^{-itx} \phi''_\rho(t) dt = -\frac{1}{2\pi} \int_{-1/h}^{1/h} e^{-itx} \phi_\rho(x) dx + O(h^\beta),$$

where the $O(h^\beta)$ term on the righthand side is uniform in ρ. With this in mind and by the fact that $\phi_\rho(t) = \nu(\mathbb{R})\phi_f(t)$, we can bound T_1 using the c_2-inequality as

$$\begin{aligned}
T_1 &\lesssim \frac{\Lambda^2}{4\pi^2} \left(\int_{\mathbb{R}\setminus[-h^{-1}, h^{-1}]} |\phi_f(t)| dt \right)^2 + h^{2\beta} \\
&\lesssim \left(\int_{\mathbb{R}\setminus[-h^{-1}, h^{-1}]} |t|^\beta |t|^{-\beta} |\phi_f(t)| dt \right)^2 + h^{2\beta} \\
&\leq \left(\int_{-\infty}^\infty |t|^\beta |\phi_f(t)| dt \right)^2 h^{2\beta} + h^{2\beta} \\
&\lesssim h^{2\beta},
\end{aligned}$$

19

provided that $h \leq 1$. Hence by Condition 2.2 the term $\sup_{\mathcal{T}} T_1$ is of order $(\log n)^{-\beta}$. This is the term that has the dominating contribution to the risk of $\hat{\rho}$. The rest of the proof is dedicated to showing that T_2 is negligible in comparison to T_1. This involves a long series of inequalities.

By the c_2-inequality we have

$$T_2 \lesssim \frac{1}{4\pi^2 x^4} \left| \int_{-1/h}^{1/h} e^{-itx} dt \right|^2 \mathrm{E}\left[|\hat{\sigma}^2 - \sigma^2|^2\right]$$
$$+ \frac{1}{4\pi^2 x^4} \mathrm{E}\left[\left| \int_{-1/h}^{1/h} e^{-itx}(\Phi(\hat{\phi}(t))1_{G_t} - \Phi(\phi(t)))dt \right|^2 \right]$$
$$= T_3 + T_4,$$

where for a twice differentiable function ζ the mapping Φ is defined by

$$\Phi(\zeta(t)) = \frac{(\zeta'(t))^2 - \zeta''(t)\zeta(t)}{(\zeta(t))^2}.$$

By Theorem 2.1 in combination with Condition 2.4 we have $\sup_{\mathcal{T}} T_3 \lesssim (\log n)^{-\beta-2}$. Next notice that

$$T_4 \leq \frac{1}{\pi^2 x^4 h^2} \mathrm{E}\left[\left(\sup_{t \in [-h^{-1}, h^{-1}]} |\Phi(\hat{\phi}(t))1_{G_t} - \Phi(\phi(t))| \right)^2 \right]$$
$$= \frac{T_5}{\pi^2 x^4}.$$

Hence it remains to study T_5. This will be done via repeated applications of Theorem 2.2. First of all, the c_2-inequality gives

$$T_5 \lesssim \frac{1}{h^2} \mathrm{E}\left[\left(\sup_{t \in [-h^{-1}, h^{-1}]} \left| \frac{\hat{\phi}''(t)}{\hat{\phi}(t)}1_{G_t} - \frac{\phi_{X_1}''(t)}{\phi_{X_1}(t)} \right| \right)^2 \right]$$
$$+ \frac{1}{h^2} \mathrm{E}\left[\left(\sup_{t \in [-h^{-1}, h^{-1}]} \left| \frac{(\hat{\phi}'(t))^2}{(\hat{\phi}(t))^2}1_{G_t} - \frac{(\phi_{X_1}'(t))^2}{(\phi_{X_1}(t))^2} \right| \right)^2 \right]$$
$$= T_6 + T_7.$$

By another application of the c_2-inequality we obtain

$$T_6 \lesssim \frac{1}{h^2} \mathrm{E}\left[\left(\sup_{t \in [-h^{-1}, h^{-1}]} \left| \frac{\hat{\phi}''(t)}{\hat{\phi}(t)}1_{G_t} - \frac{\phi_{X_1}''(t)}{\phi_{X_1}(t)}1_{G_t} \right| \right)^2 \right]$$
$$+ \frac{1}{h^2} \mathrm{E}\left[\left(\sup_{t \in [-h^{-1}, h^{-1}]} \left(\left| \frac{\phi_{X_1}''(t)}{\phi_{X_1}(t)} \right| 1_{G_t^c} \right) \right)^2 \right]$$
$$= T_8 + T_9.$$

20

The term T_8 in the last equality can be bounded as follows,

$$T_8 \lesssim \frac{1}{h^2}\mathrm{E}\left[\left(\sup_{t\in[-h^{-1},h^{-1}]}\left|\frac{\hat{\phi}''(t)}{\hat{\phi}(t)}1_{G_t} - \frac{\hat{\phi}''(t)}{\phi_{X_1}(t)}1_{G_t}\right|\right)^2\right]$$

$$+ \frac{1}{h^2}\mathrm{E}\left[\left(\sup_{t\in[-h^{-1},h^{-1}]}\left|\frac{\hat{\phi}''(t)}{\phi_{X_1}(t)}1_{G_t} - \frac{\phi''_{X_1}(t)}{\phi_{X_1}(t)}1_{G_t}\right|\right)^2\right]$$

$$= T_{10} + T_{11}.$$

Further bounding gives

$$T_{10} \le \frac{1}{h^2}\mathrm{E}\left[\left(\sup_{t\in[-h^{-1},h^{-1}]}|\hat{\phi}''(t)|\right)^2\left(\sup_{t\in[-h^{-1},h^{-1}]}\left(\frac{|\hat{\phi}(t)-\phi_{X_1}(t)|}{|\hat{\phi}(t)||\phi_{X_1}(t)|}1_{G_t}\right)\right)^2\right].$$

Now apply the Cauchy-Schwarz inequality to the righthand side to obtain

$$T_{10} \le \frac{1}{h^2}\left(\mathrm{E}\left[\left(\sup_{t\in[-h^{-1},h^{-1}]}|\hat{\phi}''(t)|\right)^4\right]\right)^{1/2}$$

$$\times \left(\mathrm{E}\left[\left(\sup_{t\in[-h^{-1},h^{-1}]}\left(\frac{|\hat{\phi}(t)-\phi_{X_1}(t)|}{|\hat{\phi}(t)||\phi_{X_1}(t)|}1_{G_t}\right)\right)^4\right]\right)^{1/2}$$

$$= \frac{1}{h^2}\sqrt{T_{12}}\sqrt{T_{13}}.$$

Observe that by the fact that $|\hat{\phi}''(t)| \le n^{-1}\sum_{j=1}^n Z_j^2$ and by the c_4-inequality

$$T_{12} \le \mathrm{E}\left[\left|\frac{1}{n}\sum_{j=1}^n Z_j^2\right|^4\right]$$

$$\le \frac{c_4}{n^4}\mathrm{E}\left[\left|\sum_{j=1}^n(Z_j^2 - \mathrm{E}[Z_j^2])\right|^4\right] + c_4(\mathrm{E}[Z_1^2])^4$$

$$\le (3\sqrt{2})^4 4^{4/2}\frac{c_4}{n^2}\mathrm{E}[(Z_1^2 - \mathrm{E}[Z_1^2])^4] + c_4(\mathrm{E}[Z_1^2])^4,$$

where the last inequality follows from the Marcinkiewicz-Zygmund inequality as given in Theorem 2 of [50]. By the Lyapunov inequality $(\mathrm{E}[Z_1^2])^4 \le \mathrm{E}[Z_1^8]$. This in combination with the c_4-inequality gives $\mathrm{E}[(Z_1^2 - \mathrm{E}[Z_1^2])^4] \lesssim \mathrm{E}[Z_1^8]$. It remains to bound $\mathrm{E}[Z_1^8]$ uniformly in Lévy triplets. The most direct way of doing this is to notice that

$$\mathrm{E}[Z_1^8] = \mathrm{E}[(\gamma + \sigma W + Y)^8] \lesssim \Gamma^8 + \Sigma^8\mathrm{E}[W^8] + \mathrm{E}[Y^8],$$

where W is a standard normal random variable, while Y has a compound Poisson distribution with intensity $\nu(\mathbb{R})$ and jump size density f. Observe that $\mathrm{E}\left[Y^8\right] = \phi_Y^{(8)}(0)$ and that under Condition 2.1 and with the Lyapunov inequality it is laborious, though straightforward to show that $\phi_Y^{(8)}(0)$ is bounded by a universal constant uniformly in Lévy triplets. Hence the term $\sup_{\mathcal{T}} \mathrm{E}\left[Z_1^8\right]$ is also bounded and then so is $\sup_{\mathcal{T}} \sqrt{T_{12}}$. As far as T_{13} is concerned, we have

$$T_{13} \lesssim \frac{e^{4\Sigma^2/h^2}}{\kappa_n^4} \mathrm{E}\left[\left(\sup_{t\in[-h^{-1},h^{-1}]} |\hat{\phi}(t) - \phi_{X_1}(t)|\right)^4\right],$$

which follows from Conditions 2.1 and 2.2. Inequality (13) with $k = 0$ and $r = 4$ then yields

$$T_{13} \lesssim \||x|\|_{\mathbb{L}_4(\mathrm{P})}^4 \frac{e^{4\Sigma^2/h^2}}{\kappa_n^4 h^4 n^2}.$$

Since $\||x|\|_{\mathbb{L}_4(\mathrm{P})}$ is bounded by a constant uniformly in Lévy triplets (this can be proved by essentially the same argument as we used for $\sup_{\mathcal{T}} \mathrm{E}\left[Z_1^8\right]$ above), it follows that $\sup_{\mathcal{T}} T_{13}$ is negligible in comparison to $(\log n)^{-\beta}$. This is also true for $h^{-2} \sup_{\mathcal{T}} \sqrt{T_{13}}$ and then also for $\sup_{\mathcal{T}} T_{10}$. To complete the study of T_8, we need to study T_{11}. The latter can be bounded as follows:

$$T_{11} \lesssim \frac{e^{\Sigma^2/h^2}}{h^2} \mathrm{E}\left[\left(\sup_{t\in[-h^{-1},h^{-1}]} |\hat{\phi}''(t) - \phi_{X_1}''(t)|\right)^2\right].$$

By the same reasoning as above one can show that $\sup_{\mathcal{T}} T_{11}$ is negligible compared to $(\log n)^{-\beta}$. Consequently, so is $\sup_{\mathcal{T}} T_8$. Next we deal with T_9. Notice that by our conditions and the Lyapunov inequality

$$\left|\frac{\phi_{X_1}''(t)}{\phi_{X_1}(t)}\right| \leq \left|\frac{\phi_{X_1}'(t)}{\phi_{X_1}(t)}\right|^2 + \sigma^2 + \int_{-\infty}^{\infty} x^2 \rho(x) dx$$

$$\leq \left(\Gamma + \Sigma^2 \frac{1}{h} + \Lambda K^{1/12}\right)^2 + \Sigma^2 + \Lambda K^{1/6}$$

$$\lesssim \frac{1}{h^2}.$$

Hence it holds that

$$\sup_{\mathcal{T}} \sup_{t\in[-h^{-1},h^{-1}]} \left|\frac{\phi_{X_1}''(t)}{\phi_{X_1}(t)}\right| \lesssim \frac{1}{h^2}. \tag{25}$$

Consequently, we have

$$T_9 \lesssim \frac{1}{h^4} \mathrm{E}\left[\left(\sup_{t\in[-h^{-1},h^{-1}]} 1_{G_t^c}\right)^2\right].$$

We study the expectation on the righthand side. First of all, for $t \in [-h^{-1}, h^{-1}]$ and all n large enough we have

$$
\begin{aligned}
G_t^c &= \left\{ |\hat{\phi}(t)| - |\phi_{X_1}(t)| < \kappa_n e^{-\Sigma^2/(2h^2)} - |\phi_{X_1}(t)| \right\} \\
&= \left\{ |\phi_{X_1}(t)| - |\hat{\phi}(t)| > |\phi_{X_1}(t)| - \kappa_n e^{-\Sigma^2/(2h^2)} \right\} \\
&\subseteq \left\{ |\phi_{X_1}(t) - \hat{\phi}(t)| > (e^{-2\Lambda} - \kappa_n) e^{-\Sigma^2/(2h^2)} \right\} \\
&\subseteq \left\{ \sup_{t \in [-h^{-1}, h^{-1}]} |\phi_{X_1}(t) - \hat{\phi}(t)| > (e^{-2\Lambda} - \kappa_n) e^{-\Sigma^2/(2h^2)} \right\} \\
&= G^*.
\end{aligned}
$$

Therefore $\sup_{t \in [-h^{-1}, h^{-1}]} 1_{G_t^c} \le 1_{G^*}$ and then by Chebyshev's inequality we obtain

$$
T_9 \lesssim \frac{1}{h^4} \, \mathrm{P}(G^*) \lesssim \frac{e^{2\Sigma^2/h^2}}{h^4} \mathrm{E} \left[\left(\sup_{t \in [-h^{-1}, h^{-1}]} |\phi_{X_1}(t) - \hat{\phi}(t)| \right)^4 \right]. \tag{26}
$$

Next apply (13) with $k = 0$ and $r = 4$ to the expectation in the rightmost inequality to conclude that $\sup_{\mathcal{T}} T_9$ is negligible in comparison to $(\log n)^{-\beta}$. This shows that also $\sup_{\mathcal{T}} T_6$ is negligible in comparison to $(\log n)^{-\beta}$. To complete bounding T_5 and eventually T_4, we need to bound T_7. By the c_2-inequality

$$
\begin{aligned}
T_7 &\lesssim \mathrm{E} \left[\left(\sup_{t \in [-h^{-1}, h^{-1}]} \left(\left| \frac{(\phi_{X_1}'(t))^2}{(\phi_{X_1}(t))^2} \right| 1_{G_t^c} \right) \right)^2 \right] \\
&+ \mathrm{E} \left[\left(\sup_{t \in [-h^{-1}, h^{-1}]} \left| \frac{(\hat{\phi}'(t))^2}{(\hat{\phi}(t))^2} 1_{G_t} - \frac{(\phi_{X_1}'(t))^2}{(\phi_{X_1}(t))^2} 1_{G_t} \right| \right)^2 \right] \\
&= T_{14} + T_{15}.
\end{aligned}
$$

Observe that for $h \to 0$ we have

$$
\sup_{\mathcal{T}} \sup_{t \in [-h^{-1}, h^{-1}]} \left| \frac{(\phi_{X_1}'(t))^2}{(\phi_{X_1}(t))^2} \right| \lesssim \frac{1}{h^2},
$$

which can be shown by the same arguments that led to (25). We also have $T_{14} \le h^{-4} \mathrm{P}(G^*)$ by the above display. It then follows from (26) that $\sup_{\mathcal{T}} T_{14}$ is negligible in comparison to $(\log n)^{-\beta}$. We turn to T_{15}. By the

23

c_2-inequality

$$T_{15} \lesssim \mathrm{E}\left[\left(\sup_{t\in[-h^{-1},h^{-1}]}\left|\frac{(\hat{\phi}'(t))^2}{(\hat{\phi}(t))^2}1_{G_t} - \frac{(\hat{\phi}'(t))^2}{(\phi_{X_1}(t))^2}1_{G_t}\right|\right)^2\right]$$

$$+ \mathrm{E}\left[\left(\sup_{t\in[-h^{-1},h^{-1}]}\left|\frac{(\hat{\phi}'(t))^2}{(\phi_{X_1}(t))^2}1_{G_t} - \frac{(\phi'_{X_1}(t))^2}{(\phi_{X_1}(t))^2}1_{G_t}\right|\right)^2\right]$$

$$= T_{16} + T_{17}.$$

Notice that by the Cauchy-Schwarz inequality

$$T_{16} \leq \mathrm{E}\left[\left(\sup_{t\in[-h^{-1},h^{-1}]}|(\hat{\phi}'(t))^2| \sup_{t\in[-h^{-1},h^{-1}]}\left(\left|\frac{1}{(\hat{\phi}(t))^2} - \frac{1}{(\phi_{X_1}(t))^2}\right|1_{G_t}\right)\right)^2\right]$$

$$= \mathrm{E}\left[\left(\sup_{t\in[-h^{-1},h^{-1}]}|(\hat{\phi}'(t))^2| \sup_{t\in[-h^{-1},h^{-1}]}\left(\frac{|(\phi_{X_1}(t))^2 - (\hat{\phi}(t))^2|}{|\hat{\phi}(t)|^2|\phi_{X_1}(t)|^2}1_{G_t}\right)\right)^2\right]$$

$$\leq \left(\mathrm{E}\left[\left(\sup_{t\in[-h^{-1},h^{-1}]}|(\hat{\phi}'(t))^2|\right)^4\right]\right)^{1/2}$$

$$\times \left(\mathrm{E}\left[\left(\sup_{t\in[-h^{-1},h^{-1}]}\left(\frac{|(\phi_{X_1}(t))^2 - (\hat{\phi}(t))^2|}{|\hat{\phi}(t)|^2|\phi_{X_1}(t)|^2}1_{G_t}\right)\right)^4\right]\right)^{1/2}$$

$$= \sqrt{T_{18}}\sqrt{T_{19}}.$$

Since $|\hat{\phi}'(t)| \leq n^{-1}\sum_{j=1}^{n}|Z_j|$, it follows that the term T_{18} is bounded by $\mathrm{E}[(n^{-1}\sum_{j=1}^{n}|Z_j|)^8]$. By the c_8-inequality we then get

$$\mathrm{E}\left[\left(\frac{1}{n}\sum_{j=1}^{n}|Z_j|\right)^8\right] \lesssim \frac{1}{n^8}\mathrm{E}\left[\left|\sum_{j=1}^{n}(|Z_j| - \mathrm{E}\left[|Z_j|\right])\right|^8\right] + (\mathrm{E}\left[|Z_j|\right])^8.$$

Hence $\sup_{\mathcal{T}} T_{18}$ is bounded by a constant, which can be proved by the same argument as we used for $\sup_{\mathcal{T}} T_{12}$. Finally, we consider T_{19}. We have

$$T_{19} \lesssim \frac{e^{4\Sigma^2/h^2}}{k_n^8}\mathrm{E}\left[\left(\sup_{t\in[-h^{-1},h^{-1}]}|\hat{\phi}(t) - \phi_{X_1}(t)|\right)^4\right],$$

because

$$|(\phi_{X_1}(t))^2 - (\hat{\phi}(t))^2| \leq 2|\phi_{X_1}(t) - \hat{\phi}(t)|,$$

because $|\phi_{X_1}(t)|$ is bounded from below by $e^{-2\Lambda - \Sigma^2/(2h^2)}$ for $t \in [-h^{-1}, h^{-1}]$, and because of the definition of G_t. Using (13), we conclude that $\sup_{\mathcal{T}} T_{19}$

is negligible in comparison to $(\log n)^{-\beta}$. Hence so is $\sup_{\mathcal{T}} T_{16}$. It remains to study T_{17}. Since

$$T_{17} \lesssim e^{2\Sigma^2/h^2} \mathrm{E} \left[\left(\sup_{t \in [-h^{-1}, h^{-1}]} |\hat{\phi}'(t) - \phi'_{X_1}(t)| \right)^2 \right],$$

it follows from (13) and Condition 2.4 that $\sup_{\mathcal{T}} T_{17}$ is negligible in comparison to $(\log n)^{-\beta}$. Consequently, so are $\sup_{\mathcal{T}} T_{15}$ and $\sup_{\mathcal{T}} T_7$. Combination of all the above results completes the proof of the theorem. □

Proof of Theorem 2.4 . The statement of the theorem is for estimators based on observations $X_1, \ldots, X_n$, but the relationship $Z_j = X_j - X_{j-1}$ and the stationary independent increments property of a Lévy process allows us to work with $Z_1, \ldots, Z_n$ instead. We adapt the proof of Theorem 4.1 in [38] to the present case. A general idea of the proof is as follows: we will consider two Lévy triplets $T_1 = (0, \sigma^2, \rho_1)$ and $T_2 = (0, \sigma^2, \rho_2)$ depending on n and such that the Lévy densities ρ_1 and ρ_2 are separated as much as possible at a point x, while at the same time the corresponding product densities $q_1^{\otimes n}$ and $q_2^{\otimes n}$ of observations $Z_1, \ldots, Z_n$ are close in the χ^2-divergence and hence cannot be distinguished well using the observations $Z_1, \ldots, Z_n$. Up to a constant, the squared distance between $\rho_1(x)$ and $\rho_2(x)$ will then give the desired lower bound (15) for estimation of a Lévy density ρ at a fixed point x. This is a standard technique and we refer to Chapter 2 of [54] for a good exposition of methods for deriving lower bounds in nonparametric curve estimation.

Consider two Lévy triplets $T_1 = (0, \sigma^2, \rho_1)$ and $T_2 = (0, \sigma^2, \rho_2)$, where $\rho_j(u) = \nu(\mathbb{R}) f_j(u)$ for $j = 1, 2$ and constants $0 < \nu(\mathbb{R}) < \Lambda$ and $0 < \sigma^2 < \Sigma^2$. Let

$$f_1(u) = \frac{1}{2}(r_1(u) + r_2(u)),$$

where two densities r_1 and r_2 are defined through their characteristic functions as follows:

$$r_1(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itu} \frac{1}{(1 + t^2/\beta_1^2)^{(\beta_2+1)/2}} dt,$$

$$r_2(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itu} e^{-\alpha_1|t|^{\alpha_2}} dt.$$

With a proper selection of $\beta_1, \beta_2, \alpha_1$ and α_2 one can achieve that f_1 satisfies (14) with constants $L/2$, $L'/2$ and $L''/2$ instead of L, L' and L''. We also assume that $1 < \alpha_2 < 2$. Next define f_2 by

$$f_2(u) = f_1(u) + \delta_n^\beta H((u - x)/\delta_n),$$

where $\delta_n \to 0$ as $n \to \infty$, and the function H satisfies the following conditions:

1. $H(0) > 0$;

2. $\phi_H(t)$ is twice continuously differentiable;

3. $\int_{-\infty}^{\infty} |t|^{\beta} |\phi_H(t)| dt \le L/2, \quad |\phi_H(t)| \le L'/(2|t|^{\beta}), \quad |\phi_H'(t)| \le L''/(2|t|^{\beta})$;

4. $\int_{-\infty}^{\infty} H(x) dx = 0$;

5. $\int_{-\infty}^{0} H(x) dx \ne 0$;

6. $\phi_H(t) = 0$ for t outside $[1, 2]$.

Since $f_1(u)$ decays as $r_2(u)$ at infinity, and consequently as $|u|^{-1-\alpha_2}$, see formula (14.37) in [52], with a proper selection of H, e.g. by the reasoning similar to the one on p. 1268 in [35], the function f_2 will be nonnegative, at least for all small enough δ_n. Consequently, f_2 will be a probability density and one can also achieve that it satisfies (14) for all small enough δ_n.

Now notice that

$$|\rho_2(x) - \rho_1(x)|^2 \asymp \delta_n^{2\beta}. \tag{27}$$

The statement of the theorem will follow from (27) and Lemma 8 of [19], if we prove that for $\delta_n \asymp (\log n)^{-1/2}$ we have

$$n\chi^2(q_2, q_1) = n \int_{-\infty}^{\infty} \frac{(q_2(u) - q_1(u))^2}{q_1(u)} du \le c, \tag{28}$$

where a positive constant $c < 1$ is independent of n. Here $\chi^2(\cdot, \cdot)$ denotes the χ^2-divergence, see p. 86 in [54] for the definition.

Denote by p_i a density of a Poisson sum $Y = \sum_{j=1}^{N(\nu(\mathbb{R}))} W_j$ conditional on the fact that its number of summands $N(\nu(\mathbb{R})) > 0$. Here W_j are i.i.d. with $W_1 \sim f_i$. Now rewrite the characteristic function of Y as

$$\phi_Y(t) = e^{-\nu(\mathbb{R})} + (1 - e^{-\nu(\mathbb{R})}) \frac{1}{e^{\nu(\mathbb{R})} - 1} \left(e^{\nu(\mathbb{R})\phi_{f_i}(t)} - 1 \right), \tag{29}$$

to see that

$$\phi_{p_i}(t) = \frac{1}{e^{\nu(\mathbb{R})} - 1} \left(e^{\nu(\mathbb{R})\phi_{f_i}(t)} - 1 \right).$$

Furthermore,

$$p_i(u) = \sum_{n=1}^{\infty} f_i^{*n}(u) P(N(\nu(\mathbb{R})) = n | N(\nu(\mathbb{R})) > 0). \tag{30}$$

By convolving the law of Y with a normal density ϕ_{0,σ^2} with mean zero and variance σ^2 and using (29), we obtain that

$$q_1(u) \ge (1 - e^{-\nu(\mathbb{R})}) \phi_{0,\sigma^2} * p_1(u).$$

Since by Lemma 2 of [17] there exists a large enough constant A, such that the right-hand side of the above display is not less than $(1-e^{-\nu(\mathbb{R})})p_1(|u|+A)$, we have

$$n\chi^2(q_2, q_1) \lesssim n\int_{-\infty}^{\infty} \frac{(q_2(u) - q_1(u))^2}{p_1(|u| + A)}dx \lesssim n\int_{-\infty}^{\infty} \frac{(q_2(u) - q_1(u))^2}{f_1(|u| + A)}dx.$$

The last inequality is true because by (30) it holds that $p_1(|u| + A) \gtrsim f_1(|u|+A)$. Splitting the integration region in the rightmost term of the last display into two parts, we get that

$$n\chi^2(q_2, q_1) \lesssim n\int_{|u|\leq A} (q_2(u) - q_1(u))^2 du + n\int_{|u|>A} u^4(q_2(u) - q_1(u))^2 dx$$
$$= T_1 + T_2.$$

Here we used the facts that $f_1(u)$ decays as $|u|^{-1-\alpha_2}$ at infinity and that $1 < \alpha_2 < 2$. Parseval's identity then gives

$$T_1 \leq n\frac{1}{2\pi}\int_{-\infty}^{\infty} |\phi_{q_2}(t) - \phi_{q_1}(t)|^2 dt$$
$$= n\frac{(1 - e^{-\nu(\mathbb{R})})^2}{2\pi}\int_{-\infty}^{\infty} |\phi_{p_2}(t) - \phi_{p_1}(t)|^2 e^{-\sigma^2 t^2} dt$$
$$= n\frac{(1 - e^{-\nu(\mathbb{R})})^2}{(e^{\nu(\mathbb{R})} - 1)^2}\frac{1}{2\pi}\int_{-\infty}^{\infty} |e^{\nu(\mathbb{R})\phi_{f_2}(t)} - e^{\nu(\mathbb{R})\phi_{f_1}(t)}|^2 e^{-\sigma^2 t^2} dt$$
$$\lesssim n\int_{-\infty}^{\infty} |\phi_{f_2}(t) - \phi_{f_1}(t)|^2 e^{-\sigma^2 t^2} dt,$$

where the last inequality is a consequence of the mean-value theorem applied to the function e^x and the fact that $|\nu(\mathbb{R})\phi_{f_i}(t)| \leq \Lambda < \infty$. Now notice that

$$\int_{-\infty}^{\infty} e^{itu}\delta_n^\beta H((u - x)/\delta_n)dx = \delta_n^{\beta+1}e^{itx}\phi_H(\delta_n t).$$

By definition of f_1 and f_2 it follows that

$$T_1 \lesssim n\delta_n^{2\beta+2}\int_{-\infty}^{\infty} |\phi_H(\delta_n t)|^2 e^{-\sigma^2 t^2} dt$$
$$= n\delta_n^{2\beta+1}\int_{-\infty}^{\infty} |\phi_H(s)|^2 e^{-\sigma^2 s^2/\delta_n^2} ds$$
$$= O\left(n\delta_n^{2\beta+1}e^{-\sigma^2/\delta_n^2}\right).$$

Hence a choice $\delta_n \asymp (\log n)^{-1/2}$ with an appropriate constant will imply that $T_1 \to 0$ as $n \to \infty$.

To complete the proof, we need to show that $T_2 \to 0$ under a suitable condition on δ_n. To this end first notice that even though ϕ_{f_1} and ϕ_{f_2} are

not twice differentiable at zero, the difference $\phi_{q_2}(t) - \phi_{q_1}(t)$ still is, because ϕ_H is identically zero outside the interval $[1, 2]$, and hence $\phi_{q_2}(t) - \phi_{q_1}(t)$ is zero for t in a neighbourhood of zero. Then by Parseval's identity we obtain that

$$T_2 \leq n \frac{1}{2\pi} \int_{-\infty}^{\infty} |(\phi_{q_2}(t) - \phi_{q_1}(t))''|^2 dt.$$

By the same arguments as we used for T_1, one can show that $T_2 \to 0$ as $n \to \infty$, provided $\delta_n \asymp (\log n)^{-1/2}$ with an appropriate constant. This entails the statement of the theorem. $\square$

The following technical lemma is used in the proof of Theorem 2.5.

Lemma 3.1. *Let the sets B_n and B_n^c be defined as*

$$
\begin{aligned}
B_n &= \left\{ \sup_{t \in [-\sqrt{2}h^{-1}, \sqrt{2}h^{-1}]} \left| \hat{\phi}(t) - \phi_{X_1}(t) \right| > \delta \right\}, \\
B_n^c &= \left\{ \sup_{t \in [-\sqrt{2}h^{-1}, \sqrt{2}h^{-1}]} \left| \hat{\phi}(t) - \phi_{X_1}(t) \right| \leq \delta \right\},
\end{aligned}
\tag{31}
$$

where $\delta = (1/4)e^{-2\Lambda - \Sigma^2/h^2}$. Suppose that $\nu(\mathbb{R}) \leq \Lambda < \infty$ and that Conditions 2.2, 2.3, 2.8 and 2.10 hold. Then there exists a universal n_0 not depending on the Lévy triplet (γ, σ, ρ), such that for all $n \geq n_0$ on the set B_n^c we have

$$\max\{\min\{M_n, \log(|\hat{\phi}(t)|)\}, -M_n\} = \log(|\hat{\phi}(t)|)$$

for t restricted to the interval $[-\sqrt{2}h^{-1}, \sqrt{2}h^{-1}]$.

Proof. The proof is similar to the proof of Lemma 5.1 in [38]. On the set B_n^c and for t restricted to the interval $[-\sqrt{2}h^{-1}, \sqrt{2}h^{-1}]$ we have

$$\left| \left| \frac{\hat{\phi}(t)}{\phi_{X_1}(t)} \right| - 1 \right| \leq \left| \frac{\hat{\phi}(t)}{\phi_{X_1}(t)} - 1 \right| < \frac{1}{2}. \tag{32}$$

Furthermore, on the same set and for $t \in [-\sqrt{2}h^{-1}, \sqrt{2}h^{-1}]$ the inequality

$$
\begin{aligned}
|\log(|\hat{\phi}(t)|)| &\leq |\log(|\phi_{X_1}(t)|)| + \left| \log\left(\left| \frac{\hat{\phi}(t)}{\phi_{X_1}(t)} \right| \right) \right| \\
&\leq |\log(|\phi_{X_1}(t)|)| + \left| \frac{\hat{\phi}(t)}{\phi_{X_1}(t)} - 1 \right| + \left| \frac{\hat{\phi}(t)}{\phi_{X_1}(t)} - 1 \right|^2 \\
&\leq |\log(|\phi_{X_1}(t)|)| + \frac{3}{4} \\
&\leq 2\Lambda + \frac{\Sigma^2}{h^2} + \frac{3}{4}
\end{aligned}
$$

28

holds. Here in the second line we used an elementary inequality $|\log(1 + z) - z| \leq |z|^2$ valid for $|z| < 1/2$, the third line follows from (32), while in the last line we used the bound

$$|\log|\phi_X(t)|| \leq 2\Lambda + \Sigma^2/h^2$$

which holds for $t \in [-\sqrt{2}h^{-1}, \sqrt{2}h^{-1}]$. The result is now immediate from Conditions 2.4 and 2.8, because on the set B_n^c an upper bound on $|\log(|\hat{\phi}(t)|)|$ grows slower than M_n. $\qquad\square$

Proof of Theorem 2.5. A general line of the proof is similar to that of Theorem 2.1 in [34], although the details and actual computations are different. We have

$$\mathrm{E}\left[(\hat{\sigma}_n^2 - \sigma^2)^2\right] = \mathrm{E}\left[(\hat{\sigma}_n^2 - \sigma^2)^2 1_{B_n}\right] + \mathrm{E}\left[(\hat{\sigma}_n^2 - \sigma^2)^2 1_{B_n^c}\right] = S_1 + S_2,$$

where the two sets B_n and B_n^c are defined in (31) and δ in their definition is given by $\delta = (1/4)e^{-2\Lambda - \Sigma^2/h^2}$. The term S_1 in the above display can be bounded as follows,

$$
\begin{aligned}
S_1 &\lesssim \left(M_n^2 \left(\int_{\mathbb{R}} |v_h(t)| dt \right)^2 + \Sigma^4 \right) \mathrm{P}(B_n) \\
&\lesssim \left(M_n^2 \left(\int_{\mathbb{R}} |v_h(t)| dt \right)^2 + \Sigma^4 \right) \frac{e^{2\Sigma^2/h^2}}{nh^2} \\
&= \left(M_n^2 h^4 \left(\int_{\mathbb{R}} |v(t)| dt \right)^2 + \Sigma^4 \right) \frac{e^{2\Sigma^2/h^2}}{nh^2} \\
&\lesssim m_n^2 \frac{e^{2\Sigma^2/h^2}}{nh^2},
\end{aligned}
$$

where we used Chebyshev's inequality and Theorem 2.2 with $r = 2$ to see the second line. Next we consider S_2. By Lemma 3.1 on the set B_n^c for all large enough n truncation in the definition of $\hat{\sigma}_n^2$ becomes unimportant and we have

$$
\begin{aligned}
S_2 &= \mathrm{E}\left[\left(\int_{\mathbb{R}} \log(|\hat{\phi}(t)|) v_h(t) dt - \sigma^2 \right)^2 1_{B_n^c} \right] \\
&= \mathrm{E}\left[\left(\int_{\mathbb{R}} \log\left(\left| \frac{\hat{\phi}(t)}{\phi_{X_1}(t)} \right| \right) v_h(t) dt + \int_{\mathbb{R}} \log(|\phi_{X_1}(t)|) v_h(t) dt - \sigma^2 \right)^2 1_{B_n^c} \right].
\end{aligned}
$$

Hence by equation (4) in [38], the c_2-inequality and Conditions 2.9 and 2.11

29

we obtain that

$$S_2 \lesssim \Lambda^2 \left(\int_{\mathbb{R}} \Re(\phi_f(t)) v_h(t) dt \right)^2$$

$$+ \mathrm{E} \left[\left(\int_{\mathbb{R}} \log \left(\left| \frac{\hat{\phi}(t)}{\phi_X(t)} \right| \right) v_h(t) dt \right)^2 1_{B_n^c} \right]$$

$$= S_3 + S_4.$$

To bound S_3, we proceed as follows,

$$S_3 \lesssim h^6 \int_{\mathbb{R}} |\phi_f(t)|^2 e^{2\alpha|t|^s} dt \int_{\mathbb{R} \setminus [-h^{-1}, h^{-1}]} e^{-2\alpha|t|^s} dt$$

$$\lesssim h^6 \int_{1/h}^{\infty} e^{-2\alpha t^s} dt$$

$$\lesssim h^{s+5} e^{-2\alpha/h^s},$$

where we used the Cauchy-Schwarz inequality, the fact that $|\Re(\phi_f(t))| \leq |\phi_f(t)|$ and Condition 2.9. As far as S_4 is concerned, we have

$$S_4 \lesssim \mathrm{E} \left[\left(\int_{\mathbb{R}} \left| \frac{\hat{\phi}(t)}{\phi_{X_1}(t)} - 1 \right| |v_h(t)| dt \right)^2 1_{B_n^c} \right]$$

$$+ \mathrm{E} \left[\left(\int_{\mathbb{R}} \left\{ \log \left(\left| \frac{\hat{\phi}(t)}{\phi_{X_1}(t)} \right| \right) - \left(\left| \frac{\hat{\phi}(t)}{\phi_{X_1}(t)} \right| - 1 \right) \right\} v_h(t) dt \right)^2 1_{B_n^c} \right]$$

$$= S_5 + S_6.$$

An application of the Cauchy-Schwarz inequality and Conditions 2.2 and 2.9 give

$$S_5 \lesssim e^{4\Lambda + 2\Sigma^2/h^2} \int_{\mathbb{R}} (v_h(t))^2 dt \, \mathrm{E} \left[\int_{-\sqrt{2}/h}^{\sqrt{2}/h} |\hat{\phi}(t) - \phi_{X_1}(t)|^2 dt \right], \qquad (33)$$

where we also used the fact that on the set B_n^c the inequality (32) holds. Parseval's identity and Proposition 1.7 of [54] (notice that in the latter it is actually not necessary to have a positive kernel) applied to the sinc kernel then yield

$$\mathrm{E} \left[\int_{-\sqrt{2}/h}^{\sqrt{2}/h} |\hat{\phi}(t) - \phi_{X_1}(t)|^2 dt \right] \lesssim \frac{1}{nh},$$

from which and from (33) we obtain

$$S_5 \lesssim e^{2\Sigma^2/h^2} h^4 \frac{1}{n}.$$

30

Using the fact that on the set B_n^c the inequality (32) holds and combining it with an inequality $|\log(1+z) - z| \leq |z|^2$ valid for $|z| < 1/2$, one sees that $S_6 \lesssim S_5$. Furthermore, by a standard argument under Condition 2.10 the term S_3 dominates other terms. For instance, we have

$$e^{2\Sigma^2/h^2} h^4 \frac{1}{n} h^{-s-5} e^{2\alpha/h^s} \to 0,$$

because

$$\frac{2\Sigma^2}{h^2} + \frac{2\alpha}{h^s} - \log n + \log h^{-s-1} = -(\log\log n)^2 - (s+1)\log h \to -\infty.$$

This follows from (16) and the fact that under Condition 2.10 it holds that $h \asymp (\log n)^{-1/2}$. The latter can be shown as formula (27) in [18]. Hence S_3 dominates S_5. Combination of all the above bounds on the terms S_i completes the proof. $\square$

Proof of Theorem 2.6. A general line of the proof is the same as in the proof of Theorem 2.3. With the same notation for the individual terms T_i as in the latter, by the the same argument as for the term T_1 in the proof of Theorem 2.3 and term S_3 in the proof of Theorem 2.5 we have

$$
\begin{aligned}
T_1 &\lesssim \left(\int_{\mathbb{R} \setminus [-h^{-1}, h^{-1}]} |\phi_f(t)| dt \right)^2 + h^{s-1} e^{-2\alpha/h^s} \\
&= \left(\int_{\mathbb{R} \setminus [-h^{-1}, h^{-1}]} e^{-\alpha|t|^s} e^{\alpha|t|^s} |\phi_f(t)| dt \right)^2 + h^{s-1} e^{-2\alpha/h^s} \\
&\lesssim \int_{\mathbb{R} \setminus [-h^{-1}, h^{-1}]} e^{-2\alpha|t|^s} dt \int_{\mathbb{R} \setminus [-h^{-1}, h^{-1}]} e^{2\alpha|t|^s} |\phi_f(t)|^2 dt + h^{s-1} e^{-2\alpha/h^s} \\
&\lesssim \int_{\mathbb{R} \setminus [-h^{-1}, h^{-1}]} e^{-2\alpha|t|^s} dt + h^{s-1} e^{-2\alpha/h^s} \\
&\lesssim \int_{1/h}^{\infty} e^{-2\alpha t^s} dt + h^{s-1} e^{-2\alpha/h^s} \\
&\lesssim h^{s-1} e^{-2\alpha/h^s}.
\end{aligned}
$$

Denote by $\mathrm{MSE}[\hat{\sigma}^2]$ the mean square error of $\hat{\sigma}^2$. From the proof of Theorem 2.3 and by Theorem 2.5 we have

$$
\begin{aligned}
T_2 &\lesssim \frac{1}{h^2} \mathrm{MSE}[\hat{\sigma}^2] + T_4 \\
&\lesssim h^{s-1} e^{-2\alpha/h^s} + \frac{e^{2\Sigma^2/h^2}}{nh^8}.
\end{aligned}
$$

We then obtain

$$T_1 + T_2 \lesssim h^{s-1} e^{-2\alpha/h^s} + \frac{e^{2\Sigma^2/h^2}}{nh^8} \lesssim h^{s-1} e^{-2\alpha/h^s},$$

31

because
$$\frac{e^{2\Sigma^2/h^2}}{nh^8}h^{1-s}e^{2\alpha/h^s} \to 0,$$

which can be seen by taking the logarithm of the lefthand side and then using Condition 2.9 and the fact that $h \asymp (\log n)^{-1/2}$, cf. formula (27) in [18], to conclude that the lefthand side in the above display diverges to minus infinity. This entails the first statement of the theorem.

Before proving the second statement of the theorem, we will show that the choice of h as in Condition 2.10 is optimal in a sense that it asymptotically minimises the order bound on the mean square error of $\hat{\rho}$. This follows in essence by arguments similar to those used in the proof of Lemma 4 in [18]: a minimiser h_* with respect to h of the expression

$$h^{s-1}e^{-2\alpha/h^s} + \frac{e^{2\Sigma^2/h^2}}{nh^8},$$

which up to a constant is an upper bound on the risk of the estimator $\hat{\rho}$, can be found from the equation

$$\frac{d}{dh}\left[h^{s-1}e^{-2\alpha/h^s} + \frac{e^{2\Sigma^2/h^2}}{nh^8} \right] = 0.$$

After neglecting lower order terms (here we assume that $h \to 0$ as $n \to \infty$), one can deduce that h_* has to satisfy

$$\frac{2\alpha s}{4\Sigma^2}nh_*^9(1+o(1)) = e^{2\Sigma^2/h_*^2+2\alpha/h_*^s}. \tag{34}$$

Taking logarithm of the both sides of (34) yields that h_* satisfies

$$a\log h_* + \frac{2\alpha}{h_*^s} + \frac{2\Sigma^2}{h_*^2} = \log n + C(1+o(1)) \tag{35}$$

for some constants a and C, cf. equation (11) in [18]. With h_* chosen as in (34) or (35), the term $h_*^{s-1}e^{-2\alpha/h_*^s}$ dominates the term $e^{2\Sigma^2/h^2}/(nh_*^8)$, cf. pp. 30–31 in [18] for a similar result for the kernel-type deconvolution density estimator in a particular deconvolution problem. Indeed, for h_* satisfying (34) we have

$$h_*^{s-1}e^{-2\alpha/h_*^s} \asymp \frac{e^{2\Sigma^2/h_*^2}}{nh_*^8}h_*^{s-2},$$

and it suffices to observe that $s < 2$ by our assumptions. Let $\tilde{h}$ be as in (16). For any $b \in \mathbb{R}$ the formula

$$h_*^b e^{-2\alpha/h_*^s} = \tilde{h}^b e^{-2\alpha/\tilde{h}^s}(1+o(1))$$

holds, which can be proved exactly as formula (28) of [18]. Furthermore,

$$\frac{e^{2\Sigma^2/\tilde{h}^2}}{n\tilde{h}^8} = o\left(e^{-2\alpha/\tilde{h}^s}\right),$$

which is a direct consequence of (16) and the fact that $\tilde{h} \asymp (\log n)^{-1/2}$, cf. formula (27) of [18]. Finally,

$$\frac{e^{2\Sigma^2/\tilde{h}^2}}{n\tilde{h}^8} \leq \frac{e^{2\Sigma^2/h_*^2}}{nh_*^8},$$

for n large enough, which can be shown as formula (30) of [18]. These facts together imply that h as in (16) defines an optimal bandwidth, for an upper bound on the risk of $\hat{\rho}(x)$ computed with such an h is of the same order as the one computed with h_*. Combination of the above results proves the first statement of the theorem.

To complete the proof of the theorem, it remains to prove (17). Assuming n is large enough, by (16) it holds in the case $s = 1$ that

$$\frac{1}{h} = \frac{-\alpha + \sqrt{2\Sigma^2(\log n - (\log\log n)^2) + \alpha^2}}{2\Sigma^2}.$$

From this is follows that

$$\exp\left(-\frac{2\alpha}{h}\right) \lesssim \exp\left(-2\alpha\sqrt{\frac{1}{2\Sigma^2}(\log n - (\log\log n)^2)}\right).$$

The righthand side is of order $\exp(-2\alpha\sqrt{(\log n)/(2\Sigma^2)})$, as can be seen by some straightforward manipulations: we have

$$\exp\left(-2\alpha\sqrt{\frac{1}{2\Sigma^2}(\log n - (\log\log n)^2)}\right) = \exp\left(-2\alpha\sqrt{\frac{1}{2\Sigma^2}\log n}\right) \quad (36)$$
$$\times \exp\left(2\alpha\sqrt{\frac{1}{2\Sigma^2}\log n} - 2\alpha\sqrt{\frac{1}{2\Sigma^2}(\log n - (\log\log n)^2)}\right)$$

and

$$\sqrt{\frac{1}{2\Sigma^2}\log n} - \sqrt{\frac{1}{2\Sigma^2}(\log n - (\log\log n)^2)} \to 0,$$

because the lefthand side of the latter can be rewritten as

$$\sqrt{\frac{1}{2\Sigma^2}}\frac{(\log\log n)^2}{\sqrt{\log n}}\left[\left(1 - \sqrt{1 - \frac{(\log\log n)^2}{\log n}}\right)\frac{\log n}{(\log\log n)^2}\right].$$

The term in the square brackets converges to $-1/2$, because it converges to a derivative of the function $\sqrt{1-t}$ at $t = 0$, while for the first factor we have

$$\sqrt{\frac{1}{\sqrt{2\Sigma^2}}}\frac{(\log\log n)^2}{\sqrt{\log n}} \to 0.$$

33

Consequently, the righthand side of (36) is of order $\exp(-2\alpha\sqrt{(\log n)/(2\Sigma^2)})$. This concludes the proof of the theorem. $\square$

Acknowledgments

References

[1] Y. Aït-Sahalia and J. Jacod. Volatility estimators for discretely sampled Lévy processes. *Ann. Statist.*, 35:355–392, 2007.

[2] M.G. Akritas and R.A. Johnson. Asymptotic inference in Lévy processes of the discontinuous type. *Ann. Statist.*, 9:604–614, 1981.

[3] M.G. Akritas. Asymptotic theory for estimating the parameters of a Lévy process. *Ann. Inst. Statist. Math.*, 34:259–280, 1982.

[4] I.V. Basawa and P.J. Brockwell. Inference for gamma and stable processes. *Biometrika*, 65:129–133, 1978.

[5] I.V. Basawa and P.J. Brockwell. A note on estimation for gamma and stable processes. *Biometrika*, 67:234–236, 1980.

[6] D. Belomestny and M. Reiß. Spectral calibration of exponential Lévy models. *Finance Stoch.*, 10:449-474, 2006.

[7] D. Belomestny and M. Reiß. Spectral calibration of exponential Lévy models [2]. SFB 649 Discussion Paper 2006-035, 2006.

[8] J. Bertoin. *Lévy Processes*. Cambridge University Press, Cambridge, 1996.

[9] B.M. Bibby and M. Sørensen. A hyperbolic diffusion model for stock prices. *Finance Stoch.*, 1:25–41, 1997.

[10] P. Blæsild and M. Sørensen. HYP — a computer program for analyzing data by means of the hyperbolic distribution. Research report No. 248, Department of Mathematical Statistics, Aarhus University, 1992.

[11] S. Borak, W. Härdle and R. Weron. Stable distributions. In P. Cizek, W. Härdle and R. Weron (editors), *Statistical Tools for Finance and Insurance*, 21–44, Springer, Berlin, 2005.

[12] L.D. Brown, M.G. Low and L.H. Zhao. Superefficiency in nonparametric function estimation. *Ann. Statist.*, 25:2607-2625, 1997.

[13] B. Buchmann. Weighted empirical processes in the nonparametric inference for Lévy processes. *Math. Methods Statist.*, 18:281–309, 2009.

[14] B. Buchmann and R. Grübel. Decompounding: an estimation problem for Poisson random sums. *Ann. Statist.*, 31:1054–1074, 2003.

[15] B. Buchmann and R. Grübel. Decompounding Poisson random sums: recursively truncated estimates in the discrete case. *Ann. Inst. Statist. Math.*, 56:743–756, 2004.

[16] E.V. Burnaev. Inversion formula for infinitely divisible distributions. *Russ. Math. Surv.*, 61:772–774, 2006.

[17] C. Butucea and C. Matias. Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model. *Bernoulli*, 11:309–340, 2005.

[18] C. Butucea and A.B. Tsybakov. Sharp optimality for density deconvolution with dominating bias, I. *Theory Probab. Appl.*, 52:24–39, 2008.

[19] C. Butucea and A.B. Tsybakov. Sharp optimality for density deconvolution with dominating bias, II. *Theory Probab. Appl.*, 52:237–249, 2008.

[20] P. Carr, H. Geman, D.B. Madan, and M. Yor. The fine structure of asset returns: an empirical investigation. *J. Bus.*, 75:305–332, 2002.

[21] S.X. Chen, A. Delaigle and P. Hall. Nonparametric estimation for a class of Lévy processes. *J. Econometrics*, 157:257–271, 2010.

[22] K.L. Chung. *A Course in Probability Theory*, 3rd edition. Academic Press, New York, 2001.

[23] F. Comte and V. Genon-Catalot. Nonparametric estimation for pure jump Lévy processes based on high frequency data. *Stochastic Process. Appl.*, 119:4088-4123, 2009.

[24] F. Comte and V. Genon-Catalot. Nonparametric adaptive estimation for pure jump Lévy processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 46:595–617, 2010.

[25] F. Comte and V. Genon-Catalot. Non-parametric estimation for pure jump irregularly sampled or noisy Lévy processes. *Stat. Neerl.*, 64: 290–313, 2010.

[26] F. Comte and V. Genon-Catalot. Estimation for Lévy processes from high frequency data within a long time interval. *Ann. Statist.*, 39: 803–837, 2011.

[27] F. Comte and C. Lacour. Data driven density estimation in presence of unknown convolution operator. To appear in *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 2011.

[28] R. Cont and P. Tankov. *Financial Modelling with Jump Processes.* Chapman & Hall/CRC, Boca Raton, 2003.

[29] R. Cont and P. Tankov. Retrieving Lévy processes from option prices: regularization of an ill-posed inverse problem. *SIAM J. Control Optim.*, 45:1–25, 2006.

[30] A. Delaigle. An alternative view of the deconvolution problem. *Statist. Sinica*, 18:1025–1045, 2008.

[31] L. Devroye. On the non-consistency of an estimate of Chiu. *Stat. Probab. Lett.*, 20:183–188, 1994.

[32] L. Devroye and L. Györfi. *Nonparametric Density Estimation: the L_1 View.* John Wiley & Sons, New York, 1985.

[33] B. van Es and S. Gugushvili. Asymptotic normality of the deconvolution kernel density estimator under the vanishing error variance. *J. Korean Statist. Soc.*, 39: 102–115, 2010.

[34] B. van Es, S. Gugushvili and P. Spreij. A kernel type nonparametric density estimator for decompounding. *Bernoulli*, 13:672–694, 2007.

[35] J. Fan. On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, 19:1257–1272, 1991.

[36] J. Fan. Deconvolution with supersmooth distributions. *Canad. J. Statist.*, 20:155-169, 1992.

[37] E. Figueroa-López. Sieve-based confidence intervals and bands for Lévy densities. *Bernoulli*, 17:643–670, 2011.

[38] S. Gugushvili. Nonparametric estimation of the characteristic triplet of a discretely observed Lévy process. *J. Nonparametr. Stat.*, 21:321–343, 2009.

[39] S. Gugushvili, B. van Es and P. Spreij. Deconvolution for an atomic distribution: rates of convergence. To appear in *J. Nonparametr. Stat.*, 2011.

[40] S. Gugushvili, C. Klaassen and P. Spreij (editors). *Statistical Inference for Lévy Processes with Applications to Finance. Stat. Neerl.*, 64(3), 2010.

[41] G. Jongbloed and F.H. van der Meulen. Parametric estimation for sub-ordinators and induced OU processes. *Scand. J. Statist.*, 33:825–847, 2006.

[42] G. Jongbloed, F.H. van der Meulen and A.W. van der Vaart. Non-parametric inference for Lévy-driven Ornstein-Uhlenbeck processes. *Bernoulli*, 11:759-791, 2005.

[43] J. Kappus and M. Reiß. Estimation of the characteristics of a Lévy process observed at arbitrary frequency. *Stat. Neerl.*, 64: 314–328, 2010.

[44] A.E. Kyprianou. *Introductory Lectures on Fluctuations of Lévy Processes with Applications.* Springer, Berlin, 2006.

[45] A. Meister. Density estimation with normal measurement error with unknown variance. *Statist. Sinica*, 16:195–211, 2006.

[46] R.C. Merton. Option pricing when underlying stock returns are discontinuous. *J. Financ. Econ.*, 3:125–144, 1976.

[47] M.H. Neumann. On the effect of estimating the error density in non-parametric deconvolution. *J. Nonparametr. Statist.*, 7:307–330, 1997.

[48] M.H. Neumann and M. Reiß. Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli*, 15:223–248, 2009.

[49] J.P. Nolan. Maximum likelihood estimation and diagnostics for stable distributions. In O.E. Barndorff-Nielsen, T. Mikosch, and S.I. Resnick (editors), *Lévy Processes: Theory and Applications*, 379–400, Birkhäuser, Boston, 2001.

[50] Y.-F. Ren and H.-Y. Liang. On the best constant in Marcinkiewicz-Zygmund inequality. *Statist. Probab. Lett.*, 53:227-233, 1999.

[51] T.H. Rydberg. The normal inverse Gaussian Lévy process: simulation and approximation. *Stoch. Models*, 13:887–910, 1997.

[52] K.-I. Sato. *Lévy Processes and Infinitely Divisible Distributions.* Cambridge University Press, Cambridge, 2004.

[53] J. Söhl. Polar sets for anisotropic Gaussian random fields. *Statist. Probab. Lett.*, 80:840–847, 2010.

[54] A.B. Tsybakov. *Introduction to Nonparametric Estimation.* Springer, New York, 2009.

[55] A.W. van der Vaart. *Asymptotic Statistics.* Cambridge University Press, Cambridge, 1998.

[56] A.W. van der Vaart and J.A. Wellner. *Weak Convergence and Empirical Processes with Applications to Statistics.* Springer, New York, 1996.

[57] M.P. Wand. Finite sample performance of deconvolving density estimators. *Statist. Probab. Lett.*, 37:131–139, 1998.

[58] R.N. Watteel and R.J. Kulperger. Nonparametric estimation of the canonical measure for infinitely divisible distributions. *J. Stat. Comput. Simul.*, 73:525–542, 2003.

[59] V.M. Zolotarev. *One-Dimensional Stable Distributions.* American Mathematical Society, Providence, 1986.