Adaptive nonparametric estimation for Lévy processes observed at low frequency

Johanna Kappus

Institut für Mathematik, Universität Rostock, 18051 Rostock, Germany
Email: johanna.kappus@uni-rostock.de, Phone: 49 381 498-6604

Abstract
This article deals with adaptive nonparametric estimation for Lévy processes observed at low frequency. For general linear functionals of the Lévy measure, we construct kernel estimators, provide upper risk bounds and derive rates of convergence under regularity assumptions.

Our focus lies on the adaptive choice of the bandwidth, using model selection techniques. We face here a non-standard problem of model selection with unknown variance. A new approach towards this problem is proposed, which also allows a straightforward generalization to a classical density deconvolution framework.

1 Introduction

Lévy processes, continuous time stochastic processes with independent and stationary increments, are the building blocks for a large number of continuous time models with jumps which play an important role, for example, in the modeling of financial data. See Cont and Tankov (2004) for an overview of the topic. The problem of estimating the characteristics of a Lévy process is thus not only a topic of great theoretical relevance, but also an important issue for practitioners and has received considerable attention over the past decade. Starting from the work by Belomestny and Reiß (2006), nonparametric estimation methods for Lévy processes have been considered in a number of articles in the past few years. Let us mention Neumann and Reiß (2009), Gugushvili (2009, 2012), Comte and Genon-Catalot (2010, 2009, 2011), Figueroa-López and Houdré (2006) and Figueroa-López (2009), Nickl and Reiß (2012), and Belomestny (2010). For results on time changed Lévy processes, see Belomestny (2011).

In the present work, we focus on the adaptive estimation of the jump measure for Lévy processes observed at low frequency. The following statistical model is being considered: A Lévy process $X$ having finite variation on compact sets and finite first moments is observed at discrete, equidistant time points. We investigate the nonparametric estimation of linear functionals of the finite signed measure $\mu(dx) = x\nu(dx)$, with $\nu$ denoting the Lévy measure. Kernel estimators are constructed and upper bounds on the corresponding risk are derived.

Our main concern is to provide a strategy for the data driven choice of the smoothing parameter, using techniques of model selection via penalization.

The model selection approach to adaptive estimation has been extensively studied in the literature, starting from the work by Birgé and Massart in the late 90’s, see, for example, Birgé and Massart (1993) and Birgé (1999). The model selection point of view essentially differs from other existing methods, typically in the spirit of Lepski, see e.g. Lepski (1990, 1991), in the sense that the problem is considered from a non-asymptotic perspective. We refer to Birgé and Massart (1993) for a detailed discussion and systematic comparison.
Recently, there is a strong tendency to apply model selection techniques in adaptive estimation problems for Lévy processes, see Figueroa-López and Houdré (2006) and subsequent papers, and Comte and Genon-Catalot (2009, 2010, 2011). However, the above mentioned papers have mainly focused on a situation where continuous time or high frequency observations of the process are available. In the present work, we consider and fully solve the problem of adaptation in a low frequency framework.

It is well known that, depending on the nature of the observations, there are two fundamentally different approaches to the estimation of the Lévy measure: For continuous time or high frequency data, the jumps are directly observable or observable in the limit, so it is possible to use the empirical jump measure as an estimator of the true underlying jump measure. The estimation procedure can then be understood in analogy with a density estimation problem. When low frequency observations of the process are available, one has to exploit the structural properties of infinitely divisible laws and faces a more complicated deconvolution type structure.

In a density deconvolution framework, adaptive estimation by model selection has been considered by Comte, Rosenholc, and Taupin (2006) and by Comte and Lacour (2011). For estimating linear functionals in a density deconvolution model, see Butucea and Comte (2009). However, in those papers, the distribution of the noise is assumed to be known and the adaptive procedure crucially depends on the fact that the variance term is feasible, which is no longer true in the present model. Indeed, the situation which we consider in the present work can be understood in analogy with a deconvolution problem with unknown distribution of the noise.

Rates of convergence for deconvolution problems with unknown error density have been studied, since the late 90’s, for example, in Neumann (1997, 2006), Meister (2006, 2009) or Johannes (2009). However, few literature is available on adaptive estimation in deconvolution problems with unknown distribution of the errors. We are aware of the work by Johannes and Schwarz (2012) in a circular deconvolution model and of Comte and Lacour (2011). We propose here a new approach to dealing with the unknown variance. Although our results are formulated for estimation problems in a Lévy model, they can be generalized to a classical density deconvolution framework.

Compared to the reasoning in Comte and Lacour (2011), we can avoid the loss of a polynomial factor. Moreover, unlike in Comte and Lacour (2011) and Johannes and Schwarz (2012), our reasoning does not rely on certain semiparametric assumptions on the decay of the characteristic function. This allows a fully general treatment.

The main technical step involved in our arguments relies on an application of the Talagrand inequality, which allows to give control on the fluctuation of the empirical characteristic function in the denominator around its target uniformly on the real line, thus improving a classical pointwise result presented in Neumann (1997).

This paper is organized as follows: In Section 2, we start by specifying the statistical model and technical assumptions. In Section 3, kernel estimators are introduced and upper bounds on the corresponding risk are provided. From this, rates of convergence are derived under regularity assumptions. The adaptive estimation procedure and hence the main result of this paper is then presented in Section 4. All proofs are postponed to Section 5.

## 2 Statistical model

A Lévy process $X = \{X_t : t \in \mathbb{R}^+\}$ taking values in $\mathbb{R}$ is observed at equidistant time points $\Delta, \cdots, 2n\Delta$, with $\Delta > 0$ fixed.

We work under the following technical assumptions:

(A1) $X$ is of pure jump type.

(A2) $X$ has moderate activity of small jumps in the sense that the following holds true for
the Lévy measure $\nu$:

$$\int_{|x| \leq 1} |x| \nu(\,dx) < \infty. \quad (2.1)$$

(A3) $X$ has no drift component.

(A4) For one and hence for any $t > 0$, $X_t$ has a finite second moment. This is equivalent to stating that

$$\int |x|^2 \nu(\,dx) < \infty. \quad (2.2)$$

Imposing the assumptions (A1) and (A2) is equivalent to stating that $X$ has finite variation on compact sets.

Under (A1)-(A4), the characteristic function of $X_\Delta$ is given by

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

with characteristic exponent

$$\Psi(u) = \int (e^{iu x} - 1) \nu(\,dx) = \int e^{iu x} \frac{1}{x} \nu(\,dx). \quad (2.3)$$

(A proof can be found, for example, in Chapter 2 in [Sato (1999)]. The process is thus fully described by the signed measure $\mu(\,dx) := x \nu(\,dx)$, which is finite thanks to (2.1) and (2.2).

It is worth mentioning that the conditions (A1)-(A3) can be relaxed, but at the cost of substantially complicating the structure of the estimator and proofs without allowing much further insight into the nature of the problem. See [Neumann and Reiß (2009)] or [Belomestny (2011)] for estimation strategies in a more general framework.

Our goal is to estimate a linear functional of $\mu$. Let $f$ be a distribution and assume that one of the following conditions is met:

(F1) $f$ is regular and can hence be identified with a function. $f \in L^1(\mathbb{R})$ and $\sup_{x \in \mathbb{R}} |f(x)| < \infty$.

(F2) $f$ is compactly supported, with order $k$, and for some open interval $D = (d_1, d_2)$ with $\text{supp}(f) \subseteq D$, the restriction $\mu|_D$ possesses a density $g_D \in C^k(D)$.

Then the parameter of interest is

$$\theta := \langle f, \mu \rangle.$$

For the definition of the order of a distribution and a concise overview of the theory of distributions, we refer to [Rudin (1991)]. For a detailed introduction, see [Jantscher (1971)].

To better understand the meaning of the abstract assumptions (F1) and (F2), let us mention that under (F1), we simply have

$$\theta \equiv \int f(x) \mu(\,dx)$$

and that (F2) covers typical problems such as point estimation or the estimation of derivatives. For notational convenience, we will often write, formally, $\int f(x) \mu(\,dx) := \langle f, \mu \rangle$ even though $f$ may be non-regular.

The special case of testing $\mu$ with smooth functions has been considered in [Neumann and Reiß (2009)] and point estimation has been dealt with in [Belomestny (2011)]. However, to the best of our knowledge, the estimation of arbitrary linear functionals of $\mu$ has not yet been treated in full generality.

Moreover, the problem of adaptive estimation, which is the main concern of the present work, has not been treated in the above mentioned papers.
3 Estimation procedure, risk bounds and rates of convergence

3.1 Construction of the estimator and non-asymptotic risk bounds

Formula (2.3) allows to recover the Fourier transform $\mathcal{F}\mu$ of $\mu$ by derivating the characteristic exponent,

$$\Psi'(u) = i \int e^{iuX} \mu(dx) = i\mathcal{F}\mu(u).$$

In terms of the characteristic function and its derivative,

$$\mathcal{F}\mu(u) = \frac{\sum \varphi_{\Delta}(u)}{i \varphi_{\Delta}(u)}.$$  \hspace{1cm} (3.1)

Recall that the characteristic function of an infinitely divisible law possesses no zeros (see Lemma 7.5 in Sato (1999)), so dividing by $\varphi_{\Delta}$ is not critical.

By the definition of a Lévy process, the increments $Z_{\Delta,k} := X_{k\Delta} - X_{(k-1)\Delta}$, $k = 1, \ldots, 2n$ form i.i.d. copies of $X_{\Delta}$. We can thus define the empirical versions of $\varphi_{\Delta}$ and $\varphi'_{\Delta}$, setting

$$\hat{\varphi}_{\Delta,n}(u) := \frac{1}{n} \sum_{k=1}^{n} e^{iuZ_{\Delta,k}}$$

and

$$\hat{\varphi'}_{\Delta,n}(u) := \frac{1}{n} \sum_{k=n+1}^{2n} iZ_{\Delta,k}e^{iuZ_{\Delta,k}}.$$  

Splitting the observations into two independent samples will be crucial for the adaptive procedure, see the proof of Proposition 5.11 in Section 5.2.

Following the approach in Comte and Genon-Catalot (2010) and Comte and Lacour (2011), we replace the empirical characteristic function in the denominator by its truncated version, setting

$$\tilde{\varphi}_{\Delta,n}(u) := \frac{1}{\varphi_{\Delta,n}(u)}$$

and

$$\tilde{\varphi'}_{\Delta,n}(u) := \frac{1}{\varphi_{\Delta,n}(u)} 1(|\hat{\varphi}_{\Delta,n}(u)| \geq (\Delta n)^{-1/2}).$$

This definition has originally been introduced in Neumann (1997).

Parseval’s identity permits to express the quantity of interest in the Fourier domain,

$$\theta = \frac{1}{2\pi} \int \mathcal{F}f(-u)\mathcal{F}\mu(u) \, du.$$  \hspace{1cm} (3.2)

Recall, at this point, that the Fourier transform of a compactly supported distribution is a function, see Theorem 7.23 in Rudin (1991).

Let $K$ be a kernel function. We use the notation $K_{h}(x) := h^{-1}K(x/h)$. $K$ is chosen such that

(K1) For any $h > 0$, $\mathcal{F}K_{h}\mathcal{F}(-\cdot) \in L^{1}(\mathbb{R})$.

(K2) If $f$ is non-regular, with order $k$, $K$ is $k$-times continuously differentiable.

Formula (3.1) and formula (3.2) suggest to define the kernel estimator

$$\hat{\theta}_{\Delta,h,n} := \frac{1}{2\pi} \int \mathcal{F}f(-u)K_{h}(u) \frac{\sum \varphi_{\Delta,n}(u)}{i \varphi_{\Delta,n}(u)} \, du.$$  

The following upper bound can be derived for the squared risk:
Theorem 3.1. Assume that (A1)-(A4), (F1) or (F2) and (K1)+(K2) are met. Then
\[
E \left[ |\theta - \hat{\theta}_{\Delta,n}|^2 \right] 
\leq 2 \left( \int f(x) \mu(dx) - \int f(x) (K_h * \mu)(dx) \right)^2 
+ \frac{T}{2\pi^2} \left\{ C_1 \int |\mathcal{F}K(hu)|^2 \left| \mathcal{F}f(-u) \right|^2 \varphi_\Delta(u) du \right. 
\left. + C_2 \left( \int |\mathcal{F}K(hu)| \left| \mathcal{F}f(-u) \right| \varphi_\Delta(u) du \right)^2 \right\},
\]
with \( T := \Delta n \) and with constants
\[
C_1 = C \left( \int |\Psi''(x)| dx + 2 \int |\Psi'(x)|^2 dx \right) \leq \infty \tag{3.3}
\]
and
\[
C_2 = C \left( \|\Psi''\|_\infty + 2\|\Psi'\|_\infty^2 \right) < \infty \tag{3.4}
\]
for a universal positive constant \( C \).

For the special case of point estimation, Theorem 3.1 is in accordance with the results found in Belomestny (2011). It should also be compared to the deconvolution framework which is considered in Butucea and Comte (2009).

3.2 Rates of convergence

We investigate here the rates of convergence which can be derived from the upper risk bounds given in Theorem 3.1 under regularity assumptions on \( \mu \) and \( f \) and on the decay of the characteristic function.

Recall the following definitions: For \( a \in \mathbb{R} \) and \( M > 0 \), the Sobolev class \( S(a, M) \) consists of those tempered distributions, for which
\[
\int \left| \mathcal{F}f(-u) \right|^2 \left( 1 + |u|^2 \right)^a du \leq M.
\]
Let \( \langle a \rangle := \max\{k \in \mathbb{N} : k < a\} \). For an open subset \( D \subseteq \mathbb{R} \), and positive constants \( a, L \) and \( R \), the Hölder class \( \mathcal{H}_D(a, L, R) \) consists of those functions \( f \) for which \( \sup_{x \in D} |f(x)| \leq R \) holds and \( f|_D \) is \( \langle a \rangle \)-times continuously differentiable, with
\[
\sup_{x,y \in D, x \neq y} \frac{|f^{(\langle a \rangle)}(x) - f^{(\langle a \rangle)}(y)|}{|x - y|^{a-\langle a \rangle}} \leq L.
\]
A kernel \( K \) is called a \( k \)-th order kernel, if for all integers \( 1 \leq m < k \),
\[
\int x^m K(x) dx = 0 \tag{3.5}
\]
and moreover,
\[
\int |x|^k |K(x)| dx < \infty.
\]
Formula (3.5) is equivalent to stating that \( (\mathcal{F}K)^{(m)}(0) \) vanishes for \( 1 \leq m < k \).

3.2.1 Rate results under global regularity assumptions

We start by providing rate results under global regularity assumptions on the signed measure \( \mu \) and on the test function \( f \), measured in a Sobolev sense. This point of view is appropriate when one is interested in estimating integrals of the form \( \theta = \int f(x) \mu(dx) \) with some function \( f \) which does not vanish near the origin.

Let us introduce the following nonparametric classes of signed measures:
Definition 3.2. For constants $\tilde{C}_1, \tilde{C}_2, C_\varphi, M_\mu > 0$, $a \in \mathbb{R}$ and $c_\varphi, \beta, \rho \geq 0$, let $\mathcal{M}(\tilde{C}_1, \tilde{C}_2, C_\varphi, c_\varphi, \beta, \rho, a, M_\mu)$ be the collection of finite signed measures $\mu$ for which the following holds:

(i) There is a Lévy process $X$ such that the assumptions (A1)-(A4) are satisfied and for the corresponding Lévy measure $\nu$, $\mu(dx) = x\nu(dx)$.

(ii) For the characteristic function $\varphi$ of $X_1$, the following holds:

$$\forall u \in \mathbb{R} : |\varphi(u)| \geq C_\varphi(1 + |u|^2)^{-\frac{\rho}{2}} e^{-c_\varphi|u|^\rho}.$$ 

(iii) For $C_1$ and $C_2$ defined in (3.3) and (3.4), $C_1 \leq \tilde{C}_1$ and $C_2 \leq \tilde{C}_2$.

We denote by $P_\mu$ the distribution of $X_1$ and by $E_\mu$ the expectation with respect to $P_\mu$.

Theorem 3.3. Assume that $f \in S(s, M_f)$ for some $s \in \mathbb{R}$ and some positive constant $M_f$.

Consider the nonparametric class $\mathcal{M} := \mathcal{M}(\tilde{C}_1, \tilde{C}_2, C_\varphi, c_\varphi, \beta, a, M_\mu)$, with $a > -s$. Assume that $\mathcal{F}K$ is supported on $[-\pi, \pi]$ and that either $\mathcal{F}K = 1_{[-\pi, \pi]}$ (sinc kernel) or $K$ has order $a + s$ and $\mathcal{F}K \in \mathcal{H}_2(a + s, L_K, R_K)$ for positive constants $L_K$ and $R_K$.

Then, selecting $h_{\Delta,n}^*$ in an optimal way, we derive that

$$\sup_{\mu \in \mathcal{M}} E_\mu \left[ \theta - \hat{\theta}_{h_{\Delta,n}^*} \right]^2 = O(r_{\Delta,n}) ,$$

with $(r_{\Delta,n})$ denoting the sequences which are summarized in the following table:

| $C_1$ | $C_1 < \infty$ | $C_1 = \infty$ |
|-------|----------------|----------------|
| $\rho = 0$ | $s \geq \Delta \beta$ | $T^{-1}$ | $s \geq \Delta \beta + \frac{1}{2}$ | $T^{-1}$ |
|       | $s = \Delta \beta$ | $T^{-1}$ | $s = \Delta \beta + \frac{1}{2}$ | $(\log T)T^{-1}$ |
|       | $s < \Delta \beta$ | $T^{-\frac{2s+2\rho}{2s+2\rho+1}}$ | $s < \Delta \beta + \frac{1}{2}$ | $n^{-\frac{2s+2\rho}{2s+2\rho+1}}$ |
| $\rho > 0$ | $(\log T)^{-\frac{2s+2\rho}{\rho}}$ | | | |

Discussion: These rates of convergence should be compared to the results which are known in a deconvolution framework, see Butucea and Comte (2009).

However, it is important to keep in mind that there are striking structural differences between the Lévy setting and a density deconvolution problem.

Very much unlike in a deconvolution framework, the parameters $a$, $\beta$ and $\rho$ are by no means independent of each other. A fast decay of the characteristic functions will always indicate a high activity of small jumps, so the jump measure cannot have a globally smooth Lebesgue density, but will be ill-behaved near the origin. Large values of $\beta$ and $\rho$ will hence not only in themselves lead to slow rates of convergence, but also result in small values of $a$. For a detailed discussion, we refer to Orey (1968), Belomestny (2011) and Kappus and Reiß (2014).

3.2.2 Rate results under local regularity assumptions

If $f$ has a compact support which is bounded away from the origin, the point of view of measuring the smoothness of $\mu$ in a global Sobolev sense is inappropriate. This is true, for example, for point estimation and the estimation of derivatives. In this setting, we consider classes of measures which have a Hölder regular density in a neighbourhood of the point or interval of interest.
**Definition 3.4.** For constants $C_1, C_2, C_3, L, R, a > 0$ and $c_\varphi, \beta, \rho \geq 0$ and a bounded open interval $D = (d_1, d_2)$, let $\mathcal{M}(C_1, C_2, C_3, c_\varphi, \beta, \rho, a, D, L, R)$ be the collection of finite signed measures $\mu$, for which the following holds: The items (i)-(iii) from Definition 3.2 are true and

(iii) The restriction $\mu|_D$ possesses a Lebesgue density $g_D \in \mathcal{H}_D(a, L, R)$.

**Theorem 3.5.** Assume that $f$ is compactly supported, with $\text{supp}(f) =: [a, b] \subset \mathbb{R} \setminus \{0\}$. Assume, moreover, that for some $s \in \mathbb{Z}$ and some positive integer $C_f$,

$$\forall u \in \mathbb{R} : |\mathcal{F}f(-u)| \leq C_f(1 + |u|^2)^{-s}.$$

Assume that the order of $K$ is $a + s$, $K$ is $-s \vee 0$-times continuously differentiable and there is a constant $C_K > 0$ such that for any nonnegative integer $m \leq (0 \vee -s)$,

$$\forall z \in \mathbb{R} : \left|K^{(m)}(z)\right| \leq C_K(1 + |z|)^{-(a+s)-m-1}.$$

Consider the nonparametric class $\mathcal{M} = \mathcal{M}(C_1, C_2, C_3, c_\varphi, \beta, \rho, a, D, L, R)$ with $D = (d_1, d_2) \supseteq [a, b]$ and $a > -s$. Then, selecting $h_{\Delta,n}^*$ in an optimal way, we derive that

$$\sup_{\mu \in \mathcal{M}} \mathbb{E}_\mu \left[ |\hat{\theta} - \hat{\theta}_{h_{\Delta,n}^*, \Delta,n}|^2 \right] = O(r_{\Delta,n})$$

with the rates $r_{\Delta,n}$ collected in the following table:

| $C_1$ | $C_1 < \infty$ | $C_1 = \infty$ |
|-------|----------------|----------------|
| $\rho = 0$ | $s > \Delta \beta + \frac{1}{2}$ $T^{-1}$ | $s > \Delta \beta + 1$ $T^{-1}$ |
| $s = \Delta \beta + \frac{1}{2}$ | $(\log T)T^{-1}$ | $s = \Delta \beta + 1$ $(\log T)T^{-1}$ |
| $s < \Delta \beta + \frac{1}{2}$ | $T^{-\frac{2\rho+2\beta}{\alpha+\rho+2\beta+1}}$ | $s < \Delta \beta + 1$ $T^{-\frac{2\rho+2\beta}{\alpha+\rho+2\beta+1}}$ |
| $\rho > 0$ | $(\log T)\left(\frac{\Delta}{T}\right)^{-\frac{\rho+2\beta}{\rho+\alpha+2\beta+1}}$ | $(\log T)\left(\frac{\Delta}{T}\right)^{-\frac{\rho+2\beta}{\rho+\alpha+2\beta+1}}$ |

**Diskussion:** For point estimation ($s=0$) or the estimation of the k-th derivative ($s=-k$), we recover the rates of convergence which are classical and known to be optimal in a density deconvolution framework, see [Fan (1991)].

In the continuous limit, for $\Delta$ close to zero, we recover the rates which are typical for density estimation with pointwise loss. The rates of convergence found for pointwise loss should also be compared to the results found in [Belomestny (2011)].

When the estimation of $\mu(A) = \int 1_A \mu(dx)$ for some $A$ bounded away from zero is being considered, we have $s = 1$.

For the particular cases of estimating integrals and point estimation, the rates of convergence are known to be minimax optimal, see [Kappus (2012)] for lower bound results.

### 4 Adaptive estimation

Let a finite collection $\mathcal{M} = \{m_1, \ldots, m_n\} \subseteq \mathbb{N}$ of indices be given and let $\mathcal{H} := \{1/m_1, \ldots, 1/m_n\}$ be a collection of bandwidths associated with $\mathcal{M}$.

For notational convenience, we suppress, in this section, the dependence on $\Delta$ and assume $\Delta = 1$. Moreover, we slightly change the notation and denote the kernel estimator by $\hat{\theta}_{m,n}$ instead of $\hat{\theta}_{1,1/m,n}$.
The goal of this section is to provide a strategy for the optimal data driven choice of the smoothing parameter $\hat{m}$ within the collection $\mathcal{M}_n$ and to derive, for the corresponding estimator $\hat{\theta}_{m,n}$, the oracle inequality

$$\mathbb{E} \left[ (\theta - \hat{\theta}_{m,n})^2 \right] \leq C^{ad} \inf_{m \in \mathcal{M}} \left\{ (\theta - \theta_{m,n})^2 + \sup_{k \geq m} |\theta_k - \theta_{m}|^2 + \text{pen}(m) \right\} + Cn^{-1}, \quad (4.1)$$

with

$$\theta_{m} := \int f(x) \left( K \ast \mu \right)(x) \, dx,$$

with positive constants $C^{ad}$ and $C$ which do not depend on the unknown underlying smoothness parameters and a penalty term $\text{pen}(m)$ to be specified, which equals, up to some logarithmic factor, the quantity

$$\frac{1}{n} \sigma_{m}^2 := \frac{n^{-1}}{2\pi^2} \left\{ C_1 \int \left| \mathcal{F} \left( -u \right) \mathcal{F} \left( \frac{u}{m} \right) \right|^2 \, du \right\} \wedge C_2 \left\{ \int \left| \mathcal{F} \left( -u \right) \mathcal{F} \left( \frac{u}{m} \right) \right|^2 \, du \right\},$$

which bounds the error in the model. The occurrence of the supremum term in the oracle-inequality (4.1) is typical for the adaptive estimation of linear functionals, see Cai and Low (2005a) for lower bound results.

In a deconvolution framework with known distribution of the noise, see Butucea and Comte (2009), the way to go is to estimate the quantities $|\theta_k - \theta_m|^2$ involved in the oracle bound by their bias-corrected version, that is, to consider $|\hat{\theta}_k - \hat{\theta}_m|^2 - H^2(m, k)$, with some deterministic correction term $H^2(m, k)$ which is chosen large enough to ensure that with high probability,

$$|\hat{\theta}_k - \hat{\theta}_m|^2 - H^2(m, k) \leq |\theta_k - \theta_m|^2 \quad \forall m, k \in \mathcal{M}.$$

On the other hand, $H^2(m, k)$ should ideally not be much larger than the variance term. The appropriate choice then turns out to be

$$H^2(m, k) := \frac{1}{n} \rho \lambda_{m,k}^2 (\sigma_{m,k}^2 + x_{m,k}^2)$$

with some positive constant $\rho$ to be appropriately chosen and

$$\sigma_{m,k}^2 := \frac{1}{2\pi^2} \left\{ C_1 \int \left| \mathcal{F} \left( -u \right) \mathcal{F} \left( \frac{u}{m} \right) \right|^2 \left| \mathcal{F} \left( \frac{u}{k} \right) - \mathcal{F} \left( \frac{u}{m} \right) \right|^2 \, du \right\} \wedge C_2 \left\{ \int \left| \mathcal{F} \left( -u \right) \mathcal{F} \left( \frac{u}{m} \right) \right|^2 \left| \mathcal{F} \left( \frac{u}{k} \right) - \mathcal{F} \left( \frac{u}{m} \right) \right|^2 \, du \right\},$$

which involves in the oracle bound by

and $x_{m,k} := \frac{1}{\sqrt{n}} \frac{1}{2\pi} \int \left| \mathcal{F} \left( -u \right) \mathcal{F} \left( \frac{u}{m} \right) \right| \left| \mathcal{F} \left( \frac{u}{k} \right) - \mathcal{F} \left( \frac{u}{m} \right) \right| \, du$

and with logarithmic weights $\lambda_{m,k}$ chosen large enough to ensure $\sum_{m \in \mathcal{M}, k > m} e^{-\lambda_{m,k}} < \infty$. See also Laurent, Ludeña, and Prieur (2008) for the underlying ideas.

Indeed, this is the fundamental idea about model selection via penalization: Deterministic terms are applied in order to control the fluctuation of certain stochastic quantities, uniformly over some countable index set. In a white noise framework, this principle is illustrated in Birgé (1999).

The situation is different in the present framework, since the optimal theoretical penalty is no longer feasible, but depends on the characteristic function in the denominator, which is unknown.
It is intuitive to work with a stochastic penalty term and replace the unknown characteristic function by its empirical version. The model selection procedure will then crucially depend on a precise control of the fluctuation of the empirical characteristic function in the denominator around its target.

In a similar setting, Comte and Genon-Catalot (2010) have dealt with this problem by proposing an additional a priori assumption on the size of the collection \( \mathcal{M} \). However, this assumption is critical and highly restrictive, since it depends on the decay behaviour of the unknown characteristic function and hence involves some prior knowledge of the underlying smoothness parameters. Comte and Lacour (2011) have proposed another approach towards model selection with unknown variance, which does not depend on any prior knowledge of the smoothness parameters. However, this approach is designed for \( L^2 \)-loss and spectral cutoff estimation and the generalization to the estimation of linear functionals with general kernels is not straightforward. Moreover, the strategy proposed in that paper would lead, in the present case, to a polynomial loss. For this reason we propose a different strategy, which will also allow to drop certain semi-parametric assumptions on \( \varphi \).

In what follows, we introduce a newly defined estimator of the characteristic function in the denominator. The fluctuation of this object can be controlled not only pointwise, but uniformly on the real line, which is the key to making the model selection procedure work under very weak assumptions.

**Definition 4.1.** Let the weight function \( w \) be

\[
w(u) = (\log(e + |u|))^{-1/2 - \delta}
\]

for some \( \delta > 0 \). For a constant \( \kappa \) to be chosen, let the truncated version of \( \hat{\varphi}_n(u) \) be

\[
\hat{\varphi}_n(u) := \begin{cases} 
\hat{\varphi}_n(u), & \text{if } |\hat{\varphi}_n(u)| \geq \kappa(\log n)^{1/2}w(u)^{-1}n^{-1/2} \\
\kappa(\log n)^{1/2}w(u)^{-1}n^{-1/2}, & \text{else.}
\end{cases}
\]

Let the corresponding estimator of \( 1/\varphi \) be \( 1/\hat{\varphi}_n \).

The definition of the weight function originates from Neumann and Reiβ (2009) and the considerations presented therein will play an important role for our arguments. Introducing the extra logarithmic factor in the definition of \( 1/\hat{\varphi} \) will enable us to apply concentration inequalities of Talagrand type. This is the key to proving the following uniform version of the pointwise result which has been stated in Lemma 2.1 in Neumann (1997):

**Lemma 4.2.** Let \( c_1 \) be the constant appearing in Talagrand’s inequality (see Lemma 5.4). Let \( \kappa \) be chosen such that for some \( \gamma > 0 \), \( \kappa \geq 2(\sqrt{2c_1} + \gamma) \). Then we have for some constant \( C \) depending on the choice of \( \kappa, \gamma \) and \( \delta \),

\[
E \sup_{u \in \mathbb{R}} \left[ \frac{1}{(\log n)^2} \frac{\left| \frac{\hat{f}(u)}{\varphi_n(u)} - 1 \right|^2}{\left| \varphi_n(u) \right|^4} \wedge \frac{1}{\left| \varphi_n(u) \right|^2} \right] \leq C.
\]

The above definition gives rise to the following redefinition of the kernel estimator: In what follows, we set

\[
\hat{\theta}_{m,n} := \frac{1}{2\pi} \int \mathcal{F}(-u) \frac{\hat{\varphi}_n(u)}{\hat{\varphi}_n(u)} \mathcal{F}(K(u/m)) \, du.
\]

For \( m, k \in \mathcal{M} \), we define the stochastic correction term

\[
\hat{H}^2(m, k) := n^{-1} \left\{ c_{\text{pen}} c_1 \tilde{\sigma}_{m,k}^2 + 16 \left( \frac{5}{2} \kappa \right)^2 (\log n) \right\} \left( \tilde{\sigma}_{m,k}^2 + \tilde{x}_{m,k}^2 \right)
\]

where
with
\[
\hat{\sigma}_{m,k}^2 := \frac{1}{2\pi^2} \left\{ C_1 \int \left| \frac{\mathcal{F}f(u)}{\phi_n(u)} \right|^2 |\mathcal{F}K(u/k) - \mathcal{F}K(u/m)|^2 w(u)^{-2} \, du \right. \\
\left. \wedge C_2 \left( \int \left| \frac{\mathcal{F}f(u)}{\phi_n(u)} \right| |\mathcal{F}K(u/k) - \mathcal{F}K(u/m)| w(u)^{-1} \, du \right)^2 \right\}
\]
for positive constants $C_1$ and $C_2$ and with
\[
\tilde{x}_{m,k} := \frac{1}{\sqrt{n} 2\pi} \int \left| \frac{\mathcal{F}f(u)}{\phi_n(u)} \right| |\mathcal{F}K(u/k) - \mathcal{F}K(u/m)| w(u)^{-1} \, du.
\]
Let the weights be defined as follows: For some $\eta > 0$,
\[
\tilde{\lambda}_{m,k} := \frac{8}{\eta} \log \left( \log(n\tilde{x}_{m,k}(k - m)) \right)^2 \log(n\tilde{x}_{m,k}(k - m)) \log \left( \tilde{\sigma}_{m,k}^2(k - m)^2 \right)
\]
\[
\eta \left( \tilde{\sigma}_{m,k}^2(k - m)^2 \right).
\]
For some $\gamma > 0$, let $e^{\text{pen}} = 64 \vee 16(2c_1 + \gamma)$ and and $\kappa = 2(\sqrt{\alpha} + \gamma)$. Finally, let
\[
\tilde{\text{pen}}(m) := \tilde{H}^2(0, m).
\]
We understand by $\text{pen}(m)$ and $H^2(m, k)$ the deterministic versions of $\tilde{\text{pen}}(m)$ and $\tilde{H}^2(m, k)$, that is, the definitions are the same as in formula (4) and formula (4), apart from the fact that $1/\varphi_n$ is replaced by $1/\varphi$.

These definitions give rise to the following choice of the cutoff parameter:
\[
\tilde{m} := \arg\inf_{m \in \mathcal{M}} \sup_{k > m, k \in \mathcal{M}} \left\{ |\tilde{\theta}_k - \tilde{\theta}_m|^2 - \tilde{H}^2(m, k) + \tilde{\text{pen}}(m) \right\}
\]
We are now ready to state the following oracle bound and hence the main result of this section:

**Theorem 4.3.** Assume that (A1)-(A4), (F1) or (F2) and (K1)+(K2) hold. Assume that $C_1 \leq \tilde{C}_1$ and $C_2 \leq \tilde{C}_2$. Assume, moreover, that $E[\exp(\eta|X_1|)] < \infty$. Then
\[
E \left[ |\theta - \tilde{\theta}_m|^2 \right] \leq C^{\text{ad}} \inf_{m \in \mathcal{M}} \left\{ \left| \theta - \theta_{m_n} \right|^2 + \sup_{k > m} |\tilde{\theta}_k - \tilde{\theta}_m|^2 + \text{pen}(m) \right\} + Cn^{-1}
\]
holds with some $C^{\text{ad}} > 0$ depending on the particular choices of the constants, but not on the unknown parameters, and with some $C > 0$ depending on the choice of the constants and on $\eta^{-1}E[\exp(\eta|X_1|)]$.

It is worth mentioning that the exponential moment assumption can be relaxed, but at the cost of losing a polynomial factor and complicating the proofs. For sake of simplicity, we omit the details.

In comparison with the adaptive results obtained in [Comte and Lacour (2011)] or [Johannes and Schwarz (2012)], it is remarkable that our procedure is completely model free in the sense that it no longer depends on any semiparametric assumption on the decay of the characteristic function, be it exponential or polynomial decay or the assumption that $\varphi$ can, up to some constant, be bounded from above and below by some monotonically decreasing function. Moreover, apart from the condition that the linear functional is well defined, the measure $\mu$ may be fairly arbitrary and is not assumed to belong to any prescribed nonparametric class.

From an asymptotic point of view, the above results tells us that the procedure attains, up to a logarithmic loss, the minimax rates of convergence for the particular cases of point estimation, estimating derivatives and for estimating integrals.

It is well known that for estimating linear functionals, the loss of a logarithmic factor due to adaptation can typically not be avoided, see [Levskii (1992)]. Consequently, it is not surprising that a logarithmic loss is also found in the Lévy model.
5 Proofs

5.1 Proofs of the main results of Section 3

5.1.1 Risk bounds

The proof of Theorem 3.1 essentially relies on the following auxiliary result:

Lemma 5.1. For some universal positive constant C,

\[
\mathbb{E} \left[ \left( \frac{1}{\hat{\varphi}_{\Delta,n}(u)} - \frac{1}{\hat{\varphi}_{\Delta}(u)} \right) \left( \frac{1}{\hat{\varphi}_{\Delta,n}(v)} - \frac{1}{\hat{\varphi}_{\Delta}(v)} \right) \right] \\
\leq C \left( \frac{T^{-1}}{|\varphi_\Delta(u)|} \wedge 1 \right) \left( |\Psi''(u - v)| + |\Psi'(u - v)|^2 + |\Psi'(u)|\Psi'(v) \right).
\]

Proof. We start by noticing that for some constant \( C_k \) depending only on \( k \),

\[
\mathbb{E} \left[ \left( \frac{1}{\hat{\varphi}_{\Delta,n}(u)} - \frac{1}{\hat{\varphi}_{\Delta}(u)} \right)^k \right] \leq C_k \left( \frac{T^{-\frac{k}{2}}}{|\varphi_\Delta(u)|^2} \wedge \frac{1}{|\varphi_\Delta(u)|^k} \right). \tag{5.1}
\]

This is a direct consequence of Lemma 2.1 in [Neumann 1997], tracing back the dependence on \( \Delta \).

In the sequel, let \( R_{\Delta,n}(u) := \frac{1}{\hat{\varphi}_{\Delta,n}(u)} - \frac{1}{\varphi_\Delta(u)} \). Using the fact that \( \hat{\varphi}_{\Delta,n} \) and \( 1/\hat{\varphi}_{\Delta,n} \) are independent by construction and that \( \hat{\varphi}_{\Delta,n}(u) - \varphi_\Delta(u) \) is centred, we find that

\[
\mathbb{E} \left[ \left( \frac{\hat{\varphi}_{\Delta,n}(u)}{\hat{\varphi}_{\Delta,n}(u)} - \frac{\varphi_\Delta(u)}{\varphi_\Delta(u)} \right) \left( \frac{\hat{\varphi}_{\Delta,n}(v)}{\hat{\varphi}_{\Delta,n}(v)} - \frac{\varphi_\Delta(v)}{\varphi_\Delta(v)} \right) \right] \\
= \mathbb{E} \left[ \left( \frac{\hat{\varphi}_{\Delta,n}(u) - \varphi_\Delta(u)}{\hat{\varphi}_{\Delta,n}(u)} + \varphi_\Delta(u) R_{\Delta,n}(u) \right) \left( \frac{\hat{\varphi}_{\Delta,n}(v) - \varphi_\Delta(v)}{\hat{\varphi}_{\Delta,n}(v)} + \varphi_\Delta(v) R_{\Delta,n}(v) \right) \right] \\
= \text{Cov}(\hat{\varphi}_{\Delta,n}(u), \hat{\varphi}_{\Delta,n}(v)) \mathbb{E} \left[ \frac{1}{\hat{\varphi}_{\Delta,n}(u) \hat{\varphi}_{\Delta,n}(v)} \right] + \varphi_\Delta(u) \varphi_\Delta(v) \mathbb{E} \left[ R_{\Delta,n}(u) R_{\Delta,n}(v) \right].
\]

The Cauchy-Schwarz inequality and then an application of (5.1) imply

\[
\mathbb{E} |R_{\Delta,n}(u) R_{\Delta,n}(v)| \leq \left( \mathbb{E} \left[ \left| R_{\Delta,n}(u) \right|^2 \right] \right)^{\frac{1}{2}} \left( \mathbb{E} \left[ \left| R_{\Delta,n}(v) \right|^2 \right] \right)^{\frac{1}{2}} \\
\leq C_2 \left( \frac{T^{-1}}{|\varphi_\Delta(u)|^2 |\varphi_\Delta(v)|^2} \wedge \frac{1}{|\varphi_\Delta(u)| |\varphi_\Delta(v)|} \right). \tag{5.2}
\]

Using the triangle inequality, again (5.1) and then (5.2), we find that

\[
\mathbb{E} \left[ \left( \frac{1}{\hat{\varphi}_{\Delta,n}(u) \hat{\varphi}_{\Delta,n}(v)} \right) \right] \leq \left( \mathbb{E} \left[ \left| R_{\Delta,n}(u) \right| + \frac{1}{|\varphi_\Delta(u)|} \right] \right) \left( \mathbb{E} \left[ \left| R_{\Delta,n}(v) \right| + \frac{1}{|\varphi_\Delta(v)|} \right] \right) \\
\leq (1 + 2C_1 + C_2) \frac{1}{|\varphi_\Delta(u)| |\varphi_\Delta(v)|}.
\]

On the other hand, by definition of \( \frac{1}{\varphi_\Delta} \),

\[
\mathbb{E} \left[ \left( \frac{1}{\hat{\varphi}_{\Delta,n}(u) \hat{\varphi}_{\Delta,n}(v)} \right) \right] \leq (\Delta n) = T.
\]

Next,

\[
\left| \text{Cov}(\hat{\varphi}_{\Delta,n}(u), \hat{\varphi}_{\Delta,n}(v)) \right| = n^4 \mathbb{E} \left[ (iZ_\Delta)^2 e^{i(u-v)Z_\Delta} \right] - \mathbb{E} \left[ iZ_\Delta e^{iuZ_\Delta} \right] \mathbb{E} \left[ iZ_\Delta e^{ivZ_\Delta} \right] \\
= n^4 \left| \varphi_\Delta(u - v) - \varphi_\Delta(u) \varphi_\Delta(v) \right| \\
\leq n^4 \left( |\Delta \Psi''(u - v)| + |\Delta \Psi'(u - v)|^2 + \Delta^2 |\Psi'(u)| \Psi'(v) \right). \tag{5.3}
\]
Putting (5.2)-(5.3) together, we have shown

\[
\left| \text{Cov} \left( \hat{\varphi}_{\Delta,n}(u), \hat{\varphi}_{\Delta,n}(v) \right) \right| \leq C' \left( \frac{T^{-1}}{|\varphi_{\Delta}(u)||\varphi_{\Delta}(v)|} \wedge 1 \right) \Delta^2 \left( |\Psi''(u-v)| + |\Psi'(u-v)|^2 + |\Psi'(u)|^2 \right).
\]

With constant \( C' := 1 + 2C_2 + C_1 \).

Another application of (5.2) gives

\[
|\mathbb{E}[R_{\Delta,n}(u)R_{\Delta,n}(v)]| \leq 2 \mathbb{E}[R_{\Delta,n}(u)R_{\Delta,n}(v)] \Delta^2 |\varphi'(u-v)\Psi'(u)\Psi'(v)|
\]

\[
\leq C_2 \left( \frac{T^{-1}}{|\varphi_{\Delta}(u)||\varphi_{\Delta}(v)|} \wedge 1 \right) \Delta^2 |\Psi'(u)|^2 |\Psi'(v)|^2.
\]

This completes the proof. \( \square \)

**Proof of Theorem 5.1**. Given the kernel function \( K \) and bandwidth \( h \), let \( \theta_h := \int f(x)(K_h * \mu)(x) \, dx \). We can estimate

\[
\mathbb{E} \left[ \left| \theta - \tilde{\theta}_{\Delta,h,n} \right|^2 \right] \leq 2 |\theta - \theta_h|^2 + 2 \mathbb{E} \left[ \left| \theta_h - \tilde{\theta}_{\Delta,h,n} \right|^2 \right].
\]

In what follows, we use the notation

\[
T(u,v) := \left( \frac{\hat{\varphi}_{\Delta,n}(u)}{\varphi_{\Delta,n}(u)} - \frac{\varphi'_\Delta(u)}{\varphi_{\Delta}(u)} \right) \left( \frac{\hat{\varphi}_{\Delta,n}(v)}{\varphi_{\Delta,n}(v)} - \frac{\varphi'_\Delta(v)}{\varphi_{\Delta}(v)} \right).
\]

Passing to the Fourier domain and applying Fubini’s theorem yields

\[
\mathbb{E} \left[ \left| \theta - \tilde{\theta}_{\Delta,h,n} \right|^2 \right] = \mathbb{E} \left[ \left| \frac{1}{2\pi} \int \mathcal{F} f(-u) \mathcal{F} K_h(u) \frac{1}{\Delta} \left( \frac{\hat{\varphi}_{\Delta,n}(u)}{\varphi_{\Delta,n}(u)} - \frac{\varphi'_\Delta(u)}{\varphi_{\Delta}(u)} \right) \, du \right|^2 \right] = \frac{1}{4\pi^2\Delta^2} \int \int \mathcal{F} f(-u) \mathcal{F} f(v) \mathcal{F} K_h(u) \mathcal{F} K_h(v) \mathbb{E}[T(u,v)] \, du \, dv.
\]

Thanks to Lemma 5.1

\[
\int \int \mathcal{F} f(-u) \mathcal{F} f(v) \mathcal{F} K_h(u) \mathcal{F} K_h(-v) \frac{1}{\Delta^2} \mathbb{E}[T(u,v)] \, du \, dv \leq C T^{-1} \left( \int \left| \mathcal{F} f(-u) \mathcal{F} f(v) \right| \left| \mathcal{F} K_h(u) \mathcal{F} K_h(v) \right| \left( |\Psi''(u-v)| + |\Psi'(u-v)|^2 \right) \, du \, dv + \int \int \left| \mathcal{F} f(-u) \mathcal{F} f(v) \right| \left| \mathcal{F} K_h(u) \mathcal{F} K_h(v) \right| \left| \Psi'(u)\Psi'(v) \right| \, du \, dv \right).
\]

If \( \Psi'' \in L^1(\mathbb{R}) \) and \( \Psi' \in L^2(\mathbb{R}) \), we apply the Cauchy-Schwarz inequality and Fubini’s theorem to find

\[
\int \int \left| \mathcal{F} f(-u) \mathcal{F} f(v) \right| \left| \mathcal{F} K_h(u) \mathcal{F} K_h(v) \right| \left( |\Psi''(u-v)| + |\Psi'(u-v)|^2 \right) \, du \, dv \leq \int \left| \mathcal{F} f(-u) \right|^2 \left| \mathcal{F} K_h(u) \right|^2 \int (|\Psi''(u-v)| + |\Psi'(u-v)|^2) \, du \, dv = \left( ||\Psi''||_{L^1} + ||\Psi'||_{L^2}^2 \right) \int \left| \mathcal{F} f(-u) \right|^2 \left| \mathcal{F} K_h(u) \right|^2 \, du.
\]
and
\[ \int \int \frac{|\mathcal{F}(u)\mathcal{F}(v)|}{|\mathcal{F}(u)\varphi_{\Delta}(\cdot-v)|} |K_h(u)K_h(\cdot-v)\Psi'(u)\Psi'(-v)| \, du \, dv \leq \|\Psi'\|_2^2 \int \frac{|\mathcal{F}(u)\mathcal{F}(u)|^2}{|\varphi_{\Delta}(u)|^2} \, du. \]

On the other hand, if \( \Psi'' \in L^1(\mathbb{R}) \) or \( \Psi' \in L^2(\mathbb{R}) \) fails to hold, we can always use the estimate
\[ \int \int \frac{|\mathcal{F}(u)\mathcal{F}(v)|}{|\varphi_{\Delta}(u)\varphi_{\Delta}(\cdot-v)|} |K_h(u)K_h(\cdot-v)(|\Psi''(u-v)| + |\Psi'(u-v)|) \, du \, dv \leq (\|\Psi''\|_{\infty} + \|\Psi'\|_{\infty}) \left( \int \frac{|\mathcal{F}(u)|}{|\varphi_{\Delta}(u)|} |\mathcal{F}(K_h(u))| \, du \right)^2 \]
and
\[ \int \int \frac{|\mathcal{F}(u)\mathcal{F}(v)|}{|\varphi_{\Delta}(u)\varphi_{\Delta}(\cdot-v)|} |\mathcal{F}(K_h(u))\mathcal{F}(K_h(\cdot-v))| \, du \, dv \leq (\int \frac{|\mathcal{F}(u)|}{|\varphi_{\Delta}(u)|} |\mathcal{F}(K_h(u))| \, du \right)^2 \leq \|\Psi'\|_{\infty}^2 \left( \int \frac{|\mathcal{F}(u)|}{|\varphi_{\Delta}(u)|} |\mathcal{F}(K_h(u))| \, du \right)^2. \]

This completes the proof.

5.1.2 Rates of convergence

Lemma 5.2. In the situation of Theorem 3.3, the approximation error can be estimated as follows:
\[ \left| \int f(x)\mu(dx) - \int f(x)(K_h * \mu)(x) \, dx \right|^2 \leq \frac{C_B}{(2\pi)^2} h^{2a+2s} =: b_h \]
with constant
\[ C_B = \left( \frac{2\pi}{(a+s)!} \right)^2 (2\pi)^{-s-a} + \left( \frac{L_K}{(a+s)!} \right)^2 M_f M_\mu. \]

Proof. By assumption, \( f \) is Sobolev-regular with index \( s \) and \( \mu \) is Sobolev-regular with index \( a > -s \). This implies, by duality of Sobolev spaces, that we can pass to the Fourier domain and write
\[ \left| \int f(x)\mu(dx) - \int f(x)(K_h * \mu)(x) \, dx \right|^2 = \frac{1}{2\pi} \int \left| \mathcal{F}(u) (1 - \mathcal{F}(\mathcal{F}(K_h(u)) \mathcal{F}(u)) \, du \right|^2. \]

Applying the Cauchy Schwarz inequality and then the regularity assumptions on \( f \) and on \( \mu \), we find that
\[ \left| \frac{1}{2\pi} \int \mathcal{F}(-u)(1 - \mathcal{F}(K_h(u)) \mathcal{F}(u)) \, du \right|^2 \leq \frac{1}{(2\pi)^2} \int |\mathcal{F}(u)|^2 (1 + |u|^2)^s \, du \int |1 - \mathcal{F}(K_h(u)|^2 (1 + |u|^2)^{-a-s} |\mathcal{F}(u)|^2 (1 + |u|^2)^a \, du \]
\[ \leq \frac{M_f M_\mu}{(2\pi)^2} \sup_{u \in \mathbb{R}} |1 - \mathcal{F}(K_h(u)|^2 (1 + |u|^2)^{-a-s}. \]

If \( K \) is the sinc kernel, we can immediately estimate
\[ \sup_{u \in \mathbb{R}} |1 - \mathcal{F}(K(\cdot u)|^2 (1 + |u|^2)^{-a-s} = \sup_{|u| \geq \pi} (1 + |u|^2)^{-a-s} \leq \pi^{-2a-2s} h^{2a+2s}, \]
which gives the desired result.
If $\mathcal{F}K$ is $(a+s)$-times continuously differentiable and the derivatives up to order $(a+s)$ at zero vanish, a Taylor series expansion gives for some constant $\tau \in [-hu, hu]$:

$$1 - \mathcal{F}K(hu) = \frac{\mathcal{F}K^{(a+s)}(\tau)}{(a+s)!}(hu)^{a+s}.$$  

(5.7)

By Hölder continuity of $\mathcal{F}K^{(a+s)}$,

$$|\mathcal{F}K^{(a+s)}(\tau)| = |\mathcal{F}K^{(a+s)}(\tau) - \mathcal{F}K^{(a+s)}(0)| \leq L_K|hu|^{a+s-\langle a+s \rangle}. \tag{5.8}$$

From (5.6), (5.7) and (5.8) we derive

$$\text{Lemma 5.3.} \quad \text{If } \mathcal{F}f \text{ is a distribution of order } \langle a, b \rangle \text{ and let } \mu_1 \text{ be the signed measure with density } g_1 \text{ and let } \mu_2 := \mu - \mu_1. \text{ Then the approximation error can be decomposed as follows: Since } \text{supp}(f) = [a, b] \subseteq D, \text{ we have}$$

$$\left| \int f(x)\mu(dx) - \int f(x)(\mathcal{F}K \ast \mu)(x)\,dx \right|^2 \leq C_B h^{2a+2s}$$

for some positive constant $C_B$ depending on $C_K, a - d_1 \lor d_2 - b, L$ and $R$.

**Proof.** We use the trivial observation that the local density $g_D$ can be extended to a compactly supported function $g \in \mathcal{F}\mathcal{D}(a, L', R')$ with constants $L' > L$ and $R' > R$. Let $\mu_1$ be the signed measure with density $g_1$ and let $\mu_2 := \mu - \mu_1$. Then the approximation error can be decomposed as follows: Since \text{supp}(f) = [a, b] \subseteq D, \text{ we have}

$$\left| \int f(x)\mu(dx) - \int f(x)(\mathcal{F}K \ast \mu)(x)\,dx \right|$$

$$= \left| \int f(x)(g_1(x) - K_h \ast g_1(x))\,dx - \int f(x)(\mathcal{F}K \ast \mu_2)(x)\,dx \right|$$

$$\leq \left| \int f(x)(g_1 - K_h \ast g_1)(x)\,dx \right| + \left| \int f(x)(\mathcal{F}K \ast \mu_2)(x)\,dx \right|. \tag{5.9}$$

Recall that $f$ is compactly supported and, by assumption, $|\mathcal{F}f(u)| \leq C_f(1 + |u|)^{-\alpha}$. Consequently, $f$ is a distribution of order $k := -\alpha$, see Theorem 7.23 in [Rudin 1991]. Thanks to Theorem 6.34 ibidem, we can estimate

$$\left| \int f(x)(\mathcal{F}K \ast \mu_2)(x)\,dx \right| \leq \|f\| \sup_{m \leq k} \sup_{x \in [a, b]} \left| (\mathcal{F}K \ast \mu_2)(m)(x) \right|$$

for some constant $\|f\|$ depending only on $f$. By assumption, $K_h$ is $k$-times continuously differentiable, with bounded derivatives. Since, moreover, $\mu_2$ is finite, we can derive under the integral sign and write

$$\|f\| \sup_{m \leq k} \sup_{x \in [a, b]} \left| (\mathcal{F}K \ast \mu_2)(m)(x) \right| = \|f\| \sup_{m \leq k} \sup_{x \in [a, b]} \left| (K_h(m) \ast \mu_2)(x) \right|$$

$$= \|f\| \sup_{m \leq k} \sup_{x \in [a, b]} \left| \int h^{-m-1} K^{(m)} \left( \frac{x - y}{h} \right) \mu_2(dy) \right|. \tag{5.10}$$

Using the fact that \text{supp}(f) = [a, b] \subseteq D = (d_1, d_2) and that $\mu_2|_D \equiv 0$ by construction, we continue from (5.10) by estimating

$$\|f\| \sup_{m \leq k} \sup_{x \in [a, b]} \left| \int h^{-m-1} K^{(m)} \left( \frac{x - y}{h} \right) \mu_2(dy) \right| \leq \|f\| \sup_{m \leq k} \sup_{z \geq \delta} \left| K^{(m)} \left( \frac{z}{h} \right) \right| \|\mu_2(\mathbb{R})\|.$$
with \( \delta := (a - d_1) \wedge (d_2 - b) \). Finally, the assumptions on the decay of \( K \) and its derivatives up to order \( k \) give

\[
\| f \| \sup_m \sup_{h \geq \delta} \left| \mathcal{K}^{(m)} \left( \frac{x}{h} \right) \right| \mu_2([\mathbb{R}]) \leq \| f \| C_K \mu_2([\mathbb{R}]) \delta^{a - (s + 0)} - 1 \delta^{a + s}.
\]

It remains to consider the first expression in the last line of formula (5.9). We observe that

\[
\int f(x) (K_h * g_1)(x) \, dx = \int K_h(y) \int f(x) g_1(x - y) \, dx \, dy.
\]

(5.11)

If \( f \) is regular, this is simply a consequence of Fubini’s theorem. For non-regular \( f \), a straightforward generalization of Theorem 39.3 and Theorem 39.10 in Jantscher (1971) gives formula (5.11).

Let \( \tilde{g}_1(x) := g_1(-x) \). Then

\[
\int f(x)(g_1(x) - K_h * g_1(x)) \, dx = \int K_h(y) \left( \int g_1(x)f(x) \, dx - \int g_1(x - y)f(x) \, dx \right) \, dy
\]

\[
= \int K_h(y) \left( f * \tilde{g}_1(0) - f * \tilde{g}_1(y) \right) \, dy.
\]

We observe that \( f * \tilde{g}_1 \) is \( (a + s) \)-times continuously differentiable and that the derivative of order \( \langle a \rangle + s \) is \( a - \langle a \rangle \)-Hölder continuous. To see this, we use the fact that \( (f * \tilde{g}_1)^{(\langle a \rangle + s)} = f^{(s)} * \tilde{g}_1^{(\langle a \rangle)} \), where \( f^{(s)} \) is understood to be a distributional derivative. (See Theorem 41.3 in Jantscher (1971) for explanation.) Now, since \( \| f^{(s)}(-u) \| = |u|^{s} \| f(-u) \| \leq C_f \), we can use Theorem 7.23 in Rudin (1991) to conclude that \( f^{(s)} \) is a compactly supported distribution of order 0. From this, and Theorem 6.34 in Jantscher (1971), we derive that

\[
\left| f^{(s)} * \tilde{g}_1^{(\langle a \rangle)}(x) - f^{(s)} * \tilde{g}_1^{(\langle a \rangle)}(y) \right| \leq \| f^{(s)} \| \sup_{\tau \in [a,b]} \left| \tilde{g}_1^{(\langle a \rangle)}(x - \tau) - \tilde{g}_1^{(\langle a \rangle)}(y - \tau) \right| \leq \| f^{(s)} \| L' |x - y|^{a - \langle a \rangle}
\]

for a constant \( \| f^{(s)} \| \) depending on \( f \).

Now, a Taylor series expansion of \( f * \tilde{g}_1 \) and an application of the order of \( K \) yields for some \( \tau_y \in [0, y] \):

\[
\left| \int K_h(y) ((f * \tilde{g}_1)(0) - (f * \tilde{g}_1)(y)) \, dy \right| = \frac{1}{(a + s)!} \int K_h(y) \left( (f * \tilde{g}_1)^{(a + s)}(\tau_y) - (f * \tilde{g}_1)^{(a + s)}(0) \right) y^{a + s} \, dy
\]

\[
\leq \frac{1}{(a + s)!} \| f^{(s)} \| L' \int |K_h(y)||y|^{a - \langle a \rangle} |y|^{a + s} \, dy
\]

\[
= \frac{1}{(a + s)!} \| f^{(s)} \| L' \int |K(z)||z|^{a + s} \, dz \, h^{a + s}.
\]

This completes the proof.

Theorem 5.2.1 and Theorem 5.2.3 are immediate consequences of Lemma 5.3 and Lemma 5.3 combined with the assumptions on \( K \) and \( f \) and on the decay of \( \varphi \).

5.2 Adaptive estimation

5.2.1 Preliminaries

Recall the following well known result:
Lemma 5.4 (Talagrand’s inequality). Let $I$ be some countable index set. For each $i \in I$, let $X^{(i)}_n, \ldots, X^{(i)}_n$ be centred i.i.d. complex valued random variables, defined on the same probability space, with $\|X^{(i)}_n\| \leq B$ for some $B < \infty$. Let $v^2 := \sup_{i \in I} \text{Var}(X^{(i)}_n)$. Then for arbitrary $\varepsilon > 0$, there are positive constants $c_1$ and $c_2 = c_2(\varepsilon)$ depending only on $\varepsilon$ such that for any $\kappa > 0$:

$$
P\left( \sup_{i \in I} |S_n^{(i)}| \geq (1 + \varepsilon) \mathbb{E} \left[ \sup_{i \in I} |S_n^{(i)}| \right] + \kappa \right) \leq 2 \exp \left( -n \left( \frac{\kappa^2}{c_1 v^2} \wedge \frac{\kappa}{c_2 B} \right) \right).$$

A proof can be found for example in Massart (2003).

The main objective of the present subsection is to prove Lemma 4.2. We start by providing a series of technical results. The arguments presented here are fairly general and not particular to the Lévy model.

Lemma 5.5. Let $\tau > 0$ be given. Let $\delta$ be the constant appearing in the definition of the weight function $w$ and let $c_1$ be the constant in Talagrand’s inequality. Then, for arbitrary $\gamma > 0$, there is a positive constant $C = C_{\tau, \gamma, \delta}$ depending only on the choice of $\tau, \gamma$ and $\delta$ such that we have for $n \geq 1$:

$$
P\left( \exists u \in \mathbb{R} : |\hat{\varphi}_n(u) - \varphi(u)| \geq \tau (\log n)^{1/2} w(u)^{-1/2} \right) \leq C n^{-\frac{(\gamma - \delta)^2}{\delta c_1}}.
$$

Proof. We prove the claim for the countable set of rational numbers. By continuity of the characteristic function and of $w$, it carries over to the whole range of real numbers.

By Theorem 4.1 in Neumann and Reiß (2009), we have for some positive constant $C$:

$$
\mathbb{E} \left[ \sup_{u \in \mathbb{R}} |\hat{\varphi}_n(u) - \varphi(u)| w(u) \right] \leq C n^{-1/2}.
$$

Since moreover, we trivially have $\sup_{u \in \mathbb{R}} \text{Var}[\hat{\varphi}_1(u)] \leq 1$ and $\sup_{u \in \mathbb{R}} \|\hat{\varphi}_1(u) w(u)\| \leq 1$, we can apply Talagrand’s inequality. Setting

$$
\kappa_n := \tau (\log n)^{1/2} n^{-1/2} - (1 + \varepsilon) C n^{-1/2},
$$

for some $\varepsilon > 0$, we can estimate

$$
P\left( \sup_{q \in \mathbb{Q}} |\hat{\varphi}_n(q) - \varphi(q)| w(q) \geq \tau (\log n)^{1/2} n^{-1/2} \right)

= \P\left( \sup_{q \in \mathbb{Q}} |\hat{\varphi}_n(q) - \varphi(q)| w(q) \geq \tau (\log n)^{1/2} n^{-1/2} \right)

\leq \P\left( \sup_{q \in \mathbb{Q}} |\hat{\varphi}_n(q) - \varphi(q)| w(q) \geq (1 + \varepsilon) \mathbb{E} \left[ \sup_{q \in \mathbb{Q}} |\hat{\varphi}_n(q) - \varphi(q)| w(q) \right] + \kappa_n \right)

\leq 2 \exp \left( -n \left( \frac{\kappa_n^2}{c_1} \wedge \frac{\kappa_n}{c_2} \right) \right).
$$

By definition of $\kappa_n$, we have for $C$ large enough and arbitrary $n \geq 1$:

$$
2 \exp \left( -n \left( \frac{\kappa_n^2}{c_1} \wedge \frac{\kappa_n}{c_2} \right) \right)

\leq 2 \exp \left( - \frac{(\tau (\log n)^{1/2} - (1 + \varepsilon) C)^2}{c_1} \right) \vee 2 \exp \left( - \frac{n^{1/2} (\tau (\log n)^{1/2} - (1 + \varepsilon) C)}{c_2} \right)

\leq C \exp \left( - \frac{(\tau - \gamma)^2}{c_1} (\log n) \right) = C n^{-\frac{(\gamma - \delta)^2}{\delta c_1}}.
$$

This is the desired result for the rational numbers and hence, by continuity, for the real line.
We can now use Lemma 5.5 to analyse the deviation of $1/\tilde{\varphi}_n$ from $1/\varphi$.

**Lemma 5.6.** Assume that for some $\gamma > 0$ and some $p > 0$, we have $\kappa \geq 2(\sqrt{pc_1} + \gamma)$, where $c_1$ denotes the constant in Talagrand’s inequality. Then

$$\mathbb{P}\left( \exists u \in \mathbb{R} : \left| \frac{1}{\tilde{\varphi}_n(u)} - \frac{1}{\varphi(u)} \right|^2 > \left( \frac{4\kappa}{2} \frac{(\log n)w(u)^{-1}n^{-1}}{|\varphi(u)|^4} \wedge \frac{(5/2)^2}{|\varphi(u)|^2} \right) \right) \leq Cn^{-p}.$$ 

**Proof.** We use the notation $a_n(u) := \kappa(\log n)^{1/2}w(u)^{-1/2}$. Let us introduce the favourable set

$$C := \left\{ \forall u \in \mathbb{R} : |\tilde{\varphi}_n(u) - \varphi(u)| \leq a_n(u)/2 \right\}.$$

Thanks to the preceding Lemma and the choice of $\kappa$,

$$\mathbb{P}(C^c) \leq Cn^{-p},$$

so it is enough to consider $C$. We have

$$\left| \frac{1}{\tilde{\varphi}_n(u)} - \frac{1}{\varphi(u)} \right|^2 = \frac{|\tilde{\varphi}_n(u) - \varphi(u)|^2}{|\tilde{\varphi}_n(u)|^2|\varphi(u)|^2}.$$ 

Consider first the case where $|\varphi(u)| > 3a_n(u)/2$. Then, by definition of $C$,

$$|\tilde{\varphi}_n(u)| \geq |\varphi(u)| - |\varphi(u) - \tilde{\varphi}_n(u)| > \frac{2}{3}|\varphi(u)| \geq a_n(u).$$

Consequently,

$$\frac{|\tilde{\varphi}_n(u) - \varphi(u)|^2}{|\tilde{\varphi}_n(u)|^2|\varphi(u)|^2} \leq \frac{(3a_n(u)/4)^2}{|\varphi(u)|^4} \leq \frac{1}{4|\varphi(u)|^2}.$$ 

Now, consider $|\varphi(u)| \leq 3a_n(u)/2$. By definition of $C$,

$$|\varphi_n(u) - \varphi(u)| \leq \max\{|\tilde{\varphi}_n(u) - \varphi(u)|, |\varphi(u)| + a_n(u)\} \leq 5a_n(u)/2.$$ 

Since, moreover, $|\varphi_n(u)| \geq a_n(u)$ holds by definition of $\varphi_n$,

$$\frac{|\varphi_n(u) - \varphi(u)|^2}{|\varphi_n(u)|^2|\varphi(u)|^2} \leq \frac{(5a_n(u)/2)^2}{|\varphi_n(u)|^2|\varphi(u)|^2} \leq \frac{(5/2)^2}{|\varphi(u)|^2} \leq \frac{(15a_n(u)/4)^2}{|\varphi(u)|^4}.$$ 

This is the desired result. \qed

The following useful corollary is an immediate consequence of the preceding statement.

**Corollary 5.7.** In the situation of the preceding Lemma,

$$\mathbb{P}\left( \exists u \in \mathbb{R} : \left| \frac{1}{\varphi_n(u)} - \frac{1}{\varphi(u)} \right|^2 > \left( \frac{5}{2} \kappa \right)^2 \frac{(\log n)w(u)^{-2}n^{-1}}{|\varphi_n(u)|^2|\varphi(u)|^2} \right) \leq Cn^{-p}.$$ 

The uniform version of the classical Neumann Lemma can now be stated as an easy consequence of Lemma 5.6.

**Proof of Lemma 4.2.** We use the notation

$$A_n(u) := \frac{\left| \frac{1}{\varphi_n(u)} - \frac{1}{\varphi(u)} \right|^2}{(\log n)w(u)^{-2}n^{-1} \wedge \frac{1}{|\varphi(u)|^2}}.$$
Let the set $C$ be defined as in the proof of Lemma 5.6. We can decompose
\[
E \left[ \sup_{u \in \mathbb{R}} A_n(u) \right] = E \left[ \sup_{u \in \mathbb{R}} A_n(u)1_C \right] + E \left[ \sup_{u \in \mathbb{R}} A_n(u)1_{C^c} \right]
\]
The definition of $C$, together with Proposition 5.6, readily implies
\[
E \left[ \sup_{u \in \mathbb{R}} A_n(u)1_C \right] \leq E \left[ \sup_{u \in \mathbb{R}} \frac{(4\kappa)^2 (\log n)^{-1} a_n(u)}{|\varphi(u)|^4} \wedge \frac{(\frac{2}{\kappa})^2}{|\varphi(u)|^2} \right] \leq 16\kappa^2.
\]
On the other hand, since we have, by definition $1/\hat{\varphi}_n(u) \leq a_n(u)^{-1}$, we can estimate for arbitrary $u \in \mathbb{R}$:
\[
A_n(u) \leq |\varphi(u)|^2 \frac{a_n(u)^{-1} + \frac{1}{\varphi(u)}}{a_n(u)^2 \vee |\varphi(u)|^2} \leq (2a_n(u)^{-2} + 1)^2 \leq (2\kappa^{-2} + 1)^2 n^2,
\]
so
\[
E \left[ \sup_{u \in \mathbb{R}} A_n(u)1_{C^c} \right] \leq (2\kappa^{-2} + 1)^2 n^2 P(C^c).
\]
Since Lemma 5.6 implies $P(C^c) \leq Cn^{-2}$, this gives the desired result.

The result immediately extends to values different from 2. The following Corollary can be obtained, replacing in each step 2 by $2q$:

**Corollary 5.8.** In the situation of the preceding statement, let $\kappa \geq 2 \left( \sqrt{2q} + 1 \right)$ for some $q \in \mathbb{R}^+$. Then we have for some constant $C_c = C_c^{q, \delta, \gamma}$ depending on $\gamma$, $q$, $\delta$ and $\epsilon$:
\[
E \left[ \sup_{u \in \mathbb{R}} \frac{|\varphi_n(u)|^2}{(\log n)^{3q} a_n(u)^{2-2q} \varphi(u) + 1} \right] \leq C_c.
\]

**5.2.2 Auxiliary results**

The main result of the present subsection is Proposition 5.11, which is then the most important technical tool for analysing the adaptive bandwidth selection.

We use, in the prequel, the short notation
\[
\Delta_{m,k} \mathcal{F} K(u) := \mathcal{F} K(u/k) - \mathcal{F} K(u/m).
\]

**Lemma 5.9.** For $k, m \in \mathbb{N}$, let
\[
x^2_{f_{m,k}} := \frac{1}{2\pi} \left\{ C_1 \int |\mathcal{F} f(-u)| \Delta_{m,k} \mathcal{F} K(u)^2 \, du \wedge C_2 \left( \int |\mathcal{F} f(-u)| \Delta_{m,k} \mathcal{F} K(u) \, du \right)^2 \right\}.
\]

Moreover, let
\[
\lambda_{f_{m,k}} := \sqrt{\kappa} \log \left( x^2_{f_{m,k}} (k - m)^2 \right).
\]

For some $\gamma > 0$, let $\kappa = 2(\sqrt{2qC} + \gamma)$. Then we have for some constant $C$ depending on $\gamma$, $\delta$ and $\epsilon$:
\[
P \left( \exists u \in \mathbb{R} : |\hat{\varphi}_n(u) - \varphi(u)| \geq \left( \frac{K}{2} (\log n)^{1/2} + \lambda_{f_{m,k}} \right) w(u)^{-1/2} \right) \leq Cn^{-p} x^2_{f_{m,k}} (k - m)^{-2}.
\]
Proof. The proof runs along the same lines as the proof of Lemma 5.5 setting, this time
\[ \kappa_n := \left( \frac{K}{2} (\log n)^{1/2} + \lambda_{f_{m,k}} \right) n^{-1/2} - C n^{1/2}. \]

Using again continuity of the (empirical) characteristic function, the Talagrand inequality and the choice of \( \kappa \), we derive that for \( C \) chosen large enough,
\[
\mathbb{P} \left( \exists u \in \mathbb{R} : \left| \frac{1}{\hat{\varphi}_n(u)} - \frac{1}{\varphi(u)} \right| ^2 > \left( \frac{\kappa}{2} (\log n)^{1/2} + \lambda_{f_{m,k}} \right) w(n) u^{-2} n^{-1} \right) \\
\leq 2 \exp \left( - \frac{(\kappa/2 - \gamma)^2}{c_1} (\log n) - \log \left( M_{f_{m,k}(k-m)}^2 \right) \right) \\
= C n^{-p} x_{f_{m,k}}^{-2} (k-m)^{-2}.
\]

The above result implies the following extension of Corollary 5.6.

**Corollary 5.10.** In the situation of the preceding statement, we have for some constant \( C \) depending on \( \gamma \) and \( \delta \):
\[
\mathbb{P} \left( \exists u \in \mathbb{R} : \left| \frac{1}{\hat{\varphi}_n(u)} - \frac{1}{\varphi(u)} \right| ^2 > \left( \frac{\kappa}{2} (\log n)^{1/2} + \lambda_{f_{m,k}} \right) w(n) u^{-2} n^{-1} \right) \\
\leq C n^{-p} x_{f_{m,k}}^{-2} (k-m)^{-2}.
\]

**Proposition 5.11.** Assume that the conditions which are summarized in Theorem 4.3 are satisfied. Then we can estimate for arbitrary \( m \in \mathbb{N} : \)
\[
\mathbb{E} \left[ \sup_{k > m, k \in \mathbb{N}} \left\{ \left( \hat{\theta}_k - \hat{\theta}_m \right) - (\theta_k - \theta_m) \right\}^2 - \frac{1}{2} \widehat{H}(m,k)^2 \right] + \leq C n^{-1},
\]
where \( C \) is a positive constant depending on the exponential moment.

**Proof.** Let
\[
\tilde{\theta}_m := \frac{1}{2\pi} \int \mathcal{F}f(-u) \frac{\varphi'(u)}{\hat{\varphi}_n(u)} \mathcal{F}K(u/m) \ du.
\]

We use the estimate
\[
\mathbb{E} \left[ \sup_{k > m, k \in \mathbb{N}} \left\{ \left( \hat{\theta}_k - \hat{\theta}_m \right) - (\theta_k - \theta_m) \right\}^2 - \frac{1}{2} \widehat{H}(m,k)^2 \right] + \leq 2 \mathbb{E} \left[ \sup_{k > m, k \in \mathbb{N}} \left\{ \left( \hat{\theta}_k - \hat{\theta}_m \right) - (\theta_k - \theta_m) \right\}^2 - \frac{1}{8} \widehat{H}(m,k)^2 \right] + \\
+ 2 \mathbb{E} \left[ \sup_{k > m, k \in \mathbb{N}} \left\{ \left( \hat{\theta}_k - \hat{\theta}_m \right) - (\theta_k - \theta_m) \right\}^2 - \frac{1}{8} \widehat{H}(m,k)^2 \right] +.
\] (5.13)

Consider first the expression appearing in the second line of formula (5.13). We can estimate,
Moreover, we define the remainder term \( Z_j := Z_j 1_{\{|Z_j| \leq \frac{1}{\eta} (\log n + \log \bar{m}_h (k - m))\}}. \)  

In the present situation, \( \hat{\varphi}_n \) is unbounded. To be able to apply the Bernstein inequality, let us introduce the truncated version of \( Z_j \),

\[
\tilde{Z}_j := Z_j 1_{\{|Z_j| \leq \frac{1}{\eta} (\log n + \log \bar{m}_h (k - m))\}}.
\]

Moreover, we define the remainder term \( Z_j^r := Z_j - \tilde{Z}_j \). Then

\[
\begin{align*}
& \left| \hat{\theta}_k - \hat{\theta}_m \right| - \left| \tilde{\theta}_k - \tilde{\theta}_m \right| \right|^2 \\
= & \left| \frac{1}{n} \sum_{j=1}^{n} \frac{1}{2\pi} \int \mathcal{F}f(-u) \frac{Z_j e^{iuZ_j} - E[Z_j e^{iuZ_j}]}{\hat{\varphi}_n(u)} \Delta_{m,k} \mathcal{F}K(u) \, du \right|^2 \\
\leq & 2 \left| \frac{1}{2\pi} \sum_{j=1}^{n} \int \mathcal{F}f(-u) \frac{Z_j e^{iuZ_j} - E[Z_j e^{iuZ_j}]}{\hat{\varphi}_n(u)} \Delta_{m,k} \mathcal{F}K(u) \, du \right|^2 \\
+ & 2 \left| \frac{1}{2\pi} \sum_{j=1}^{n} \int \mathcal{F}f(-u) \frac{Z_j e^{iuZ_j} - E[Z_j e^{iuZ_j}]}{\hat{\varphi}_n(u)} \Delta_{m,k} \mathcal{F}K(u) \, du \right|^2 \\
= & 2 \left| \hat{\theta}_k - \hat{\theta}_m \right| - \left| \tilde{\theta}_k - \tilde{\theta}_m \right| \right|^2 + 2 \left| \tilde{\theta}_k - \hat{\varphi}_n \right| - \left| \tilde{\theta}_k - \hat{\varphi}_n \right| \right|^2.
\end{align*}
\]

Since

\[
\mathbb{E} \left[ \left| \hat{\theta}_k - \hat{\theta}_m \right| - \left| \tilde{\theta}_k - \tilde{\theta}_m \right| \right] \leq \frac{1}{n} \sigma_{m,k}^2 \ a.s.
\]

and

\[
\left| \frac{1}{2\pi} \int \mathcal{F}f(-u) \frac{Z_j e^{iuZ_j}}{\hat{\varphi}_n(u)} \Delta_{m,k} \mathcal{F}K(u) \, du \right| \leq \frac{4}{\eta} (\log(n \bar{m}_h (k - m))) \frac{1}{2\pi} \int \left| \mathcal{F}f(-u) \right| \Delta_{m,k} \mathcal{F}K(u) \, du
\]

\[
\leq \frac{4\sqrt{n}}{\eta} (\log(n \bar{m}_h (k - m))) \bar{m}_k \ a.s.
\]

the integral version of the classical Bernstein-inequality (see, for example, Dudley (2008)).
The remainder term can be estimated as follows:

\[ \mathbb{E} \left[ \left\{ \left( \tilde{\theta}_k - \tilde{\theta}_m \right) - \left( \tilde{\theta}_k - \tilde{\theta}_m \right) \right\}^2 \bigg| \tilde{\varphi}_n \right] \leq \frac{2048\sqrt{2} \log (n\tilde{x}_{m,k}(k-m))}{n} \exp \left( -\frac{\tilde{\varphi}_n \lambda_{m,k} \tilde{x}_{m,k}}{64\tilde{\sigma}_{m,k}^2} \right) \]

\[ + \frac{32\tilde{\sigma}_{m,k}^2}{n} \exp \left( -\frac{\tilde{\varphi}_n \lambda_{m,k} \tilde{x}_{m,k}}{64\tilde{\sigma}_{m,k}^2} \right) \text{ a.s.} \]

It is important to recall at this point that \( \tilde{\varphi}_n \) and \( \tilde{\varphi}_n \) are independent by construction. Using the fact that, by definition, \( \tilde{\varphi}_n \geq 64 \) and \( \lambda_{m,k} \geq \log \left( \tilde{\sigma}_{m,k}^2 (k-m)^2 \right) \), as well as

\[ \tilde{\lambda}_{m,k} \geq \frac{8}{n} \log (n\tilde{x}_{m,k}(k-m)) \log (n\tilde{x}_{m,k}(k-m))^2 \log (n\tilde{x}_{m,k}(k-m))^2 \],

we can continue by estimating

\[ \frac{32\tilde{\sigma}_{m,k}^2}{n} \exp \left( -\frac{\tilde{\varphi}_n \lambda_{m,k} \tilde{x}_{m,k}}{64\tilde{\sigma}_{m,k}^2} \right) \]

\[ + \frac{2048\sqrt{2} \log (n\tilde{x}_{m,k}(k-m))}{n} \tilde{\sigma}_{m,k}^2 \exp \left( -\frac{\tilde{\varphi}_n \lambda_{m,k} \tilde{x}_{m,k}}{64\tilde{\sigma}_{m,k}^2} \right) \]

\[ - \frac{\lambda_{m,k}}{n} \log (n\tilde{x}_{m,k}(k-m)) \]

\[ + \frac{2048\sqrt{2} \log (n\tilde{x}_{m,k}(k-m))}{n} \tilde{\sigma}_{m,k}^2 \log (n\tilde{x}_{m,k}(k-m))^{-2} \tilde{\sigma}_{m,k}^2 (k-m)^{-2} \]

We have thus shown that, almost surely,

\[ \sum_{\substack{k \geq 0 \\text{ s.t. } k \in \mathbb{N}}} \mathbb{E} \left[ \left\{ \left( \tilde{\theta}_k - \tilde{\theta}_m \right) - \left( \tilde{\theta}_k - \tilde{\theta}_m \right) \right\}^2 \bigg| \tilde{\varphi}_n \right] \leq 2(64 + \frac{2048\sqrt{2}}{n^2}) \]

The remainder term can be estimated as follows:

\[ \mathbb{E} \left[ \left( \tilde{\theta}_k - \tilde{\theta}_m \right) - \left( \tilde{\theta}_k - \tilde{\theta}_m \right) \bigg| \tilde{\varphi}_n \right] \]

\[ \leq \frac{1}{n} \mathbb{E} \left[ \left| \frac{1}{2\pi} \int_{\mathbb{R}} \mathcal{F}(-u) \frac{Z_r^i e^{iuZ_i}}{\tilde{\varphi}_n(u)} \Delta_{m,k} \mathcal{F} K(u) \, du \right|^2 \bigg| \tilde{\varphi}_n \right] \]

\[ \leq \frac{1}{n} \left| \mathbb{E} \left| Z_r^i \bigg| \tilde{\varphi}_n \right| \left( \int \left| \mathcal{F}(-u) \right| \frac{\Delta_{m,k} \mathcal{F} K(u)}{\tilde{\varphi}_n(u)} \, du \right)^2 \right| \tilde{x}_{m,k} \mathbb{E} \left| Z_r^i \bigg| \tilde{\varphi}_n \right| \text{ a.s.} \]
Now, we apply Markov’s inequality to find
\[
\mathbb{E} \left[ |Z|^2 |\hat{\varphi}_n | \right] = \mathbb{E} \left[ |Z|^2 1 \left( |Z| > \frac{4}{\eta} \log (nx_{m,k}(k-m)) \right) \right] |\hat{\varphi}_n |
\leq \mathbb{E} \left[ |Z|^2 \exp \left( \frac{\eta}{2} |Z| \right) \right] |\hat{\varphi}_n | \exp (-2 \log (nx_{m,k}(k-m)))
\leq n^{-\frac{4}{\eta}} \mathbb{E} \left[ \exp (\eta |Z|) \right] x^{-2}_{m,k} (k-m)^{-2} \ \text{a.s.}
\]

where we have used the notation 1(A) instead of 1A. We have thus shown
\[
\sum_{k \geq m, k \in \mathbb{N}} \mathbb{E} \left[ \left( |\hat{\theta}_k - \hat{\theta}_m | - (\theta_k - \theta_m) \right)^2 \right] |\hat{\varphi}_n | \leq \frac{8}{\eta^2} \mathbb{E} \left[ \exp (\eta |Z|) \right] n^{-1} \ \text{a.s.}
\]

It remains to consider the expression appearing in the last line of formula \(5.13\). Let us introduce, for arbitrary \(m \leq k\), the favourable set
\[C(m,k) := \left\{ \forall u \in \mathbb{R} : \left| \frac{1}{\hat{\varphi}_n(u)} - \frac{1}{\varphi(u)} \right|^2 \leq \left( \frac{\lambda_{f_{m,k}}}{\varphi(u)} \right)^2 (w(u))^{-1} \right\},\]

with \(\lambda_{f_{m,k}}\) defined as in Lemma 5.9. We can estimate
\[
\mathbb{E} \left[ \sup_{k > m, k \in \mathbb{N}} \left\{ \left( |\hat{\theta}_k - \hat{\theta}_m | - (\theta_k - \theta_m) \right)^2 - \frac{1}{8} \mathbb{H}^2 (m, k) \right\} \right] \\
\leq \mathbb{E} \left[ \sup_{k > m, k \in \mathbb{N}} \left\{ \left( |\hat{\theta}_k - \hat{\theta}_m | - (\theta_k - \theta_m) \right)^2 - \frac{1}{8} \mathbb{H}^2 (m, k) \right\} + \mathbb{E} \left[ 1 (C(m,k)) \right] \\
+ \mathbb{E} \left[ \sup_{k > m, k \in \mathbb{N}} \left\{ \left( |\hat{\theta}_k - \hat{\theta}_m | - (\theta_k - \theta_m) \right)^2 - \frac{1}{8} \mathbb{H}^2 (m, k) \right\} + 1 (C(m,k)^c) \right].
\]

The Cauchy Schwarz inequality and the fact that \(|\varphi(u)| \leq 1\) imply
\[
\left| (\hat{\theta}_k - \hat{\theta}_m ) - (\theta_k - \theta_m ) \right|^2 \\
= \frac{1}{2\pi} \int \mathcal{F}f(-u)\varphi(u) \left( \frac{1}{\hat{\varphi}_n(u)} - \frac{1}{\varphi(u)} \right) \mathbb{D}_{m,k}(u) |\mathbb{K}(u)| du \\
= \frac{1}{2\pi} \int \mathcal{F}f(-u)\mathcal{F}\varphi(u) \left( \frac{1}{\hat{\varphi}_n(u)} - \frac{1}{\varphi(u)} \right) \varphi(u) \mathbb{D}_{m,k}(u) |\mathbb{K}(u)| du \\
\leq \frac{1}{(2\pi)^2} \left\{ \left| \mathcal{F}f(-u) \right|^2 \left( \int \mathcal{F}f(-u) \right)^2 \left( \frac{1}{\hat{\varphi}_n(u)} - \frac{1}{\varphi(u)} \right)^2 |\varphi(u)| |\mathbb{D}_{m,k}(u) | |\mathbb{K}(u)| du \right)^{\frac{1}{2}} \\
\wedge \left( \left| \mathcal{F}\varphi(u) \right|^2 \int \left| \mathcal{F}f(-u) \right|^2 \left( \frac{1}{\hat{\varphi}_n(u)} - \frac{1}{\varphi(u)} \right)^2 |\varphi(u)|^2 |\mathbb{D}_{m,k}(u) |^2 |\mathbb{K}(u)|^2 du \right)^{\frac{1}{2}} \right\}
\]

The definition of \(C(m,k)\) readily implies that on this set,
\[
\frac{1}{(2\pi)^2} \left( \left| \mathcal{F}\varphi(u) \right|^2 \int \left| \mathcal{F}f(-u) \right|^2 \left( \frac{1}{\hat{\varphi}_n(u)} - \frac{1}{\varphi(u)} \right)^2 |\varphi(u)|^2 |\mathbb{D}_{m,k}(u) |^2 |\mathbb{K}(u)|^2 du \right)^{\frac{1}{2}} \\
\wedge \left( \left| \mathcal{F}f(-u) \right|^2 \int \left| \mathcal{F}f(-u) \right|^2 \left( \frac{1}{\hat{\varphi}_n(u)} - \frac{1}{\varphi(u)} \right)^2 |\varphi(u)|^2 |\mathbb{D}_{m,k}(u) |^2 |\mathbb{K}(u)|^2 du \right)^{\frac{1}{2}} \right) \\
\leq \left( \frac{5}{2} \kappa \right) (\log n)^{1/2} + \lambda_{f_{m,k}} \right)^2 n^{-1} \hat{\sigma}_{m,k}^2.
\]
We use the trivial observation that we always have $x_{f_{m,k}}^2 \leq \tilde{\sigma}_{m,k}^2$ and hence $\lambda_{f_{m,k}}^2 \leq c_1 \tilde{\lambda}_{m,k}^2$ as well as the fact that, by definition,

$$\tilde{H}^2(m, k) \geq \frac{5}{2} c (\log n)^{1/2} + c \sqrt{c_{m,k}}$$

to conclude that the last line is smaller than $\frac{1}{8} \tilde{H}^2(m, k)$. We have thus shown

$$\mathbb{E} \left[ \sup_{k > m, k \in \mathbb{N}} \left\{ \left( \tilde{\theta}_k - \tilde{\theta}_m \right) - (\theta_k - \theta_m) \right\}^2 - \frac{1}{8} \tilde{H}^2(m, k) \right] + 1(C(m, k)) = 0.$$

It remains to show that the remainder term is negligible. The definition of $\frac{1}{\varphi_n}$ implies that we always have $\frac{1}{\varphi_n} \leq \kappa^{-2}(\log n)^{-1} n$. We can thus estimate

$$\left( \tilde{\theta}_k - \tilde{\theta}_m \right) - (\theta_k - \theta_m) \right\}^2 \leq \frac{1}{2\pi^2} \left\{ C_1 \int |\mathcal{F}(-u)|^2 \frac{1}{\varphi_n(u)} - \frac{1}{\varphi(u)} \right\}^2 |\varphi(u)|^2 |\Delta_{m,k} \mathcal{K}(u)|^2 \, du \wedge C_2 \left( \int |\mathcal{F}(-u)| \frac{1}{\varphi_n(u)} - \frac{1}{\varphi(u)} \right| |\varphi(u)||\Delta_{m,k} \mathcal{K}(u)| \, du \right\}^2 \leq \frac{1}{2\pi^2} \left( \kappa^{-1}(\log n)^{-1/2} n^{1/2} + 1 \right)^2 x_{f_{m,k}}^2,$$

with $x_{f_{m,k}}$ defined as in Lemma 5.3. This implies

$$\mathbb{E} \left[ \sup_{k > m, k \in \mathbb{N}} \left\{ \left( \tilde{\theta}_k - \tilde{\theta}_m \right) - (\theta_k - \theta_m) \right\}^2 - \frac{1}{8} \tilde{H}^2(m, k) \right] + 1(C(m, k)^c) \leq \sum_{k > m, k \in \mathbb{N}} \mathbb{E} \left[ \left\{ \left( \tilde{\theta}_k - \tilde{\theta}_m \right) - (\theta_k - \theta_m) \right\}^2 - \frac{1}{8} \tilde{H}^2(m, k) \right] + 1(C(m, k)^c) \leq \sum_{k > m, k \in \mathbb{N}} \frac{1}{2\pi^2} \left( \kappa^{-1}(\log n)^{-1/2} n^{1/2} + 1 \right)^2 x_{f_{m,k}}^2 \mathbb{P} \left( C(m, k)^c \right).$$

(5.14)

Now, Lemma 5.10 and the choice of $\kappa$ imply

$$\mathbb{P} \left( C(m, k)^c \right) = C n^{-2} x_{f_{m,k}} (m - k)^{-2},$$

so the sum appearing in the last line of formula (5.14) is readily negligible. This completes the proof. \hfill \Box

### 5.2.3 Proof of Theorem 4.3

In what follows, let $m^*$ denote the oracle cutoff,

$$m^* := \arg\min_{m \in \mathcal{M}} \left\{ \sup_{k \geq m, k \in \mathcal{M}} |\theta_k - \theta_m|^2 + \text{pen}(m) \right\}.$$

We start by considering the loss on the set $\{ \tilde{m} \leq m^* \}$. We use the estimate

$$\left| \theta - \tilde{\theta}_{\tilde{m}} \right|^2 1(\{ \tilde{m} \leq m^* \}) \leq 2 \left| \theta - \tilde{\theta}_{m^*} \right|^2 1(\{ \tilde{m} \leq m^* \}) + 2 \left| \tilde{\theta}_{m^*} - \tilde{\theta}_{\tilde{m}} \right|^2 1(\{ \tilde{m} \leq m^* \}).$$
Lemma 4.2 and the definition of the penalty term imply
\[ E \left[ |\theta - \hat{\theta}_{m^*}|^2 \right] \leq 2|\theta - \theta_{m^*}|^2 + 2E \left[ \hat{\theta}_{m^*} - \hat{\theta}_{m^*} \right]^2 \leq 2|\theta - \theta_{m^*}|^2 + 2C \text{pen}(m^*). \]

By definition of \( \hat{m} \),
\[ |\hat{\theta}_{m^*} - \hat{\theta}_{m}|^2 1\{\{\hat{m} \leq m^*\}\} \leq \text{pen}(m^*) + \sup_{k > m^*, k \in M} \left\{ |\hat{\theta}_k - \hat{\theta}_{m^*}|^2 - \tilde{H}^2(m^*, k) \right\} + \tilde{H}^2(\hat{m}, m^*) 1\{\{\hat{m} \leq m^*\}\}. \]

We can estimate
\[ \sup_{k > m^*, k \in M} \left\{ |\hat{\theta}_k - \hat{\theta}_{m^*}|^2 - \tilde{H}^2(m^*, k) \right\} \leq 2 \sup_{k > m^*, k \in M} \left\{ |\hat{\theta}_k - \hat{\theta}_{m^*} - (\theta_k - \theta_{m^*})|^2 - \tilde{H}^2(m^*, k) \right\} + 2 \sup_{k > m^*, k \in M} |\theta_k - \theta_{m^*}|^2. \]

Proposition 5.11 readily implies for some positive constant \( C \),
\[ E \left[ \sup_{k > m^*, k \in M} \left\{ 2 \left( |\hat{\theta}_k - \hat{\theta}_{m^*} - (\theta_k - \theta_{m^*})|^2 - \tilde{H}^2(m^*, k) \right) \right\} \right] \leq Cn^{-1}. \]

Next, we observe that, by definition of \( \tilde{H}^2 \) and \( \text{pen} \),
\[ \tilde{H}(\hat{m}, m^*) 1\{\{\hat{m} \leq m^*\}\} \leq \text{pen}(m^*). \]

Since we have chosen \( \kappa \geq 2 \left( \sqrt{4C_1} + \gamma \right) \), we can apply Corollary 5.8 to find that for some positive constant \( C_2 \) depending only on the choice of the constants,
\[ E [\text{pen}(m^*)] \leq C_2 \text{pen}(m^*). \] (5.15)

To do this, we apply the Cauchy-Schwarz inequality to see that
\[ E \left[ \tilde{\lambda}_{m^*}^2 \hat{\sigma}^2 \right] \leq \left( E \left[ \tilde{\lambda}_{m^*}^4 \right] \right)^{1/2} \left( E \left[ \hat{\sigma}_{m^*}^4 \right] \right)^{1/2} \]
and then use Corollary 5.8 to derive that
\[ \left( E \left[ \tilde{\lambda}_{m^*}^4 \right] \right)^{1/2} \left( E \left[ \hat{\sigma}_{m^*}^4 \right] \right)^{1/2} \leq C_2 \tilde{\lambda}_{m^*}^2 \hat{\sigma}_{m^*}^2. \]

Putting the above results together, we have shown that for some positive constant \( C_2 \) depending only on the choice of the constants and some positive constant \( C \) specified in in Proposition 5.11,
\[ E \left[ \left| \theta - \hat{\theta}_{\hat{m}} \right|^2 1\{\{\hat{m} \leq m^*\}\} \right] \leq \max \{ 10, 8C_2 \} \inf_{m \in M} \left\{ |\theta - \theta_{m^*}|^2 + \sup_{k \geq m, k \in M} |\theta_k - \theta_{m^*}|^2 + \text{pen}(m) \right\} + Cn^{-1}. \]

This is the desired result for the expected loss on \( \{\hat{m} \leq m^*\} \).
It remains to consider the loss on the set \( \{\hat{m} > m^*\} \). We use the estimate
\[ |\theta - \hat{\theta}_{\hat{m}}|^2 \leq 3|\theta - \theta_{m^*}|^2 + 3|\theta_{\hat{m}} - \theta_{m^*}|^2 + 3|\theta_{\hat{m}} - \hat{\theta}_{\hat{m}}|^2. \]
First, we clearly have

\[ 3 \left( |\theta - \theta_{m^n}|^2 + |\hat{\theta} - \theta_{m^n}|^2 \right) 1 \{ \{ \hat{m} > m^* \} \} \leq 6 |\theta - \theta_{m_n}|^2 + 6 \sup_{k > m^*, k \in M} |\theta_k - \theta_{m^n}|^2. \]

Next, we can estimate

\[ |\hat{\theta} - \hat{\theta}|^2 1 \{ \{ \hat{m} > m^* \} \} = \sum_{k > m^*, k \in M} |\theta_k - \hat{\theta}_k|^2 + \sum_{k > m^*, k \in M} \hat{\text{pen}}(k) 1 \{ \{ \hat{m} = k \} \}. \]

Another application of Proposition 5.11 gives

\[ \sum_{k > m^*, k \in M} \frac{E}{\{ \{ \hat{m} > m^* \} \}} \left| \theta_k - \hat{\theta}_k \right|^2 - \hat{\text{pen}}(k) + \sum_{k > m^*, k \in M} \hat{\text{pen}}(k) 1 \{ \{ \hat{m} = k \} \} \leq C N^{-1}. \]

Moreover, by definition of \( \hat{m} \), we have on \( \{ \hat{m} = k \} \):

\[ \hat{\text{pen}}(k) \leq \hat{\text{pen}}(m^*) + \sup_{l > m^*, l \in M} \left| (\hat{\theta}_l - \hat{\theta}_{m^*}) - (\theta_l - \theta_{m^*}) \right|^2 - \frac{1}{2} \hat{\text{H}}^2 (m^*, l) \]

By Proposition 5.11 and formula (5.15),

\[ \frac{E}{\{ \{ \hat{m} > m^* \} \}} \left[ \left| (\hat{\theta}_l - \hat{\theta}_{m^*}) - (\theta_l - \theta_{m^*}) \right|^2 - \frac{1}{2} \hat{\text{H}}^2 (m^*, l) \right] \leq C_n^{-1}. \]

Putting the above results together, we have shown

\[ \frac{E}{\{ \{ \hat{m} > m^* \} \}} \left| \theta - \hat{\theta}_m \right|^2 \leq \max\{ 11, 3C_n \} \inf_{m \in M} \left| \theta - \theta_{m_n} \right|^2 + \sup_{k > m^*, k \in M} |\theta_k - \theta_{m^n}|^2 + \text{pen}(m) \] + C_n^{-1},

which is the desired result for \( \{ \hat{m} > m^* \} \). This completes the proof. \( \square \)

References

Denis Belomestny. Spectral estimation of the fractional order of a Lévy process. The Annals of Statistics, 38(1):317–351, 2010.

Denis Belomestny. Statistical inference for time-changed Lévy processes via composite characteristic function estimation. The Annals of Statistics, 39(4):2205–2242, 2011.

Denis Belomestny and Markus Reiß. Spectral calibration of exponential Lévy models. Finance and Stochastics, 10(4):449–474, 2006. doi: 10.1007/s00780-006-0021-5.

Lucien Birgé. An alternative point of view on Lepski’s method. In State of the art in probability and statistics: Festschrift for Willem R. van Zwet, pages 113–133. IMS Lecture Notes-Monograph Series, 1999.
Lucien Birgé and Pascal Massart. From model selection to adaptive estimation. In Festschrift for Lucien Le Cam, pages 55–87. Springer, 1997.

Christina Butucea and Fabienne Comte. Adaptive estimation of linear functionals in the convolution model and applications. Bernoulli, 15(1):69–98, 2009.

T. Tony Cai and Mark G. Low. Adaptive estimation of linear functionals under different performance measures. Bernoulli, 11(2):341–358, 2005a.

T. Tony Cai and Mark G. Low. On adaptive estimation of linear functionals. Annals of Statistics, 33(5):2311–2343, 2005b.

Fabienne Comte and Valentine Genon-Catalot. Nonparametric estimation for pure jump Lévy processes based on high frequency data. Stochastic Processes and their Applications, 119:4088–4123, 2009.

Fabienne Comte and Valentine Genon-Catalot. Nonparametric adaptive estimation for pure jump Lévy processes. Annales de l’Institut Henri Poincaré, 3(46):595–617, 2010.

Fabienne Comte and Valentine Genon-Catalot. Estimation for Lévy processes from high frequency data within a long time interval. Annals of Statistics, 39(2):803–837, 2011.

Fabienne Comte and Claire Lacour. Data-driven density estimation in the presence of additive noise with unknown distribution. Journal of the Royal Statistical Society: Series B, 73:601–627, 2011.

Fabienne Comte, Yves Rosenholc, and Marie-Luce Taupin. Penalized contrast estimator for adaptive density deconvolution. Canadian Journal of Statistics, (34):431–452, 2006.

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

Richard M. Dudley. Uniform central limit theorems. Cambridge Studies in Advanced Mathematics 63. Cambridge: Cambridge University Press, 2008.

Jianqing Fan. On the optimal rates of convergence for nonparametric deconvolution problems. The Annals of Statistics, 19(3):1257–1272, 1991.

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

José E. Figueroa-López and Christian Houdré. Risk bounds for the non-parametric estimation of Lévy processes. IMS Lecture Notes - Monograph Series; High Dimensional Probability IV, 51:96–116, 2006.

Alexander Goldenshluger and Sergei V. Pereverzev. On adaptive inverse estimation of linear functionals in Hilbert scales. Bernoulli, 9(5):783–807, 2003.

Alexander Goldenshluger. On pointwise adaptive nonparametric deconvolution. Bernoulli, (5):907–925, 1999.

Shota Gugushvili. Nonparametric estimation of the characteristic triplet of a discretely observed Lévy process. Journal of Nonparametric Statistics, 21(3):321–343, 2009. doi: 10.1080/10485250802645824.

Shota Gugushvili. Nonparametric estimation for discretely sampled Lévy processes. Annales de l’Institut Henri Poincaré, 48(1):282–307, 2012.

Lothar Jantscher. Distributionen. De Gruyter, 1971.
Jan Johannes. Deconvolution with unknown error distribution. *The Annals of Statistics*, 37 (5a):2301–2323, 2009.

Jan Johannes and Maik Schwarz. Adaptive circular deconvolution by model selection under unknown error distribution. arXiv:0912.1207v2, 2012. To appear in Bernoulli.

Johanna Kappus. *Nonparametric adaptive estimation for discretely observed Lévy processes*. PhD thesis, Humboldt Universität zu Berlin, 2012.

Johanna Kappus and Markus Reiß. Estimation of the characteristics of a Lévy process observed at arbitrary frequency. *Statistica Neerlandica*, 64:314–328, 2010.

Yitzhak Katznelson. *An introduction to harmonic analysis*. Cambridge University Press, 3rd edition, 2004.

Thierry Klein and Emmanuel Rio. Concentration around the mean for maxima of empirical processes. *The Annals of Probability*, 33(3):1060–1077, 2005.

Béatrice Laurent, Carenne Ludeña and Clémente Prieur. Adaptive estimation of linear functionals by model selection. *Electronic Journal of Statistics*, 2:993–1020, 2008.

Oleg V. Lepski. On a problem of adaptive estimation in Gaussian white noise. *Theory of Probability and Applications*, 35:454–466, 1990.

Oleg V. Lepski. Asymptotically minimax adaptive estimation. I. Upper bounds. *Theory of Probability and Applications*, 36(3):682–697, 1991.

Oleg V. Lepski. Asymptotically minimax adaptive estimation. II. Schemes without optimal adaptation. *Theory of Probability and Applications*, 37(3):433–448, 1992.

Pascal Massart. *Concentration Inequalities and Model selection*. Number 1896 in Lecture Notes in Mathematics. Springer, 2003.

Alexander Meister. Density estimation with normal measurement error with unknown variance. *Statistica Sinica*, 16:195–211, 2006.

Alexander Meister. *Deconvolution Problems in Nonparametric Statistics*. Number 193 in Lecture Notes in Statistics. Springer, 2009.

Michael Neumann. Deconvolution from panel data with unknown error distribution. *Journal of Multivariate Analysis*, 98:1955–1968, 2006.

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

Michael H. Neumann. On the effect of estimating the error density in nonparametric deconvolution. *Journal of Nonparametric Statistics*, 7(4):307–330, 1997. doi: 10.1080/10485259708832708.

Richard Nickl and Markus Reiβ. A Donsker Theorem for Lévy measures. arXiv:1201.0590v2, 2012. To appear in Journal of Functional Analysis.

Steven Orey. On continuity properties of infinitely divisible functions. *The Annals of Mathematical Statistics*, 39(3):936–937, 1968.

Walter Rudin. *Functional analysis*. McGraw-Hill, 2nd edition, 1991.

Ken-Iti Sato. *Lévy processes and infinitely divisible distributions*. Cambridge University Press, 1999.

Alexandre B. Tsybakov. *Introduction à l estimation non-parametrique*. Number 41 in Mathematiques et Applications. Springer, 2004.