Estimation and calibration of Lévy models via Fourier methods

Denis Belomestny and Markus Reiß

Abstract In this chapter we discuss different aspects of statistical estimation for Lévy-based processes based on low-frequency observations. In particular, we consider the estimation of the Lévy triplet and the Blumenthal-Getoor index in Lévy and time-changed Lévy models. Moreover, a calibration problem in exponential Lévy models based on option data is studied. The common feature of all these statistical problems is that they can be conveniently formulated in the Fourier domain. We introduce a general spectral estimation/calibration approach that can be applied to these and many other statistical problems related to Lévy processes. On the theoretical side, we provide a comprehensive convergence analysis of the proposed algorithms and address each time the question of optimality.

AMS Subject Classification 2000:

Primary:	60G10 , 60G70 , 60J10
Secondary:	91B28, 91B84

1 Introduction

Statistics for jump processes and especially Lévy processes has been attracting a lot of attention recently. This is on one hand due to the more and more

---

Denis Belomestny
Duisburg-Essen University, Faculty of Mathematics Thea-Leymann-Str. 9, D-45127 Essen, Germany, e-mail: `denis.belomestny@uni-due.de`; homepage: `http://www.uni-due.de/~hm0124`

Markus Reiß
Humboldt-Universität zu Berlin, Institut für Mathematik, Unter den Linden 6, D-10099 Berlin, Germany, e-mail: `mreiss@mathematik.hu-berlin.de`; homepage: `http://www.mathematik.hu-berlin.de/~mreiss`

refined understanding of these processes in probability theory and to the mathematical challenges posed by basic inference questions from mathematical statistics. On the other hand, jump models are very popular in diverse applications, notably in physics, biology, economics and finance. This is not surprising if one takes into account their simplicity and analytic tractability as well as the ability to reproduce many features observed in empirical data.

The problem of nonparametric statistical inference for jump processes has a long history with early works by Rubin and Tucker [51] and Basawa and Brockwell [6]. Currently, the field is developing quickly and there are two major strands of recent literature dealing with statistical inference for Lévy processes or more generally for semi-martingale models. The first type of literature considers the so-called high-frequency or infill asymptotics, where the corresponding estimates are studied under the assumption that the distance between observation times tends to zero, see Aït-Sahalia and Jacod [3] for a recent review. In the second strand of literature the frequency of observations is assumed to be fixed or to converge slowly to infinity (the so-called low-frequency setup) and the asymptotic analysis is performed under the premiss that the observational horizon tends to infinity. Clearly, none of the above asymptotic hypotheses can be perfectly realised on real data and they can only serve as a convenient approximation, as in practice the frequency of observations and the horizon are always finite. Here, we concentrate on the low-frequency setting which is significantly harder since the statistician has no access to the underlying continuous-time dynamics. Moreover, we shall adopt a general nonparametric point of view where the quantities of interest like the jump measure are only assumed to possess certain regularity properties, but no finite-dimensional parameter model is imposed. This more agnostic point of view is often essential to exclude or at least reduce errors due to model misspecification, which are not transparent within a parametric statistical analysis.

If a Lévy process is only observed at discrete time points, we do not know how many jumps have occurred between two observations and to what extent the jumps are superposed by regular continuous dynamics of a diffusive component. In this case we therefore observe the jumps only indirectly and it turns out that already in simple cases estimating the distribution of jump sizes is a complex problem which has all the difficulties of a nonparametric deconvolution problem and is in addition inherently nonlinear. Another case of major interest is when the underlying jump process is not observed, but only derived data are available. This is most pronounced in financial derivative products on assets or term structures, where the underlying risk-neutral pricing model is often supposed to allow for jumps and only information about the derived product (e.g. an option) is at our disposal. This calibration approach leads to a typical statistical inverse problem, which is nonlinear and ill-posed in the sense that additional regularisation is necessary and nonparametric convergence rates are much slower. Due to the special structure of Lévy processes the direct application of standard likelihood based approaches is not possible, since most Lévy processes do not have densities in closed form. In view of the well-known Lévy-Khintchine formula, a natural approach towards statistical inference for Lévy processes is a spectral method working with the characteristic function in the Fourier domain.

Such a spectral approach for nonparametric estimation was first introduced in Belomestny and Reiß [10] in the context of non-parametric calibration of the Lévy triplet to option prices, but in fact it is very generally applicable. In particular, it has been successfully adopted to the case of low-frequency observations. Not aiming at a full literature review, let us point out the works by Gugushvili [34] and by Figueroa-Lopez and Houdré [32] on nonparametric jump density estimation from low and high-frequency observations, respectively. The surprisingly difficult problem of adaptive nonparametric estimation (i.e., a completely data-driven choice of tuning parameters) has been addressed by Kappus [42] for low-frequency observations, building on the high-frequency model selection approach by Comte and Genon-Catalot [20]. Confidence intervals and bands for the Lévy triplet involving a jump density have been constructed by Figueroa-Lopez [31] and Söhl [54] for high-frequency and option price observations, respectively. The power of the spectral approach for both, high and low frequency observations, has been demonstrated for the important problem of nonparametric testing by Reiß [50], which shows also the close relationship with the so realized Laplace transform approach by Todorov and Tauchen [57] for high-frequency observations. The natural question whether the (generalised) distribution function of the Lévy measure allows for a Donsker-type theorem has been considered by Nickl and Reiß [49], using advanced theory for Fourier multipliers and smoothed empirical processes. In Trabs [58] semiparametric efficiency for this estimation is established. The spectral estimation method has found several applications in finance, see e.g. Belomestny and Schoenmakers [11] for Libor model calibration.

In the next section we briefly review the main facts about Lévy and more general jump processes that will be fundamental for the statistical methodology developed subsequently. For the spectral approach the empirical characteristic function, viewed as a process in the frequency argument, is a fundamental object and in Section 3 we present its main theory, in particular uniform convergence results based on exponential inequalities and entropy arguments. The basic estimation method for the Lévy triplet based on low-frequency observations is presented in Section 4. In particular, the error decomposition and the upcoming bias-variance dilemma are discussed in detail and minimax convergence rates are derived. Section 5 then introduces the methodology to establish lower bounds on the error and demonstrates that the spectral estimators are indeed rate-optimal. While the estimator in Section 4 was designed for finite jump intensity only, Section 6 reveals its quite natural behaviour under general jump measures. Moreover, an approach for general Lévy triplets which results in the estimation of the Lévy measure in a weak (negative Sobolev) norm is proposed. Following these ideas further, an estimator of the Blumenthal-Getoor index, which measures the (often infinite) activity of small jumps, is constructed and analysed in Section 7. Then in Section 8 the spectral estimation method is extended to the case of time-changed Lévy processes, often used in applications. The extension to option data is presented in Section 9, which also reports results from real data (DAX options). Finally, Section 10 points out further directions of research where still many open questions exist.

2 Lévy and related processes

In this section we gather some basic results on Lévy and related processes, most of which can be found e.g. in Sato [52].

2.1 Lévy processes

Definition 2.1. *An $\mathbb{R}^d$-valued process $X = (X_t, t \geq 0)$ defined on a filtered probability space $(\Omega, \mathfrak{F}, (\mathfrak{F}_t)_{t\geq 0}, \mathbb{P})$ is called a Lévy process if it is $(\mathfrak{F}_t)$-adapted and has the following properties:*

(a) *X is continuous in probability, i.e., for fixed $u \geqslant 0$, $\mathbb{P}(|X_t - X_u| > \varepsilon) \to 0$ holds as $t \to u$ for all $\varepsilon > 0$.*
(b) $\mathbb{P}(X_0 = 0) = 1$.
(c) *For $0 \leq s \leq t$, $X_t - X_s$ is equal in distribution to X_{t-s}.*
(d) *For $0 \leq s \leq t$, $X_t - X_s$ is independent of $\mathfrak{F}_s$.*

Definition 2.2. *A Lévy measure on $\mathbb{R}^d$ is a σ-finite measure ν on $\mathbb{R}^d$ such that*

$$\int_{\mathbb{R}^d} (1 \wedge |x|^2)\nu(dx) < \infty. \tag{2.1}$$

Sometimes we shall assume $\nu(\{0\}) = 0$ (no jumps of size zero), but usually the integrands for ν will vanish at zero anyway. Fundamental for the understanding of Lévy processes and in particular for their statistical inference is the celebrated Lévy-Khintchine formula.

Proposition 2.3 (Lévy-Khintchine Formula). *Let X be a Lévy process taking values in $\mathbb{R}^d$. Then for each $t \geq 0$ the law of X_t is infinitely divisible and its characteristic function admits the representation*

$$\mathbb{E}[e^{\mathrm{i}\langle u, X_t\rangle}] = e^{t\psi(u)}, \quad u \in \mathbb{R}^d, \tag{2.2}$$

where the characteristic exponent $\psi(u)$ is given by the Lévy-Khintchine formula

$$\psi(u) = \mathtt{i}\langle u,\gamma\rangle - \frac{1}{2}\langle u,\sigma^2 u\rangle + \int_{\mathbb{R}^d} (e^{\mathtt{i}\langle x,u\rangle} - 1 - \mathtt{i}\langle x,u\rangle\,\mathbf{1}(|x|\le 1))\,\nu(dx) \quad (2.3)$$

where $\gamma \in \mathbb{R}^d$, $\sigma^2 \in \mathbb{R}^{d\times d}$ *is a positive semi-definite matrix, and* ν *is a Lévy measure on* $\mathbb{R}^d$.

The quantity (γ,σ^2,ν) is called the *characteristic triplet* of X.

Remark 2.4. Under some additional assumptions on ν the Lévy-Khintchine formula (2.3) has a simpler form.

(a) If $\int_{\mathbb{R}^d}(1\wedge|x|)\nu(dx) < \infty$ holds, then (2.3) reduces to

$$\psi(u) = t\left(\mathtt{i}\langle u,\gamma_0\rangle - \frac{1}{2}\langle u,\sigma^2 u\rangle + \int_{\mathbb{R}^d}(e^{\mathtt{i}\langle x,u\rangle} - 1)\,\nu(dx)\right) \quad (2.4)$$

with $\gamma_0 = \gamma - \int_{\mathbb{R}^d} x\mathbf{1}(|x|\le 1)\,\nu(dx)$.

(b) If $\int_{\mathbb{R}^d}|x|\mathbf{1}(|x|>1)\nu(dx) < \infty$ holds, we can rewrite (2.3) in the form

$$\psi(u) = \left(\mathtt{i}\langle u,\gamma_1\rangle - \frac{1}{2}\langle u,\sigma^2 u\rangle + \int_{\mathbb{R}^d}(e^{\mathtt{i}\langle x,u\rangle} - 1 - \mathtt{i}\langle x,u\rangle)\,\nu(dx)\right)$$

with $\gamma_1 = \gamma + \int_{\mathbb{R}^d} x\mathbf{1}(|x|>1)\,\nu(dx)$ and we have $\mathbb{E}[X_t] = \gamma_1 t$.

(c) If $d=1$ and $\int_{\mathbb{R}} x^2\nu(dx) < \infty$ holds, then the so-called Kolmogorov representation holds:

$$\begin{aligned}\psi(u) &= \mathtt{i}u\tilde{\gamma} - \frac{\sigma^2u^2}{2} + \int_{\mathbb{R}} \frac{e^{\mathtt{i}xu} - 1 - \mathtt{i}xu}{x^2}\,\widetilde{\nu}(dx)\\ &= iu\tilde{\gamma} + \int_{\mathbb{R}} \frac{e^{\mathtt{i}xu} - 1 - \mathtt{i}xu}{x^2}\,\nu_\sigma(dx) \quad (2.5)\end{aligned}$$

with $\widetilde{\nu}(dx) = x^2\nu(dx)$ and $\nu_\sigma(dx) = \widetilde{\nu}(dx) + \sigma^2\delta_0(dx)$, using at $x=0$ the continuous extension of the integrand to $-u^2/2$ in the second representation. Here the first two moments take a particularly nice form: $\mathbb{E}[X_t] = \tilde{\gamma}t$, $\mathrm{Var}(X_t) = (\sigma^2 + \widetilde{\nu}(\mathbb{R}))t = \nu_\sigma(\mathbb{R})t$.

2.2 Affine processes

The class of affine processes extends the class of Lévy processes and encompasses many interesting processes, e.g. used in finance. It is also defined via properties of the characteristic function.

Definition 2.5. *The process* $(X_t, t\ge 0)$ *is an affine process if it is a stochastically continuous, time-homogenous Markov process with state space* $\mathscr{D} \subseteq \mathbb{R}^d$, *such that the conditional characteristic function of* X_s *given* X_0 *is an exponentially affine function of the initial state* X_0*:*

$$\varphi(u|s,x) := \mathbb{E}\left[e^{\mathrm{i}\langle u,X_s\rangle}\middle|X_0=x\right] = e^{\psi_0(u,s)+\langle x,\psi_1(u,s)\rangle},\quad u\in\mathbb{R}^d, \tag{2.6}$$

where ψ_0 and ψ_1 take values in $\mathbb{C}$ and $\mathbb{C}^d$, respectively.

The affine process X_t is called regular, if the derivatives

$$F_0(u) := \left.\frac{\partial\psi_0(u,s)}{\partial s}\right|_{s=0},\quad F_1(u) := \left.\frac{\partial\psi_1(u,s)}{\partial s}\right|_{s=0} \tag{2.7}$$

exist and are continuous at $u=0$. As was recently shown by Keller-Ressel *et al.* [44], any affine process is, in fact, regular. The following theorem provides the characterization of affine processes and is proved in Duffie et al. [25].

Theorem 2.6. *If X is an affine process, then the complex valued functions ψ_0 and ψ_1 satisfy the (generalized) Riccati equations*

$$\frac{\partial\psi_0(u,s)}{\partial s} = F_0(\psi_1(u,s)),\quad \psi_0(u,0)=0, \tag{2.8}$$
$$\frac{\partial\psi_1(u,s)}{\partial s} = F_1(\psi_1(u,s)),\quad \psi_1(u,0)=u, \tag{2.9}$$

where

$$F_0(z) = -\frac{1}{2}\langle z,\sigma_0 z\rangle + \mathrm{i}\langle z,\gamma_0\rangle + \int_{\mathscr{D}\setminus\{0\}}\left(e^{\mathrm{i}\langle z,u\rangle}-1-\mathrm{i}\langle z,u\rangle\mathbf{1}(|u|\le 1)\right)\nu_0(du),$$
$$F_{1,j}(z) = -\frac{1}{2}\langle z,\sigma_j z\rangle + \mathrm{i}\langle z,\gamma_j\rangle + \int_{\mathscr{D}\setminus\{0\}}\left(e^{\mathrm{i}\langle z,u\rangle}-1-\mathrm{i}\langle z,u\rangle\mathbf{1}(|u|\le 1)\right)\nu_j(du)$$

for $j=1,\dots,d$. Here $\sigma_j^2\in\mathbb{R}^{d\times d}$, $\gamma_j\in\mathbb{R}^d$, and ν_j is a Lévy measure on $\mathbb{R}^d$.

Under some admissibility conditions a regular affine process X is a Feller process in the domain $\mathscr{D}=\mathbb{R}^m\times\mathbb{R}_+^{d-m}$ (see Duffie *et al.* [26]), where the function F_0 corresponds to the state-independent part of the infinitesimal generator and F_1 is related to the state-dependent one.

Spectral estimation for affine processes is treated by Belomestny [8] in specific cases, but the general methodology is far from understood, see the discussion in Section 10 below.

2.3 Time-changed Lévy processes

Let X_t be a d-dimensional Lévy process with characteristic exponent $\psi(u)$. Let furthermore $t\to\mathscr{T}(t)$, $t\ge 0$ be an increasing right-continuous process with left limits such that $\mathscr{T}(0)=0$ and for each fixed t, the random variable $\mathscr{T}(t)$ is a stopping time with respect to the filtration $(\mathfrak{F}_t)_{t\ge 0}$. Suppose furthermore that $\mathscr{T}(t)$ is finite $\mathbb{P}$-a.s. for all $t\ge 0$ and that $\mathscr{T}(t)\to\infty$ as $t\to\infty$. Then the family $(\mathscr{T}(t))_{t\geq 0}$ defines a random time change. The d-dimensional process $Y_t := X_{\mathscr{T}(t)}$ is called *time-changed Lévy process.*

We compute the characteristic function of Y_t in the case of independent processes $(\mathscr{T}(t))$ and (X_t) as

$$\varphi_Y(u|t) = \mathbb{E}\left[e^{\mathrm{i}\langle u, X_{\mathscr{T}(t)}\rangle}\right] = \mathscr{L}_t(-\psi(u)), \tag{2.10}$$

where $\mathscr{L}_t$ is the Laplace transform of $\mathscr{T}(t)$:

$$\mathscr{L}_t(\lambda) = \mathbb{E}\left[e^{-\lambda \mathscr{T}(t)}\right].$$

As we shall see, the formula (2.10) plays a crucial role in statistical inference for time-changed processes based on low-frequency data.

Let us look at some examples. If $\mathscr{T}(t)$ is a Lévy process, then Y_t is another Lévy process. A more general situation is when $\mathscr{T}(t)$ is modeled by a non-decreasing semimartingale

$$\mathscr{T}(t) = b_t + \int_0^t \int_0^\infty y\rho(dy, ds),$$

where b is a drift and ρ is the counting measure of jumps in the time change. Often, the case $b_t = 0$ is considered with locally deterministic time changes

$$\mathscr{T}(t) = \int_0^t Z_{s-}\, ds, \tag{2.11}$$

where Z is a non-negative jump process (instantaneous activity rate). When X_t is Brownian motion and ρ is proportional to the instantaneous variance rate of the Brownian motion, then Y_t is a pure jump Lévy process with the Lévy measure proportional to ρ.

3 Empirical characteristic functions and processes

In this section we consider arbitrary i.i.d. random variables $(X_k)_{k\geqslant 1}$ and study their empirical characteristic function $\varphi_n(u)$ as a process in $u \in \mathbb{R}^d$. Because of their independent and identically distributed increments, we shall later apply the results to the observations of Lévy processes. The study of the empirical characteristic function as a process in the frequency variable has attracted considerable interest in the 1970s and 80s when Feuerverger and Mureika [29] have shown its usefulness for statistical questions and then the developing general theory of empirical processes was used to understand better which kind of uniform convergence on the real line can hold for the empirical characteristic process, see e.g. Csörgö [23]. Here we shall prove a basic uniformity result in detail and then present a general result which is most useful for our purposes.

Definition 3.1. *The empirical characteristic function of i.i.d. $\mathbb{R}^d$-valued random variables $X_1,\dots,X_n$ is given by*

$$\varphi_n(u)=\frac{1}{n}\sum_{k=1}^{n}e^{\mathrm{i}\langle u,X_k\rangle},\quad u\in\mathbb{R}^d,$$

and the empirical characteristic process is

$$u\mapsto\mathscr{C}_n(u)=\sqrt{n}(\varphi_n(u)-\varphi(u))\ \text{with}\ \varphi(u)=\mathbb{E}[e^{\mathrm{i}\langle u,X_1\rangle}].$$

For complex-valued random variables Z_i we define $\operatorname{Cov}_{\mathbb{C}}(Z_1,Z_2)=\mathbb{E}[Z_1\bar{Z}_2]-\mathbb{E}[Z_1]\overline{\mathbb{E}[Z_2]}$ and $\operatorname{Var}_{\mathbb{C}}(Z_1)=\mathbb{E}[|Z_1-\mathbb{E}[Z_1]|^2]$. Standard calculations in the scalar case $d=1$ yield $\mathbb{E}[\varphi_n(u)]=\varphi(u)$, $\operatorname{Cov}_{\mathbb{C}}(\varphi_n(u),\varphi_n(v))=\frac{1}{n}(\varphi(u-v)-\varphi(u)\varphi(-v))$, $\operatorname{Var}_{\mathbb{C}}(\varphi_n(u))=\frac{1}{n}(1-|\varphi(u)|^2)\leqslant\frac{1}{n}$. The standard multivariate central limit theorem therefore shows the convergence of the empirical characteristic process

$$\mathscr{C}_n\xrightarrow{fidi}\Gamma$$

(*fidi* denoting weak convergence of the finite-dimensional distributions) with a centred complex-valued Gaussian process $\Gamma(u)$ satisfying $\operatorname{Cov}_{\mathbb{C}}(\Gamma(u),\Gamma(v))=\varphi(u-v)-\varphi(u)\varphi(-v)$.

For our purposes we shall need a much stronger result, which is in particular uniform in the frequency $u\in\mathbb{R}^d$. This will be achieved by exponential inequalities and we start with a straight-forward, but slightly sub-optimal result using Hoeffding's inequality [35].

Proposition 3.2 (Hoeffding's inequality (1963)). *Suppose the real-valued and centred random variables $Y_1,\dots,Y_n$ are i.i.d. and set $S_n=\sum_{k=1}^n Y_k$. If there exists a deterministic number R with $|Y_1|\leqslant R$ almost surely, then*

$$\forall\kappa>0:\quad \mathbb{P}(|S_n|\geqslant\kappa)\leqslant 2\exp\Big(-\frac{\kappa^2}{2nR^2}\Big)$$

Proposition 3.3. *For i.i.d. random vectors $(X_k)_{k\geqslant1}$ in $\mathbb{R}^d$ with $X_k\in L^1$ and any constant $R>8\sqrt{d}$ the empirical characteristic process satisfies uniformly in $n\in\mathbb{N}$ and $K\geqslant1$*

$$\mathbb{P}\Big(\max_{u\in[-K,K]^d}|\mathscr{C}_n(u)|\geqslant R\sqrt{\log(nK^2)}\Big)\leq C(\sqrt{n}K)^{(64d-R^2)/(64d+64)}$$

for some constant C depending on d and $\mathbb{E}[|X_1|]$ only.

Proof. We consider the real part first and set $S_n(u)=\sum_{k=1}^n(\cos(\langle X_k,u\rangle)-\mathbb{E}[\cos(\langle X_k,u\rangle)])$. Then $S_n(u)$ is for each $u\in\mathbb{R}^d$ a sum of i.i.d. centred random variables, bounded by 2, and Hoeffding's inequality yields

$$\mathbb{P}(|S_n(u)| \geqslant \kappa/2) \leqslant 2\exp\Big(-\frac{(\kappa/2)^2}{8n}\Big).$$

We consider for some $J = J(n)$ the $(2J)^d$ equidistant grid points $u_j = jK/J$, $j \in G_J^d := \{-J+1, -J+2, \dots, 0, 1, \dots, J\}^d$ on the cube $[-K,K]^d$ and obtain

$$\mathbb{P}\Big(\max_{j\in G_J^d}|S_n(u_j)| \geqslant \kappa/2\Big) \leqslant \sum_{j\in G_J^d} 2\exp\Big(-\frac{(\kappa/2)^2}{8n}\Big) = 2(2J)^d \exp\Big(-\frac{(\kappa/2)^2}{8n}\Big).$$

For arbitrary $u, v \in \mathbb{R}^d$ we have

$$|\cos(\langle X_k, u\rangle) - \cos(\langle X_k, v\rangle)| \leqslant |X_k||u-v|.$$

From $\mathbb{E}[|X_k|] < \infty$ we infer $|S_n(u) - S_n(v)| \leqslant \sum_{k=1}^n (|X_k| + \mathbb{E}[|X_k|])|u-v|$ and because of $\max_{u\in[-K,K]^d} \min_j |u-u_j| \leqslant \sqrt{d}K/J$

$$\mathbb{P}\Big(\max_{u\in[-K,K]^d}|S_n(u)| \geqslant \kappa\Big) \leqslant \mathbb{P}\Big(\max_{j\in G_J^d}|S_n(u_j)| + \sum_{k=1}^n (|X_k| + \mathbb{E}[|X_k|])\sqrt{d}KJ^{-1} \geqslant \kappa\Big).$$

By Markov's inequality we obtain

$$\begin{aligned}
&\mathbb{P}\Big(\max_{u\in[-K,K]^d}|S_n(u)| \geqslant \kappa\Big)\\
&\leqslant \mathbb{P}\Big(\max_{j\in G_J^d}|S_n(u_j)| \geqslant \kappa/2\Big) + \mathbb{P}\Big(\sum_{k=1}^n (|X_k| + \mathbb{E}[|X_k|])\sqrt{d}KJ^{-1} \geqslant \kappa/2\Big)\\
&\leqslant 2(2J)^d \exp\Big(-\frac{(\kappa/2)^2}{8n}\Big) + \sqrt{d}KJ^{-1}(\kappa/2)^{-1}\sum_{k=1}^n \mathbb{E}[|X_k| + \mathbb{E}[|X_k|]]\\
&= 2^{d+1}J^d \exp\Big(-\frac{\kappa^2}{32n}\Big) + 4\sqrt{d}nKJ^{-1}\kappa^{-1}\mathbb{E}[|X_k|].
\end{aligned}$$

The choice $J = (nK/\kappa)^{1/(d+1)} \exp(\kappa^2/32(d+1)n)$ yields the order

$$\mathbb{P}\Big(\max_{u\in[-K,K]^d}|S_n(u)| \geqslant \kappa\Big) \leq C(nK/\kappa)^{d/(d+1)} \exp\Big(-\frac{\kappa^2}{32(d+1)n}\Big)$$

with $C = 2^{d+1} + 4\sqrt{d}\mathbb{E}(|X_1|)$. For $R > 8\sqrt{d}$ and $nK^2 \to \infty$ we arrive at

$$\begin{aligned}
\mathbb{P}\Big(\max_{u\in[-K,K]}|S_n(u)| \geqslant \tfrac{R}{2}\sqrt{n\log(nK^2)}\Big) &\leq C(\sqrt{n}K)^{d/(d+1)} \exp\Big(-\frac{R^2\log(nK^2)}{128(d+1)}\Big)\\
&\leq C(\sqrt{n}K)^{d/(d+1)-R^2/(64(d+1))}.
\end{aligned}$$

An analogous bound for the imaginary part of φ_n then yields the result due to $\{|\varphi_n - \varphi| \geqslant R\} \subseteq \{|\mathrm{Re}(\varphi_n - \varphi)| \geqslant R/2\} \cup \{|\mathrm{Im}(\varphi_n - \varphi)| \geqslant R/2\}$. □

The result implies that the empirical characteristic function converges uniformly on compact intervals in probability (or even in L^p, $p \geqslant 1$) to the true characteristic function with rate $(\log(n)/n)^{1/2}$. Using the theory of empirical processes, in particular a bracketing entropy argument, it is possible to improve the rate to $1/n^{1/2}$ and to obtain also a bound for any derivative and on the entire real axis.

Based on the proof in Neumann and Reiß [48], the following theorem is derived in Kappus and Reiß [42].

Theorem 3.4. *Let X be a one-dimensional Lévy process with finite $(2k+\gamma)$-th moment and choose $w(u) = (\log(e + |u|))^{-1/2-\delta}$ for some constants $\gamma, \delta > 0$ and $k \in \mathbb{N}_0$. Then for the k-th derivative $\mathscr{C}_{n,\Delta}^{(k)}$ of the characteristic process*

$$\mathscr{C}_{n,\Delta}(u) = \sqrt{n}\Big(\frac{1}{n}\sum_{k=1}^{n} e^{\mathtt{i}u(X_{k\Delta}-X_{(k-1)\Delta})} - \mathbb{E}[e^{\mathtt{i}uX_\Delta}]\Big), \quad u \in \mathbb{R}, \Delta > 0,$$

we have

$$\sup_{n\geqslant 1, \Delta \leqslant 1} \Delta^{-(k\wedge 1)/2}\mathbb{E}\left[\sup_{u\in\mathbb{R}}|\mathscr{C}_{n,\Delta}^{(k)}(u)|w(u)\right] < \infty.$$

From this bound in the mean it is nowadays standard to derive more powerful uniform concentration results of the empirical characteristic process, using Talagrand's concentration inequality, see the nice exposition by Massart and Picard [47] or Chapter "Adaptive Estimation for Lévy processes" by F. Comte and V. Genon-Catalot in this book.

4 Spectral estimation of the Lévy triplet in the finite intensity case

4.1 Basic ideas

Since the characteristics of a Lévy process X appear linearly in the characteristic exponent of X, it is reasonable to work in the Fourier domain to estimate them.

The main idea of the spectral approach can be well illustrated in the case of one-dimensional Lévy processes with finite Lévy measures. We have due to the Lévy-Khintchine formula for the characteristic exponent of X :

$$\psi(u) = \mathtt{i}u\gamma - \frac{1}{2}u^2\sigma^2 - \lambda + \mathscr{F}[\nu](u), \tag{4.1}$$

where $\lambda = \int \nu(dx) < \infty$ is the jump intensity and $\mathscr{F}[\nu] = \int_{\mathbb{R}} e^{iux}\nu(dx)$ stands for the Fourier transform of ν. If ν is absolutely continuous with an absolutely integrable density, then by the Riemann-Lebesgue lemma (see [43], p. 43)

$\mathscr{F}[\nu](u) \to 0$ as $|u| \to \infty$, and consequently $\psi(u)$ can be viewed, at least for large $|u|$, as a second order polynomial with the coefficients $(-\lambda, \mathtt{i}\gamma, -\sigma^2/2)$. So, the parametric part of the Lévy triplet can be approximated via the solution of the following optimisation problem

$$\inf_{(\sigma^2,\gamma,\lambda)} \int_{\{|u|>A\}} w(u) \left|\psi(u) - \mathtt{i}u\gamma + \frac{1}{2}u^2\sigma^2 + \lambda\right|^2 du$$

for some nonnegative weight function w and a large $A > 0$.

Of course, the characteristic exponent ψ needs to be estimated from the data. This can be conveniently done via a plug-in estimator based on the empirical characteristic function. Let $\Delta > 0$ be fixed and let $X_0, X_\Delta, \ldots, X_{n\Delta}$ be $n+1$ equidistant observations of the Lévy process X. Define

$$\varphi_n(u) := \frac{1}{n}\sum_{j=1}^{n} e^{\mathtt{i}u(X_{j\Delta} - X_{(j-1)\Delta})},$$

and set

$$\psi_n(u) = \Delta^{-1} \log \varphi_n(u),$$

where the branch of the complex logarithm is taken in such a way that ψ_n is continuous on $(-x_{0,n}, x_{0,n})$ with $\psi_n(0) = 0$ and $x_{0,n}$ being the first zero of φ_n. In fact, since φ does not vanish on $\mathbb{R}$, we have $x_{0,n} \overset{a.s.}{\to} \infty$ (see [61], p. 156).

Figure 1 shows the plot of $\mathrm{Re}(\psi(u))$ and its polynomial approximation $-u^2\sigma^2/2 - \lambda$ in the so called Merton model, where the process is a compound Poisson process plus Brownian motion. More specifically, the triplet is $\nu(dx) = \frac{10}{\sqrt{2\pi}} e^{-x^2/2}\, dx$, $\gamma = 0$ and $\sigma = 1$. The figure shows also 10 typical estimates ψ_n in the case of $n = 500$ and $\Delta = 0.1$. As can be seen, the deviation of ψ_n from ψ becomes larger as u increases. The reason for this is that the variance of ψ_n increases exponentially in u if $\sigma > 0$, since $|\varphi(u)|$ decreases exponentially with u and $\log \varphi_n - \log \varphi \approx (\varphi - \varphi_n)/\varphi$. This means that we should restrict the range of the frequencies u used to estimate $(\sigma^2, \gamma, \lambda)$. The latter task can be accomplished by using weight functions supported on $[0, U_n]$, with a sequence $U_n, \to \infty$, $n \to \infty$. These weight functions naturally appear as the solution of a certain optimisation problem. Let

$$\widetilde{w}^{U_n}(u) := (1/U_n)\, \widetilde{w}\,(u/U_n)\,,$$

where $\widetilde{w}(u)$ is a continuous function, supported on the interval $[0,1]$ with $\widetilde{w}(u) > 0$ on $(0,1)$. Consider the optimisation problem

$$(\sigma_n^2, \lambda_n) := \underset{(\sigma^2,\lambda)}{\operatorname{argmin}} \int_0^\infty \widetilde{w}^{U_n}(u)(\mathrm{Re}\,\psi_n(u) + \sigma^2 u^2/2 + \lambda)^2\, du. \tag{4.2}$$

By straight-forward calculations, the solution σ_n^2 of (4.2) is found to be

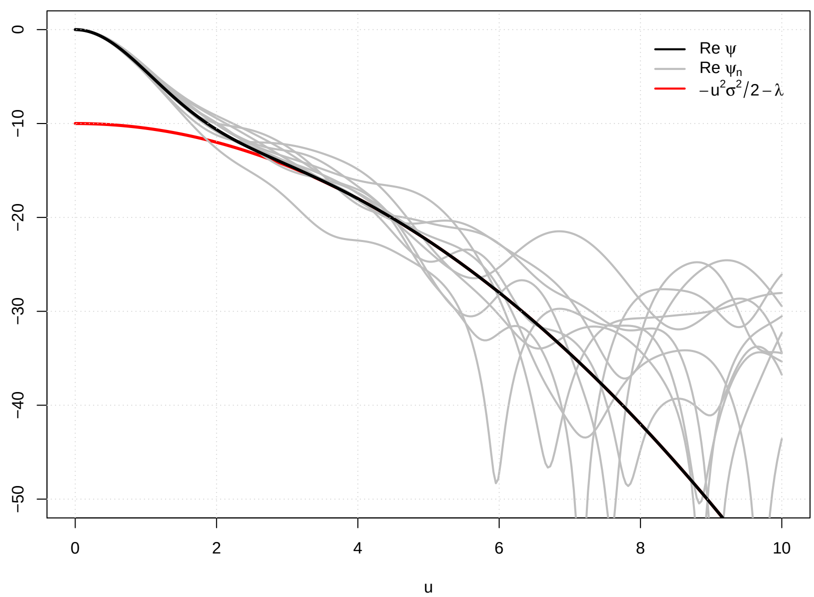

Fig. 1 The real part of the characteristic exponent $\psi(u)$ (black solid line) together with 10 realisations of its estimate ψ_n (grey dashed line) and the polynomial $-u^2\sigma^2/2-\lambda$ (red solid line) for the compound Poisson process with normally distributed jump sizes.

$$\sigma_n^2 = \int_0^\infty w_\sigma^{U_n}(u)\,\mathrm{Re}\,\psi_n(u)\,du, \tag{4.3}$$

where

$$w_\sigma^{U_n}(u) := \widetilde{w}^{U_n}(u)\frac{2\left[\left(\int_0^\infty \widetilde{w}^{U_n}(s)\,ds\right)u^2 - \int_0^\infty \widetilde{w}^{U_n}(s)s^2\,ds\right]}{\left(\int_0^\infty \widetilde{w}^{U_n}(s)s^2\,ds\right)^2 - \int_0^\infty \widetilde{w}^{U_n}(s)s^4\,ds \bullet \int_0^\infty \widetilde{w}^{U_n}(s)\,ds.} \tag{4.4}$$

As a result $w_\sigma^{U_n}(u)$ satisfies the rather natural conditions

$$\int_0^{U_n}(-u^2/2)w_\sigma^{U_n}(u)\,du = 1, \quad \int_0^{U_n} w_\sigma^{U_n}(u)\,du = 0. \tag{4.5}$$

Moreover, $w_\sigma^{U_n}(u) = U_n^{-3}w_\sigma^1(u/U_n)$. Note that w_σ^1 is supported on $[0,1]$ and bounded since $\left(\int_0^1 \widetilde{w}(s)s^2\,ds\right)^2 - \int_0^1 \widetilde{w}(s)s^4\,ds \bullet \int_0^1 \widetilde{w}(s)\,ds > 0$ due to the Cauchy-Schwarz inequality. Analogously,

$$\lambda_n = \int_0^\infty w_\lambda^{U_n}(u)\,\mathrm{Re}\,\psi_n(u)\,du \tag{4.6}$$

holds where $w_\lambda^{U_n}(u) = U_n^{-1} w_\lambda^1(u/U_n)$ with

$$w_\lambda^1(u) := \widetilde{w}(u) \frac{\left(\int_0^1 \widetilde{w}(s)\, s^2\, ds\right) u^2 - \int_0^1 \widetilde{w}(s) s^4\, ds}{\int_0^1 \widetilde{w}(s)\, s^4\, ds \bullet \int_0^1 \widetilde{w}(s)\, ds - \left(\int_0^1 \widetilde{w}(s) s^2\, ds\right)^2}. \tag{4.7}$$

The weight function $w_\lambda^{U_n}$ obviously fulfills

$$\int_0^{U_n} (-1) w_\lambda^{U_n}(u)\, du = 1, \quad \int_0^{U_n} (-u^2/2) w_\lambda^{U_n}(u)\, du = 0.$$

By considering the optimisation problem

$$\gamma_n := \operatorname*{argmin}_\gamma \int_0^\infty \widetilde{w}^{U_n}(u) (\operatorname{Im} \psi_n(u) - \gamma u)^2\, du, \tag{4.8}$$

we arrive at

$$\gamma_n = \int_0^\infty w_\gamma^{U_n}(u) \operatorname{Im} \psi_n(u)\, du, \tag{4.9}$$

where $w_\gamma^{U_n}(u) = U_n^{-2} w_\gamma^1(u/U_n)$ fulfills

$$\int_0^{U_n} u\, w_\gamma^{U_n}(u)\, du = 1.$$

All functions w_σ^1, w_γ^1 and w_λ^1 are bounded and supported on $[0,1]$. Assume now that the Lévy measure ν possesses a density, which we denote, with a slight abuse of notation, by $\nu(x)$. Then we define the estimate for ν as a regularised inverse Fourier transform of the remainder:

$$\nu_n(x) := \mathscr{F}^{-1}\left[\left(\psi_n(\bullet) + \tfrac{\sigma_n^2}{2}(\bullet)^2 - \mathtt{i}\gamma_n(\bullet) + \lambda_n\right) w_\nu(\bullet/U_n)\right](x), \quad x \in \mathbb{R}, \tag{4.10}$$

where w_ν is a weight function supported on $[-1,1]$. Note that $\int_{\mathbb{R}} \nu_n(x)\, dx = \lambda_n$, if $w_\nu(0) = 1$. Due to the estimation error and as a result of the cut-off procedure in (4.10), the estimate ν_n can take negative values and needs correcting. A corrected version ν_n^+ can be constructed via finding a density which is closest to ν_n, i.e., we need to solve the optimisation problem

$$\|\nu_n^+ - \nu_n\|_{L^2(\mathbb{R})}^2 \to \min, \quad \inf_{x \in \mathbb{R}} \nu_n^+ \geq 0$$

subject to

$$\int \nu_n^+(x)\, dx = \int \nu_n(x)\, dx = \lambda_n.$$

It turns out that the above optimisation problem can be solved explicitly:

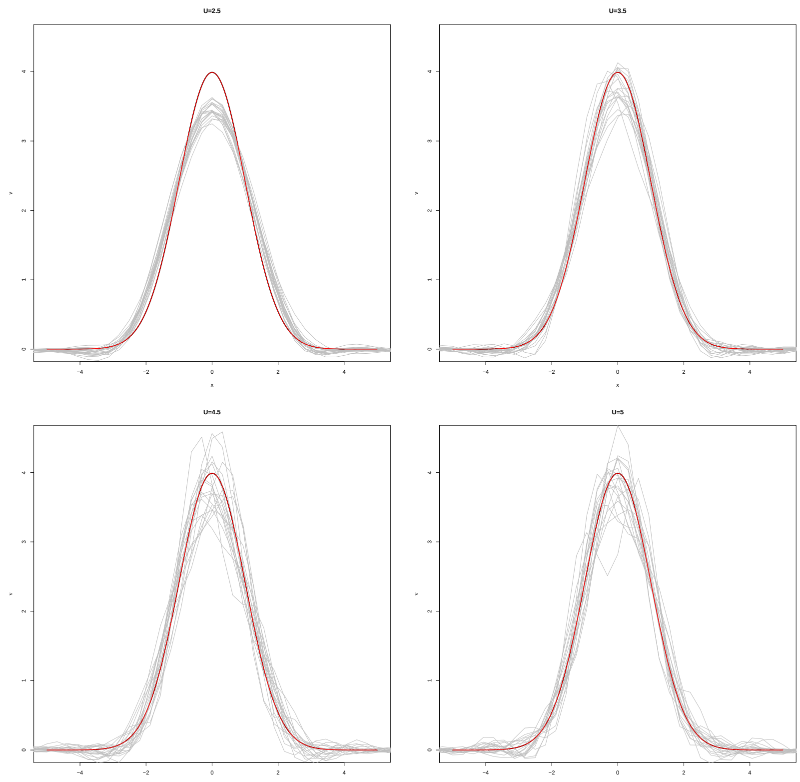

Fig. 2 The estimated Lévy densities ν_{1000} for four different cut-off parameters in the Merton model.

$$\nu_n^+(x;\xi) = \max\{0, \nu_n(x) - \xi\},$$

where ξ is chosen to satisfy the equation $\int \nu_n^+(x;\xi)\,dx = \lambda_n$.

Let us analyse the performance of the estimator ν_n from (4.10) in the Merton model with the same parameters as before (see Figure 1). Figure 2 shows 20 estimated densities ν_{1000} (grey) together with the true Lévy density $\nu(x) = \frac{10}{\sqrt{2\pi}} e^{-x^2/2}$ (red) for four different cut-off parameters $U \in \{2.5, 3.5, 4.5, 5\}$. As one can see, the larger U, the higher the variance of ν_n is. On the other hand, the approximation error or bias in estimating ν decreases with U (compare with Figure 1) and the optimal value of U should balance the bias and the variance (see Section 4.2 for the choice of U based on asymptotic considerations).

4.2 Error decomposition

For the sake of clarity we focus our analysis on the estimate σ_n. First note that by (4.3) and (4.5) the difference $\sigma_n^2 - \sigma^2$ can be decomposed as follows:

$$\begin{aligned}\sigma_n^2 - \sigma^2 &= \int_0^{U_n} w_\sigma^{U_n}(u)\,\mathrm{Re}\,(\psi_n(u) - \psi(u))\,du + \int_0^{U_n} w_\sigma^{U_n}(u)\,\mathrm{Re}\,\psi(u)\,du - \sigma^2 \\ &= \underbrace{\int_0^{U_n} w_\sigma^{U_n}(u)\,\mathrm{Re}\,(\psi_n(u) - \psi(u))\,du}_{\text{Statistical error}} + \underbrace{\int_0^{U_n} w_\sigma^{U_n}(u)\mathscr{F}[\nu](u)\,du}_{\text{Bias}}. \quad (4.11)\end{aligned}$$

While the first term in (4.11) is connected to the statistical error due to the use of ψ_n instead of ψ, the second one reflects the misspecification error (bias) due to the approximation of $\mathrm{Re}\,\psi(u)$ by $-\sigma^2 u^2/2 - \lambda$. The statistical error can be further decomposed into the first order (linear) term and a remainder:

$$\int_0^{U_n} w_\sigma^{U_n}(u)\,\mathrm{Re}\big(\psi_n(u) - \psi(u)\big)\,du = \underbrace{\int_0^{U_n} w_\sigma^{U_n}(u)\,\mathrm{Re}\big((\varphi_n(u) - \varphi(u))/\varphi(u)\big)\,du}_{\text{Linear term}} + R_n.$$

In fact, the remainder R_n contains higher order terms of the Taylor expansion of $\log(1+z)$.

We shall use throughout the notation $A \lesssim B$ if A is bounded by a constant multiple of B, independently of the parameters involved, that is, in the Landau notation $A = O(B)$. Equally $A \gtrsim B$ means $B \lesssim A$ and $A \sim B$ stands for $A \lesssim B$ and $A \gtrsim B$ simultaneously.

4.2.1 Bias

Let us first study the bias term in (4.11). Its order obviously depends on the decay of the Fourier transform $\mathscr{F}[\nu](u)$, which in turn is related to the smoothness of ν (see [43]). Suppose that the s-fold derivative $\nu^{(s)}$ of ν satisfies $\|\nu^{(s)}\|_{L^\infty(\mathbb{R})} \leqslant C$ for some $C > 0$, then by the Plancherel identity

$$\begin{aligned}\left|\int_0^\infty w_\sigma^{U_n}(u)\mathscr{F}[\nu](u)du\right| &= 2\pi\left|\int_{-\infty}^\infty \nu^{(s)}(x)\overline{\mathscr{F}^{-1}[w_\sigma^{U_n}(\bullet)/(\mathtt{i}\bullet)^s](x)}dx\right| \\ &\le U_n^{-(s+3)}\|\nu^{(s)}\|_\infty \|\mathscr{F}(w_\sigma^1(u)/u^s)\|_{L^1}.\end{aligned}$$

So

$$\left|\int_0^\infty w_\sigma^{U_n}(u)\mathscr{F}[\nu](u)du\right| \lesssim U_n^{-(s+3)}, \quad (4.12)$$

provided $\|\mathscr{F}(w_\sigma^1(\bullet)/\bullet^s)\|_{L^1} < \infty$.

4.2.2 Linear term

The linear term

$$L_n := \int_0^{U_n} w_\sigma^{U_n}(u)\,\mathrm{Re}\left(\frac{\varphi_n(u)-\varphi(u)}{\varphi(u)}\right) du$$

can be analysed using the covariance structure of φ_n (see Section 3). We have $\mathbb{E}[L_n] = 0$ and

$$\begin{aligned}
\mathrm{Var}[L_n] &\le \int_0^{U_n}\int_0^{U_n} w_\sigma^{U_n}(u) w_\sigma^{U_n}(v)\,\mathrm{Cov}_{\mathbb{C}}\big(\varphi_n(u)/\varphi(u), \varphi_n(v)/\varphi(v)\big)\,du\,dv \\
&= \frac{1}{n}\int_0^{U_n}\int_0^{U_n} w_\sigma^{U_n}(u) w_\sigma^{U_n}(v)\varphi^{-1}(u)\varphi^{-1}(-v)\,(\varphi(u-v)-\varphi(u)\varphi(-v))\,du\,dv \\
&\le \frac{2}{n}\left(\int_0^{U_n} \left|w_\sigma^{U_n}(u)/\varphi(u)\right| du\right)^2 \\
&= \frac{2}{nU_n^4}\left(\int_0^1 \left|w_\sigma^1(u)/\varphi(uU_n)\right| du\right)^2 =: \varepsilon_{1,n}^2 \qquad (4.13)
\end{aligned}$$

Hence, Markov's inequality yields

$$\mathbb{P}(|L_n| > A\varepsilon_{1,n}) \le A^{-2}. \qquad (4.14)$$

4.2.3 Remainder term

Introduce the *good event*

$$\mathscr{G}_n := \left\{\|(\varphi_n-\varphi)/\varphi\|_{U_n} \le 1/2\right\} \text{ with } \|f\|_{U_n} := \sup_{u\in[-U_n,U_n]} |f(u)|.$$

Then the simple inequality $|\log(1+z)-z| \le 2|z|^2$ for $|z| < 1/2$ yields on $\mathscr{G}_n$

$$\psi_n(u)-\psi(u) = \frac{1}{\Delta}\left\{\frac{\varphi_n(u)-\varphi(u)}{\varphi(u)} + O\left(\left|\frac{\varphi_n(u)-\varphi(u)}{\varphi(u)}\right|^2\right)\right\}. \qquad (4.15)$$

Let us estimate the probability of the complement $\mathscr{G}_n^c$. This can be conveniently done using Proposition 3.3:

$$\mathbb{P}(\mathscr{G}_n^c) \le \mathbb{P}\left(\sqrt{n/\log(nU_n^2)}\,\|\varphi_n-\varphi\|_{U_n} > \frac{1}{2}\sqrt{n/\log(nU_n^2)}\inf_{u\in[-U_n,U_n]}|\varphi(u)|\right)$$
$$= \mathbb{P}\left(\sqrt{n/\log(nU_n^2)}\,\|\varphi_n-\varphi\|_{U_n} > \varkappa_n\right) = O\left((\sqrt{n}U_n)^{(64-\varkappa_n)/128}\right),$$

provided U_n is chosen in such a way that

$$\varkappa_n := \frac{1}{2}\sqrt{n/\log(nU_n^2)}\inf_{u\in[-U_n,U_n]}|\varphi(u)| > 64.$$

The latter condition means that U_n should not increase too fast with n. In a similar way we can bound the quadratic term in (4.15). Denote $\varepsilon_{2,n} = 1/\varkappa_n$ then

$$\mathbb{P}\left(\|(\varphi_n-\varphi)/\varphi\|_{U_n}^2 > 4A\varepsilon_{2,n}^2\right) \le \mathbb{P}\left(n\,\|\varphi_n-\varphi\|_{U_n}^2 > 4A\log(nU_n^2)\right)$$
$$= O\left((\sqrt{n}U_n)^{(64-4A)/128}\right) \tag{4.16}$$

for $A > 16$. Hence, we have on $\mathscr{G}_n$

$$|R_n| \lesssim \Delta^{-1}\,\|(\varphi_n-\varphi)/\varphi\|_{U_n}^2\int_0^{U_n}|w_\sigma^{U_n}(u)|\,du \lesssim \Delta^{-1}U_n^{-2}.$$

4.3 Minimax upper bounds

In this section we derive the uniform (over a class of Lévy models) convergence of the estimators σ_n^2, γ_n, λ_n and ν_n defined in Section 4.1. First let us define the corresponding class of Lévy processes.

Definition 4.1. *For $s\in\mathbb{N}$ and $R,\sigma_{max}>0$ let $\mathscr{G}_s(R,\sigma_{max})$ denote the set of all Lévy triplets $\mathscr{T}=(\gamma,\sigma^2,\nu)$, such that ν is s-times (weakly) differentiable and*

$$\sigma\in[0,\sigma_{max}],\quad |\gamma|,\lambda\in[0,R],\quad \|\nu^{(s)}\|_{L^\infty(\mathbb{R})}\leqslant R.$$

Definition 4.2. *Let $\{\mathbb{P}_\vartheta,\vartheta\in\Theta\}$ be a family of probability measures on $(\Omega,\mathfrak{F})$. Assume that $\xi_n=\xi_n(\vartheta)$ is a sequence of random variables, possibly depending on ϑ, all defined on $(\Omega,\mathfrak{F})$. We write $\xi_n=O_{\mathrm{P},\Theta}(r_n)$ for a sequence of positive numbers r_n, if*

$$\lim_{A\to\infty}\limsup_{n\to\infty}\sup_{\vartheta\in\Theta}\mathbb{P}_\vartheta\left(|\xi_n(\vartheta)|\geqslant Ar_n\right)=0.$$

The main statement about the convergence of σ_n^2 in the case $\sigma^2>0$ is as follows.

Theorem 4.3. *Suppose that the weight function w_σ^1 satisfies*

$$\|\mathscr{F}(w^1_\sigma(u)/u^s)\|_{L^1} < \infty.$$

Choosing for some $\bar\sigma > \sigma_{max}$ *the cut-off* $U_n := \bar\sigma^{-1}(\log(n)/\Delta)^{1/2}$, *we obtain for the risk of* σ_n^2 *the uniform convergence rate*

$$\sigma_n^2 - \sigma^2 = O_{\mathrm{P},\mathscr{G}_s}\big((\Delta\bar\sigma^2/\log(n))^{(s+3)/2}\big). \tag{4.17}$$

Proof. We have for n large enough

$$\begin{aligned}\varepsilon_{1,n} &= \frac{\sqrt{2}}{\sqrt{n}U_n^2}\int_0^1 \left|w^1_\sigma(u)/\varphi(uU_n)\right| du \\ &\lesssim \frac{1}{\sqrt{n}U_n^2|\varphi(U_n)|}\int_0^1 \left|w^1_\sigma(u)\right| du \lesssim \frac{1}{\sqrt{n}\log(n)} n^{\sigma^2/(2\bar\sigma^2)}\end{aligned}$$

and

$$\begin{aligned}\varepsilon_{2,n} &= 2\sqrt{(1/n)\log(nU_n^2)}\left[\inf_{u\in[-U_n,U_n]}|\varphi(u)|\right]^{-1} \\ &\lesssim \sqrt{(1/n)\log(nU_n^2)}\,|\varphi(U_n)|^{-1} \lesssim \sqrt{\frac{\log n}{n}}\, n^{\sigma^2/(2\bar\sigma^2)}.\end{aligned}$$

Combining this with (4.12), (4.16) and (4.14), we get

$$\lim_{A\to\infty}\limsup_{n\to\infty}\sup_{(\sigma^2,\gamma,\mu)\in\mathscr{G}_s(R,\sigma_{max})}\mathbb{P}_{(\sigma^2,\gamma,\mu)}\Big(|\sigma_n^2-\sigma^2| > A(\Delta\bar\sigma^2/\log(n))^{(s+3)/2}\Big) = 0.$$

In a similar way one can derive the following minimax convergence rates for γ, λ and ν.

Theorem 4.4. *Suppose that the weight functions* w^1_γ *and* w^1_λ *satisfy*

$$\|\mathscr{F}(w^1_\gamma(u)/u^s)\|_{L^1} < \infty, \quad \|\mathscr{F}(w^1_\lambda(u)/u^s)\|_{L^1} < \infty.$$

For any $\bar\sigma > \sigma_{max}$ *we choose*

$$U_n := \bar\sigma^{-1}\big(\log(n)/\Delta\big)^{1/2}. \tag{4.18}$$

Then

$$\gamma_n - \gamma = O_{\mathrm{P},\mathscr{G}_s}\big((\Delta\bar\sigma^2/\log(n))^{(s+2)/2}\big)$$

and

$$\lambda_n - \lambda = O_{\mathrm{P},\mathscr{G}_s}\big((\Delta\bar\sigma^2/\log(n))^{(s+1)/2}\big).$$

For $\sigma = 0$ the convergence rates of γ_n and λ_n become polynomial. In this case one can prove that

$$\gamma_n - \gamma = O_{\mathrm{P},\mathscr{G}_s}\big((\Delta/n)^{(2s+4)/(2s+5)}\big)$$

and
$$\lambda_n - \lambda = O_{\mathrm{P},\mathscr{G}_s}\big((\Delta/n)^{(2s+2)/(2s+5)}\big).$$

An imminent statistical question is, of course, a fully data-driven choice of the spectral cut-off value U_n, independently of knowing the values of s or σ. In practice, the very simple quasi-optimality criterion works well, while it only allows for statistical optimality results in a Bayesian (average) sense and may fail in a minimax sense, see Bauer and Reiß [7]. Since the level of the stochastic error is not known a priori, it is highly non-standard to find a provably minimax optimal selection rule for U_n which has been achieved recently by Comte and Genon-Catalot [20] for a high-frequency setting and by Kappus [41] in a low-frequency setting, both based on a penalized empirical risk criterion, see also chapter "Adaptive Estimation for Lévy processes" by F. Comte and V. Genon-Catalot in this book. In practice, however, a much smaller penalisation constant must be chosen than necessary for the proofs.

5 Rate optimality for the triplet estimation

We want to prove that our spectral estimation method achieves optimal convergence rates. Conceptually, it is clear that we need a convergence rate over a certain nonparametric class because individual rates for each parameter are not well-defined since the parameter is then known. A general paradigm is the minimax risk approach where the uniform risk over a class is minimised.

Definition 5.1. *Consider a sequence of statistical models (i.e., measurable spaces with a family of probability measures) $(\mathscr{X}_n, \mathfrak{F}_n, (\mathbb{P}_{f,n})_{f\in\mathscr{G}})$, $n\in\mathbb{N}$, with a family $\mathscr{G}$ of unknown parameters, equipped with a semi-metric d. Let $(v_n)_{n\geqslant 1}$ be a sequence converging to zero and assume that there are estimators $\hat{f}_n$ in model n (i.e., $\hat{f}_n : \mathscr{X}_n \to \mathscr{G}$ measurable) such that $d(\hat{f}_n, f) = O_{P,\mathscr{G}}(v_n)$ holds according to Definition 4.2. Then (v_n) is called* optimal rate of convergence *in a minimax sense over $\mathscr{G}$ if also*

$$\exists \varepsilon > 0 : \liminf_{n\to\infty} \inf_{\hat{\vartheta}_n} \sup_{f\in\mathscr{G}} \mathbb{P}_{f,n}(d(\hat{\vartheta}_n, f) > \varepsilon v_n) > 0$$

holds, the infimum being taken over all estimators $\hat{\vartheta}_n$ over the observations in model n.

The limiting property in the display of the definition means that $d(\hat{\vartheta}_n, f)$ is not $o_P(v_n)$ uniformly in $f\in\mathscr{G}$ for any estimator sequence $(\hat{\vartheta}_n)$. The optimal rates of convergence are formulated in an $\mathscr{O}_P$-setting, which is the right type of convergence for the construction of confidence regions and facilitates the proof of the upper bound. A lower bound in O_P-sense, of course, yields a fortiori also a lower bound for a pth moment risk $\mathbb{E}_{f,n}[d(\hat{f}_n, f)^p]^{1/p}$.

Here, we shall first explain in detail how lower bounds for the minimax risk are proved, inspired by the exposition in Tsybakov [60]. This is in fact not too difficult, although at first sight it seems quite untractable to deal with the minimal risk over the set of all estimators, that is measurable functions of the data. Afterwards we apply this technology to derive the lower bound for estimating the scalar parameters $\sigma^2, \gamma, \lambda$ of the Lévy triplet.

5.1 A general recipe for lower bound proofs

Let us reduce the statement of the lower bound for estimators to a test problem between a finite set $\{f_1, \dots, f_M\} \subseteq \mathscr{G}$ of parameters. Suppose for some $\varepsilon > 0$

$$d(f_k, f_l) > 2\varepsilon v_n \text{ for all } k, l = 1, \dots, M \text{ with } k \neq l.$$

Then any estimator $\hat{\vartheta}_n$ in model n satisfies

$$\sup_{f \in \mathscr{G}} \mathbb{P}_{f,n}(d(\hat{\vartheta}_n, f) > \varepsilon v_n) \geqslant \max_{j=1,\dots,M} \mathbb{P}_{f_j,n}(d(\hat{\vartheta}_n, f_j) > \varepsilon v_n) \geqslant \max_{j=1,\dots,M} \mathbb{P}_{f_j,n}(\psi_n^* \neq j),$$

where $\psi_n^* := \operatorname{argmin}_{j=1,\dots,M} d(\hat{\vartheta}_n, f_j)$ denotes the minimum-distance test based upon $\hat{\vartheta}_n$. If we can show

$$\liminf_{n \to \infty} \inf_{\psi_n} \max_{j=1,\dots,M} \mathbb{P}_{f_j,n}(\psi_n \neq j) > 0$$

for all tests ψ_n in model n, then this implies in particular the lower bound for estimation.

Definition 5.2. *For measures μ and ν on $(\mathscr{X}, \mathfrak{F})$ we denote their* total variation distance *by*

$$\|\mu - \nu\|_{TV} := \sup_{A \in \mathfrak{F}} |\mu(A) - \nu(A)|.$$

Proposition 5.3. *Let $\mathbb{P}_1, \dots, \mathbb{P}_M$ be probability measures on $(\mathscr{X}, \mathfrak{F})$ with densities $p_1, \dots, p_M$ with respect to some measure μ (e.g. take $\mu = \sum_{i=1}^M \mathbb{P}_i$). Then any test (measurable map) $\psi : \mathscr{X} \to \{1, \dots, M\}$ between the M hypotheses satisfies*

$$\max_{j=1,\dots,M} \mathbb{P}_j(\psi \neq j) \geqslant \frac{1}{M} \sum_{j=1}^{M} \mathbb{P}_j(\psi \neq j) \geqslant 1 - \frac{1}{M} \int_{\mathscr{X}} \max_{j=1,\dots,M} p_j(x) \mu(dx)$$

For $M = 2$ and hypotheses H_0, H_1 we obtain in terms of the total-variation distance

$$\max_{j=0,1} \mathbb{P}_j(\psi \neq j) \geqslant \frac{1}{2}\Big(\mathbb{P}_0(\psi = 1) + \mathbb{P}_1(\psi = 0)\Big) \geqslant \frac{1}{2}\Big(1 - \|\mathbb{P}_0 - \mathbb{P}_1\|_{TV}\Big).$$

Proof. The first inequality is trivial since the average is never larger than the maximum. For the second inequality note

$$\begin{aligned}\frac{1}{M}\sum_{j=1}^{M}\mathbb{P}_j(\psi\neq j) &= 1-\frac{1}{M}\sum_{j=1}^{M}\int \mathbf{1}(\psi(x)=j)p_j(x)\,\mu(dx)\\ &= 1-\frac{1}{M}\int\Big(\sum_{j=1}^{M}\mathbf{1}(\psi(x)=j)p_j(x)\Big)\,\mu(dx).\end{aligned}$$

Remarking that the last integrand is at most $\max_j p_j(x)$, we conclude

$$\frac{1}{M}\sum_{j=1}^{M}\mathbb{P}_j(\psi\neq j)\geqslant 1-\frac{1}{M}\int \max_{j=1,\dots,M} p_j(x)\,\mu(dx).$$

For $M=2$ we use $\int p_1 d\mu = 1$ and find

$$\begin{aligned}&\mathbb{P}_0(\{x\in\mathscr{X}: p_0(x)>p_1(x)\})-\mathbb{P}_1(\{x\in\mathscr{X}: p_0(x)>p_1(x)\})\\ &=\int_{\{p_0>p_1\}}(p_0(x)-p_1(x))\,\mu(dx)+\int_{\mathscr{X}}p_1(x)\,\mu(dx)-1\\ &=\int_{\{p_0>p_1\}}p_0(x)\,\mu(dx)+\int_{\{p_0\leqslant p_1\}}p_1(x)\,\mu(dx)-1.\end{aligned}$$

This shows $\|\mathbb{P}_0-\mathbb{P}_1\|_{TV}+1\geqslant\int_{\mathscr{X}}\max(p_0(x),p_1(x))\,\mu(dx)$. Insertion in the general case yields the result. □

The case $M=2$ yields the well-known characterisation of the total variation distance in terms of the minimax error in testing:

$$\inf_{\psi:\text{ test}}\Big(\mathbb{P}_0(\psi=1)+\mathbb{P}_1(\psi=0)\Big)=1-\|\mathbb{P}_0-\mathbb{P}_1\|_{TV}.$$

A test that attains the bound on the right-hand side is the Neyman-Pearson test $\psi(x)=\mathbf{1}(p_1(x)>p_0(x))$.

For statistical purposes the total variation distance is often not very useful because it cannot easily be bounded for product measures, e.g. arriving from i.i.d. observations. For this, other methods to measure the closeness of distributions are much more convenient like Kullback-Leibler divergence (also known as relative entropy) or the Hellinger distance, see Tsybakov [60]. For estimates in the Fourier domain, the χ^2-divergence has proved to be a powerful tool.

Definition 5.4. *For probability measures* $\mathbb{P}$ *and* $\mathbb{Q}$ *on* $(\mathscr{X},\mathfrak{F})$ *we denote their* χ^2-divergence *by*

$$\chi^2(\mathbb{P}|\mathbb{Q}):=\begin{cases}\int_{\mathscr{X}}\big(\frac{d\mathbb{P}}{d\mathbb{Q}}-1\big)^2 d\mathbb{Q}, & \textit{if } \mathbb{P}\ll\mathbb{Q},\\ +\infty, & \textit{otherwise.}\end{cases}$$

Lemma 5.5. *The χ^2-divergence satisfies:*

(a) $\|\mathbb{P}-\mathbb{Q}\|_{TV} \leqslant \frac{1}{2}\sqrt{\chi^2(\mathbb{P}|\mathbb{Q})}$;
(b) $\chi^2(\mathbb{P}^{\otimes n}|\mathbb{Q}^{\otimes n}) = (1+\chi^2(\mathbb{P}|\mathbb{Q}))^n - 1 \leqslant \exp\big(n\chi^2(\mathbb{P}|\mathbb{Q})\big) - 1$.

Proof. For part (a) we may assume $\mathbb{P} \ll \mathbb{Q}$ and thus obtain by the relationship between total-variation and L^1-distance as well as by the Cauchy-Schwarz or Jensen inequality:

$$\|\mathbb{P}-\mathbb{Q}\|_{TV} = \frac{1}{2}\int_{\mathscr{X}} \Big|\frac{d\mathbb{P}}{d\mathbb{Q}} - 1\Big| d\mathbb{Q} \leqslant \frac{1}{2}\Big(\int_{\mathscr{X}} \Big|\frac{d\mathbb{P}}{d\mathbb{Q}} - 1\Big|^2 d\mathbb{Q}\Big)^{1/2}.$$

Part (b) follows from the formula

$$\begin{aligned}\chi^2(\mathbb{P}^{\otimes n}|\mathbb{Q}^{\otimes n}) &= \int_{\mathscr{X}^n} \Big(\frac{d\mathbb{P}^{\otimes n}}{d\mathbb{Q}^{\otimes n}}\Big)^2 d\mathbb{Q}^{\otimes n} - \int_{\mathscr{X}^n} 2d\mathbb{P}^{\otimes n} + \int_{\mathscr{X}^n} 1 d\mathbb{Q}^{\otimes n}\\ &= \Big(\int_{\mathscr{X}} \Big(\frac{d\mathbb{P}}{d\mathbb{Q}}\Big)^2 d\mathbb{Q}\Big)^n - 1.\end{aligned}$$

and the inequality $(1+x) \leqslant e^x$. □

The proofs in the sequel will rely on testing between $M=2$ hypotheses based on a sample of i.i.d. observations. Consequently, we gather the findings of this section for this case. Lower bound results in a global functional norm like L^2 require a high combinatorial complexity of the test problem with M_n hypotheses and $M_n \to \infty$ quickly.

Theorem 5.6. *Suppose that for some $\varepsilon > 0$ and $n \in \mathbb{N}$ there are parameters $f_{0,n}, f_{1,n} \in \mathscr{G}$ such that*

$$d(f_{0,n}, f_{1,n}) > 2\varepsilon v_n.$$

If the observations in model n follow the product law $\mathbb{P}_{f,n} = \mathbb{P}_f^{\otimes n}$ under parameter $f \in \mathscr{G}$ and

$$\chi^2(\mathbb{P}_{f_{1,n}}|\mathbb{P}_{f_{0,n}}) \leqslant n^{-1}\log(1+(2-4\delta)^2)$$

holds for some $\delta \in (0,1/2)$, then the following lower bound holds for all estimators $\hat{\vartheta}_n$ based on observations from model n:

$$\inf_{\hat{\vartheta}_n} \sup_{f\in\mathscr{G}} \mathbb{P}_{f,n}\big(d(\hat{\vartheta}_n, f) \geqslant \varepsilon v_n\big) \geqslant \delta.$$

If the above holds for fixed $\varepsilon, \delta > 0$ and all $n \in \mathbb{N}$, then the optimal rate of convergence in a minimax sense over $\mathscr{G}$ is not faster than v_n.

Proof. We infer from the preceding lemma

$$\|\mathbb{P}_{f_{1,n}}^{\otimes n} - \mathbb{P}_{f_{2,n}}^{\otimes n}\|_{TV} \leqslant \frac{1}{2}\sqrt{e^{\log(1+(2-4\delta)^2)} - 1} = 1 - 2\delta.$$

Proposition 5.3 yields the minimax lower bound for testing

$$\max_{j=0,1} \mathbb{P}_{f_j}(\psi \neq j) \geqslant \frac{1}{2}(1-(1-2\delta)) = \delta,$$

which by the reduction of estimation to testing problems yields the assertion. The asymptotic rate result then holds by definition. □

5.2 Lower bound for estimating the triplet

We now come back to the estimation problem from Section 4.1 and derive rate optimality of the estimator.

Theorem 5.7. *The rates for estimating σ^2, γ and λ, respectively, among triplets in the class $\mathscr{G}_s(R, \sigma_{max})$ with $s, R, \sigma_{max} > 0$ from Theorem 4.4 are minimax optimal.*

Proof. We need a kernel function K whose Fourier transform satisfies $\mathscr{F}K(u) = 1$ for $u \in [-1,1]$. For convenience, we require K to be a Schwartz functions, that is $K, \mathscr{F}K \in C^\infty(\mathbb{R})$ show polynomial decay of any order for the function and its derivatives. Such functions K are called *flat-top kernel* and an example is given by

$$\mathscr{F}K(u) = \begin{cases} 1, & |u| \leq 1, \\ \exp\left(-\frac{e^{-1/(|u|-1)}}{2-|u|}\right), & 1 < |u| < 2, \\ 0, & |u| \geq 2. \end{cases}$$

Introduce $K_h(x) = h^{-1}K(h^{-1}x)$ for some (bandwidth) $h > 0$. Suppose $(\sigma^2, \gamma, \nu) \in \mathscr{G}_s(R/2, \sigma_{max})$ with $\sigma = \sigma_{max}/2$, $\gamma = 0$ and a Lévy density $\nu \in C^s(\mathbb{R})$ such that $\int |x|^m \nu(dx) < \infty$ and on the other hand

$$\nu(x) \gtrsim (1+|x|)^{-2m}$$

for some $m \in \mathbb{N}$, $m \geqslant 2$, e.g. take $\nu(x) = c(1+|x|)^{-2m}$ with a sufficiently small $c > 0$.

We now perturb (σ, ν) such that for low frequencies the characteristic functions still coincide. For $\delta > 0$ we set

$$\sigma_0^2 = \sigma^2, \, \nu_0 = \nu - \delta K_h''; \quad \sigma_1^2 = \sigma^2 + 2\delta, \, \nu_1 = \nu.$$

Note that ν_0 is positive when h is small enough, provided $\delta = o(h^3)$ since then as $h \to 0$

$$\delta |K_h''(x)| = \delta h^{-3} |K''(x/h)| \lesssim \delta h^{-3}(1+|x|/h)^{-2m} = o((1+|x|)^{-2m}) = o(\nu(x))$$

(uniformly over $x \in \mathbb{R}$) follows by the polynomial decay of K'' of any order. Moreover, we have

$$\|(\nu_0 - \nu)^{(s)}\|_\infty = \delta \|K_h^{(s+2)}\|_\infty = \delta h^{-s-3} \|K^{(s+2)}\|_\infty \lesssim \delta h^{-s-3}$$

such that for $\delta = \delta' h^{s+3}$ with some small, but fixed $\delta' > 0$, both triplets are asymptotically as $h \to 0$ still in the parameter class $\mathscr{G}_s(R, \sigma_{max})$.

For the corresponding characteristic exponents we obtain (note $\mathscr{F}K_h''(u) = -u^2 \mathscr{F}K_h$, $\int K_h'' = 0$):

$$\psi_1(u) - \psi_0(u) = \delta u^2 (-1 + \mathscr{F}K(hu)),$$

which is zero for $u \in [-h^{-1}, h^{-1}]$.

The marginal density p_0 of the Lévy process with triplet $(\sigma_0^2, 0, \nu_0)$ at time $\Delta > 0$ then satisfies using the Poisson convolution exponential and $\lambda = \|\nu_0\|_{L^1}$:

$$\begin{aligned} p_0(x) &= \Big(N(0, \sigma_0^2 \Delta) * \sum_{k=0}^\infty \frac{e^{-\lambda\Delta} (\lambda \Delta)^k}{k!} \nu_0^{*k} \Big)(x) \\ &\geqslant \lambda \Delta e^{-\lambda \Delta} (N(0, \sigma_0^2 \Delta) * \nu_0)(x) \\ &\gtrsim \Delta (1 + |x|)^{-2m}, \end{aligned}$$

in view of the positivity of the summands, $\nu_0 \gtrsim \nu$ and the exponential decay of the Gaussian density (uniformly for $\Delta \lesssim 1$ and keeping λ, σ_0, ν_0 fixed). The χ^2-divergence between the two densities (i.e. their laws) then satisfies

$$\chi^2(p_1 | p_0) = \int \frac{(p_1(x) - p_0(x))^2}{p_0(x)} dx \lesssim \Delta^{-1} \int (1 + x^{2m})(p_1(x) - p_0(x))^2 dx.$$

By the Plancherel identity and $\mathscr{F}[(ix)^m p(x)] = (\mathscr{F}p)^{(m)}$ we arrive at

$$\chi^2(p_1 | p_0) \lesssim \Delta^{-1} \big(\|\varphi_1 - \varphi_0\|_{L^2}^2 + \|(\varphi_1 - \varphi_0)^{(m)}\|_{L^2}^2 \big).$$

With the inequality $|1 - e^{-z}| \leqslant 2|z|$ for $z = x + iy \in \mathbb{C}$ with $x \geqslant 0$ (use $1 - e^{-x} \leqslant x$ by concavity and $|e^{-x} - e^{-(x+\mathrm{i}y)}| \leqslant \sqrt{2}|y|$) we can estimate

$$\begin{aligned} \|\varphi_1 - \varphi_0\|_{L^2}^2 &\leqslant \int 2 \max(|\varphi_0(u)|, |\varphi_1(u)|)^2 \Delta^2 |\psi_1(u) - \psi_0(u)|^2 du \\ &\lesssim \int_{|u| > h^{-1}} e^{-\Delta \sigma^2 u^2} \Delta^2 \delta^2 u^4 |-1 + \mathscr{F}K(hu)|^2 du \\ &\lesssim \Delta^2 \delta^2 \int_{|u| > h^{-1}} e^{-\Delta \sigma^2 u^2} u^4 du \\ &\lesssim \Delta^{3/2} \sigma^{-1} \delta^2 h^{-4} e^{-\Delta \sigma^2 h^{-2}}. \end{aligned}$$

The main order is $e^{-\Delta\sigma^2 h^{-2}}$ and this we also expect for $\|(\varphi_1-\varphi_0)^{(m)}\|_{L^2}^2$. For $j=0,1$ we have by the Leibniz formula

$$\varphi_j^{(m)} = (\Delta\psi_j'\varphi_j)^{(m-1)} = \Delta\sum_{k=0}^{m-1}\binom{m-1}{k}\varphi_j^{(k)}\psi_j^{(m-k)}.$$

Now, using the (rough) bounds $|\varphi_j^{(k)}(u)| \lesssim (1+|u|)^k|\varphi_j(u)|$, $|\psi_j^{(m-k)}(u)| \lesssim (1+|u|)$ (because of $\int|x|^m\nu_j(dx)<\infty$) and the preceding formula, iteratively for k instead of m, we arrive at

$$|(\varphi_1-\varphi_0)^{(m)}(u)| \lesssim \Delta\sum_{k=0}^{m-1}|(\psi_1-\psi_0)^{(m-k)}(u)|(1+|u|^k)e^{-\Delta\sigma^2u^2/2}.$$

Together with $\|\mathscr{F}K(h\bullet)^{(k)}\|_\infty \lesssim h^k \lesssim 1$ this yields for $h<1$ the L^2-bound

$$\|(\varphi_1-\varphi_0)^{(m)}\|_{L^2}^2 \lesssim \Delta^2\delta^2\int_{|u|>h^{-1}}u^{2m+4}e^{-\Delta\sigma^2u^2}\,du \lesssim \Delta^{3/2}\sigma^{-1}\delta^2h^{-2m-4}e^{-\Delta\sigma^2h^{-2}}.$$

Inserting the choice of $\delta=\delta' h^{s+3}$, the χ^2-divergence is bounded by

$$\chi^2(p_1|p_0) \lesssim \Delta^{1/2}\sigma^{-1}h^{2(s-m+1)}e^{-\Delta\sigma^2h^{-2}}.$$

If we choose $h=\sqrt{(1-\varepsilon)\Delta\sigma^2/\log n}$ for any (small) $\varepsilon\in(0,1)$, then $\chi^2(p_1|p_0)\lesssim n^{-(1+\varepsilon)}$ follows. Applying Theorem 5.6, we have thus proved that

$$|\sigma_1^2-\sigma_0^2| = 2\delta' h^{s+3} \sim \Big(\frac{\Delta\sigma^2}{\log n}\Big)^{(s+3)/2}$$

gives a lower bound for the minimax rate for estimating σ^2.

The minimax rate for γ and λ follow in the same way. For γ we use the alternatives

$$\gamma_0=\gamma,\ \nu_0=\nu-\delta K_h'; \qquad \gamma_1=\gamma+\delta, \quad \nu_1=\nu$$

with $\sigma_1=\sigma_0>0$ and note $\mathscr{F}K_h'(u)=-iu\mathscr{F}K(hu)$, $\|(\nu_0-\nu)^{(s)}\|_\infty \lesssim \delta h^{-s-2}$. For λ we only perturb the Lévy density by considering

$$\nu_0=\nu+\delta K_h; \qquad \nu_1=\nu,$$

for which $\lambda_0-\lambda_1=\delta$ and $\|(\nu_0-\nu)^{(s)}\|_\infty \lesssim \delta h^{-s-1}$ hold. □

The L^2-lower bound for ν requires Fano's lemma or Assouad's cube technique with many alternatives, but does not use any fine interplay between ν_j and σ_j. We refer to Gugushvili [34] for the proof.

In the case $\sigma=0$, the same argument via bounding the χ^2-divergence goes through (with slightly different estimates), but it is even easier to apply the result of Thm. 2.6 in Liese [46] that the Kullback-Leibler divergence in the case $\gamma_0 = \gamma_1$ satisfies

$$KL(p_1|p_0) \leqslant KL(\nu_1|\nu_0).$$

So, we can transfer all lower bound techniques for density estimation in the natural Kullback-Leibler topology, see e.g. Tsybakov [60], to the estimation of the jump density in the compound Poisson case.

The general bounds in Liese [46] do not take the smoothing effect of a Gaussian component into account and are thus highly suboptimal in the case $\sigma > 0$ (or even for most infinite jump activity models). The idea to mimick the diffusive part at low frequencies exactly through a suitable difference of jump measures has been first applied by Jacod and Reiß [37] for infinite variation jump models. A lower bound for the estimation of functionals of the Lévy density is given in Neumann and Reiß [48].

6 Extension to the infinite intensity case

The natural statistical problem is certainly the estimation of the Lévy triplet in the general case, allowing for infinite jump intensity. The structural nonparametric problem is that, even if a jump density exists, it will have a strong singularity at zero and smoothing methods are not appropriate. One possibility is to estimate the jump density only away from zero as in Figueroa-Lopez and Houdré [32].

One can also argue that finite activity Lévy processes approximate arbitrarily well general Lévy processes in terms of their finite-dimensional distributions. Due to the statistical uncertainty we cannot recover the true object from a finite sample anyway and thus we may still use an estimator designed for the finite intensity case. In Section 6.1 we shall study this estimator in the general case and in particular see that infinite activity of small jumps is absorbed by an increase in the estimated diffusion coefficient besides a high, but finite estimated activity.

Another way to deal with the general case is that a weaker norm is used in order to assess the performance of the Lévy measure estimator, where the measure is used to integrate functions of some minimal regularity, cf. Neumann and Reiß [48]. In Section 6.2 we shall investigate another estimator, based on the second derivative of the empirical characteristic function, from this point of view. The material of this section is mainly new.

6.1 Performance of the finite intensity estimator

The class of Lévy processes with finite intensity jumps lies dense in the class of all Lévy processes with respect to the weak convergence of finite-dimensional distributions. The natural statistical question, which is treated here, is how the above spectral estimator, designed for the finite intensity case, works if the underlying Lévy process has infinite jump intensity. In statistical language we study the behaviour of the spectral method in Section 4 under *model misspecification*.

At a qualitative level the definition of the estimators σ_n^2, λ_n in (4.2) as solutions to an optimisation problem explains the behaviour. In the infinite intensity case $\mathrm{Re}(\psi(u)) \to -\infty$ holds for $u \to \infty$ even without diffusion coefficient. On the frequency interval $[0, U_n]$ the estimators σ_n^2, λ_n minimise a weighted L^2-distance of $\mathrm{Re}(\psi(u))$ to $-\sigma_n^2 u^2/2 - \lambda_n$. The jump part in $|\mathrm{Re}(\psi(u))|$ grows more slowly than u^2 in $u \to \infty$ and thus the infinite intensity jump part results in an increase of both, σ_n^2 and λ_n. In the entire argument we only need to focus on the bias part in (4.11) as it is easy to see that for $\sigma > 0$ the analysis of the statistical error in the infinite intensity case does not change: it is governed by the Gaussian decay of the characteristic function. For a precise quantitative statement we restrict, as often in the literature, to a stable-like behaviour of the jump component.

Proposition 6.1. *Suppose the triplet of the Lévy process X satisfies $\sigma > 0$ and $\int (1-\cos(ux))\nu(dx) = c_\alpha u^\alpha + O(u^\beta)$ for $0 \leqslant \beta < \alpha < 2$ and $c_\alpha > 0$ with the asymptotics $u \to \infty$. Then for any $\bar\sigma > \sigma$:*

$$\sigma_n^2 = \sigma^2 + O_P\Big(U_n^{\alpha-2} + n^{-1/2} U_n^{-2} e^{\bar\sigma^2 \Delta U_n^2/2}\Big), \quad \lambda_n \gtrsim U_n^\alpha + O_P\Big(n^{-1/2} e^{\bar\sigma^2 \Delta U_n^2/2}\Big).$$

In particular, for the choice of U_n in Theorem 7.3 the estimator σ_n^2 is still consistent with rate $(\log n)^{(\alpha-2)/2}$.

Proof. The infinite-intensity analogue of (4.11), not using $\int d\nu < \infty$ and $\int w_\sigma^{U_n} = 0$, gives the bias term in σ_n^2

$$\mathbb{E}[\sigma_n^2] - \sigma^2 = \int_0^{U_n} w_\sigma^{U_n}(u) \,\mathrm{Re}\,\psi(u)\,du - \sigma^2 = \int_0^{U_n} w_\sigma^{U_n}(u) \int (\cos(ux)-1)\nu(dx)\,du.$$

By the assumption on ν we infer that the bias term can be estimated as

$$\begin{aligned}
\Big|\mathbb{E}[\sigma_n^2] - \sigma^2\Big| &= \Big| U_n^{-2} \int_0^1 w_\sigma^1(v) \int (1-\cos(U_n v x))\nu(dx)\,dv \Big| \\
&\lesssim U_n^{-2} \int_0^1 |w_\sigma^1(v)| U_n^\alpha v^\alpha \,dv \\
&\lesssim U_n^{\alpha-2}.
\end{aligned}$$

For the infinite intensity parameter λ the difference of λ_n and the statistical error is

$$\int_0^{U_n} w_\lambda^{U_n}(u)\int(1-\cos(ux))\nu(dx)\,du = \int_0^1 w_\lambda^1(v)\int(1-\cos(U_n vx))\nu(dx)\,dv$$
$$= c_\alpha U_n^\alpha \int_0^1 w_\lambda^1(v)v^\alpha\,dv + O(U_n^\beta).$$

From (4.7) we know that with some constant $C>0$

$$\int_0^1 w_\lambda^1(v)v^\alpha dv = C\Big(\int_0^1 \tilde{w}v^2\int_0^1 \tilde{w}v^{2+\alpha} - \int_0^1 \tilde{w}v^4\int_0^1 \tilde{w}v^\alpha\Big).$$

By the Hölder inequality in $L^1(\tilde{w})$ with $p=\frac{4-\alpha}{2-\alpha}$, $q=\frac{4-\alpha}{2}$ we obtain

$$\int_0^1 \tilde{w}v^2 < \Big(\int_0^1 \tilde{w}v^4\Big)^{1/p}\Big(\int_0^1 \tilde{w}v^\alpha\Big)^{1/q}, \quad \int_0^1 \tilde{w}v^{2+\alpha} < \Big(\int_0^1 \tilde{w}v^4\Big)^{1/q}\Big(\int_0^1 \tilde{w}v^\alpha\Big)^{1/p}.$$

This shows $\int_0^1 w_\lambda^1(v)v^\alpha dv<0$ and thus $\int_0^{U_n} w_\lambda^{U_n}(u)\int(1-\cos(ux))\nu(dx)\,du \gtrsim U_n^\alpha$.

The analysis of the statistical error is exactly as in Section 4. □

Interestingly, by the optimal choice of U_n for σ_n^2 we can achieve the rate $(\log n)^{-(2-\alpha)}$, which can be shown to be minimax optimal (with respect to jump components whose characteristic function decays at most like $e^{-c|u|^\alpha}$ for $|u|\to\infty$, $c>0$), without any regularity or density assumptions on the Lévy measure. Because of $\nu_n(\mathbb{R})=\lambda_n$ the estimated Lévy measure ν_n will be large, but relatively smooth around zero. So, the quantitative estimates around zero might be bad, but the large intensity of small jumps will be captured. The analysis for γ_n is slightly more delicate. Especially, in the unbounded variation case the *drift* is not well defined and depending on the symmetry or asymmetry of ν around zero the drift estimate might remain bounded or diverge (to compensate for the small jumps).

Figure 3 shows the performance of the estimate σ_n^2 in finite and infinite intensity cases. In particular, we simulate 200 samples each of length $n=1000$ from the distribution of X_1, where

(0) X is a finite jump activity Lévy process with parameters $\sigma^2=0.25$, $\gamma=0$, $\nu=0$ (*Brownian motion*)

(1) X is a finite jump activity Lévy process with parameters $\sigma^2=0.25$, $\gamma=0$, $\nu(dx)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx$ (*Brownian motion plus compound Poisson process*)

(2) X is an infinite jump activity Lévy process with parameters $\sigma^2=0.25$, $\gamma=0$, $\nu(dx)=x^{-1}e^{-x}1_{\{x\geq 0\}}\,dx$ (*Brownian motion plus Gamma process*)

(3) X is an infinite jump activity Lévy process with parameters $\sigma^2=0.25$, $\gamma=0$, $\nu(dx)=(2\pi)^{-1}|x|^{-1}K_1(|x|)\,dx$, where K_1 is the modified Bessel function of the second kind (*Brownian motion plus Normal Inverse Gaussian process*, see [5])

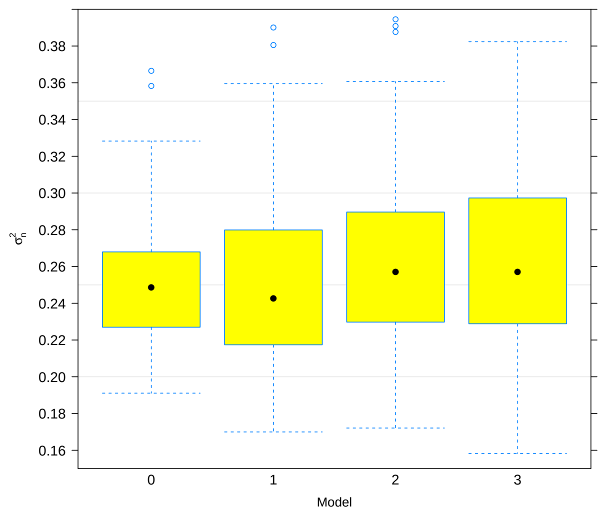

Fig. 3 The box plots of the estimate σ_n^2 in Normal (0), Merton (1), Gamma (2) and NIG (3) Lévy models.

In all four above models we construct σ_n^2 via (4.2) with $\widetilde{w}(u) = 1_{\{0.5 \le u < 1\}}$ and optimally chosen U_n. We see that the performance of σ_n^2 remains reasonable even in the case of infinite activity Lévy models.

6.2 Spectral estimation for general Lévy measures

In this section we work under the assumption of a finite second moment $\int x^2 d\nu$ for an otherwise completely general Lévy measure ν. Then the measure

$$\nu_\sigma(dx) = \sigma^2 \delta_0(dx) + x^2 \nu(dx)$$

is a finite measure. This measure is an intrinsic object in the description of a Lévy process because

$$\psi''(u) = -\sigma^2 + \int (ix)^2 e^{iux} \nu(dx) = -\mathscr{F}\nu_\sigma(u)$$

holds and the Kolmogorov representation of the characteristic function reads

$$\varphi_t(u) = e^{t\psi(u)}, \quad \psi(u) = i\gamma u + \int \frac{e^{iux} - 1 - iux}{x^2} \nu_\sigma(dx),$$

where the integrand is continuously extended to $-u^2/2$ at $x=0$. The weak topology of infinitely divisible distributions is completely described by the pair (γ,ν_σ), which follows by the convergence of the characteristic functions, see also Theorem VII.2.9 and Remark VII.2.10 in Jacod and Shiryaev [38] or Theorem 19.1 in Gnedenko and Kolmogorv [33].

Proposition 6.2. *The convergence* $\mathbb{P}_{\sigma_m^2,\gamma_m,\nu_m} \xrightarrow{w} \mathbb{P}_{\sigma^2,\gamma,\nu}$ *for a sequence of triplets* $(\sigma_m^2,\gamma_m,\nu_m)_{m\geqslant 1}$ *takes place if and only if* $\gamma_m \to \gamma$ *and* $\nu_{\sigma,m} \xrightarrow{w} \nu_\sigma$ *(weak convergence for finite measures).*

In particular, the result shows that the diffusion part σ and the small jump part in ν cannot be disentangled unless a (statistically sometimes artificial) bound on the intensity of small jumps is imposed. A natural object of inference is therefore the measure ν_σ itself, which, however, will not have any smoothness properties due to the point measure at zero (for $\sigma>0$). [48] therefore treat the estimation of functionals $\int f d\nu_\sigma$ for suitable integrands f, using a general minimum-distance estimator. In line with the approaches taken here, we consider a spectral estimator for ν_σ, but measure its performance in a weak norm, namely in the Sobolev space $H^{-1}(\mathbb{R})$, the dual of the Sobolev space $H^1(\mathbb{R})$ of functions $f\in L^2(\mathbb{R})$ with weak derivative $f'\in L^2(\mathbb{R})$. Note that all finite measures on $\mathbb{R}$ lie in $H^{-1}(\mathbb{R})$. In the spectral domain we shall use

$$\|g\|_{H^{-1}} = \frac{1}{\sqrt{2\pi}}\|(1+u^2)^{-1/2}\mathscr{F}g(u)\|_{L^2}$$

such that for $H^1(\mathbb{R})$-functions f we can analyse the error of functional estimation via (H^1,H^{-1})-duality by

$$\Big|\int f d\nu_{\sigma,n} - \int f d\nu_\sigma\Big| \leqslant \|f\|_{H^1}\|\nu_{\sigma,n}-\nu_\sigma\|_{H^{-1}}.$$

The estimator $\nu_{\sigma,n}$ is again based upon a plug-in approach. Owing to $\nu_\sigma = \mathscr{F}^{-1}[-\psi'']$, we use a kernel function $K\in L^1(\mathbb{R})$ which satisfies $\int K = 1$ and is band-limited: $supp(\mathscr{F}K)\subseteq[-1,1]$. We define for some bandwidth $h>0$

$$\nu_{\sigma,n} := -\mathscr{F}^{-1}[\psi_n''\mathscr{F}K_h] = \frac{-1}{\Delta}\mathscr{F}^{-1}\big[(\varphi_n''/\varphi_n - (\varphi_n'/\varphi_n)^2)\mathscr{F}K_h\big]. \tag{6.1}$$

Here we suppose that φ_n does not vanish on $[-h^{-1},h^{-1}]$, the maximal support of $\mathscr{F}K_h$, which for suitable h is the case with overwhelming probability, see Lemma 6.4 below.

We obtain the error decomposition for $\nu_{\sigma,n}$

$$\nu_{\sigma,n}-\nu_\sigma := \underbrace{-\mathscr{F}^{-1}[\mathscr{F}K_h(\psi_n''-\psi'')]}_{\text{statistical error}} - \underbrace{\mathscr{F}^{-1}[(\mathscr{F}K_h - 1)\psi'']}_{\text{approximation error}}. \tag{6.2}$$

The approximation error is handled by standard methods since

$$-\mathscr{F}^{-1}[(\mathscr{F}K_h - 1)\psi''] = K_h * \nu_\sigma - \nu_\sigma$$

holds. In particular, when assuming no regularity for ν_σ, we obtain the general order $O(h^{1/2})$ for this error.

Lemma 6.3. *Suppose that the kernel K satisfies $\int |\eta|^{1/2} K(\eta)d\eta < \infty$. Then we have as $h \to 0$*

$$\|K_h * \nu_\sigma - \nu_\sigma\|_{H^{-1}} \lesssim h^{1/2}.$$

Proof. We calculate by the dual definition of H^{-1}, $\int K = 1$ and by the Cauchy-Schwarz inequality:

$$\begin{aligned}
\|K_h * \nu_\sigma - \nu_\sigma\|_{H^{-1}} &= \sup_{\|f\|_{H^1}=1} \Big| \int f\, d(K_h * \nu_\sigma - \nu_\sigma) \Big| \\
&= \sup_{\|f\|_{H^1}=1} \Big| \int (f * K_h(-\bullet) - f)\, d\nu_\sigma \Big| \\
&\leqslant \sup_{\|f\|_{H^1}=1} \sup_{x\in\mathbb{R}} |(f * K_h(-\bullet) - f)(x)| \|\nu_\sigma\|_{TV} \\
&\lesssim \sup_{\|f\|_{H^1}=1} \sup_{x\in\mathbb{R}} \Big| \int (f(x) - f(x+y)) K_h(y)\, dy \Big| \\
&\leqslant \sup_{\|f'\|_{L^2}=1} \sup_{x\in\mathbb{R}} \Big| \int \Big(\int f'(z) \mathbf{1}_{[x,x+y]}(z)\, dz \Big) K_h(y)\, dy \Big| \\
&\leqslant \int |y|^{1/2} |K_h(y)|\, dy \\
&= \int h^{1/2} |\eta|^{1/2} K(\eta)\, d\eta \lesssim h^{1/2}.
\end{aligned}$$

□

The next lemma identifies the main term in the statistical error.

Lemma 6.4. *Introduce*

$$M_h := \max_{k=0,1,2} \sup_{|u|\leqslant 1/h} |(1/\varphi)^{(k)}(u)|.$$

If $M_{h_n} = o(n^{1/2} \log(h_n^{-1})^{-(1+\delta)/2})$ holds for a sequence $h_n \to 0$ and some $\delta > 0$, then we have

$$\mathscr{F}^{-1}[\mathscr{F}K_{h_n} \Delta(\psi_n'' - \psi'')](x) = \mathscr{F}^{-1}[\mathscr{F}K_{h_n}(\varphi^{-1}(\varphi_n - \varphi))''](x) + R_n(x)$$

with a second order term R_n satisfying

$$\|R_n\|_{H^{-1}} = O_P(M_{h_n}^2 n^{-1} \log(h_n^{-1})^{1+\delta}).$$

Proof. To linearise $\psi_n'' - \psi'' = \Delta^{-1}(\log(\varphi_n/\varphi))''$, we set $F(y) = \log(1+y)$, $\eta = (\varphi_n - \varphi)/\varphi$, and use

$$\begin{aligned}(F\circ\eta)''(u) &= F'(\eta(u))\eta''(u) + F''(\eta(u))\eta'(u)^2 \\ &= F'(0)\eta''(u) + O\Big(\|F''\|_\infty\Big(\|\eta\|_\infty\|\eta''\|_\infty + \|\eta'\|_\infty^2\Big)\Big),\end{aligned}$$

where the supremum norms are taken over the ranges of u and $\eta(u)$, respectively. On the event $\Omega_n := \{\|(\varphi_n-\varphi)/\varphi\|_{L^\infty[-h^{-1},h^{-1}]} \leqslant 1/2\}$ the values η are in $[-1/2, 1/2]$ and we obtain the error estimate

$$\begin{aligned}&\sup_{|u|\leqslant h^{-1}} |(\log(\varphi_n/\varphi))''(u) - ((\varphi_n-\varphi)/\varphi)''(u)| \\ &= O\Big(\max_{k=0,1,2}\|((\varphi_n-\varphi)/\varphi)^{(k)}\|^2_{L^\infty[-h^{-1},h^{-1}]}\Big) = O\Big(M_h^2 \max_{k=0,1,2}\|(\varphi_n-\varphi)^{(k)}\|^2_{L^\infty[-h^{-1},h^{-1}]}\Big).\end{aligned}$$

From Theorem 3.4 we know that under our moment assumption on ν (for $k=0,1,2$ and any $\delta>0$)

$$\|(\varphi_n-\varphi)^{(k)}\|_{L^\infty[-h^{-1},h^{-1}]} = O_P(n^{-1/2}\Delta^{(k\wedge 1)/2}\log(h^{-1})^{(1+\delta)/2}).$$

Together with the growth assumption on M_{h_n} this shows $\mathbb{P}(\Omega_n)\to 1$ and then

$$\sup_{|u|\leqslant h_n^{-1}} |\Delta(\psi_n''(u)-\psi''(u)) - ((\varphi_n-\varphi)/\varphi)''(u)| = O_P(M_{h_n}^2 n^{-1}\log(h_n^{-1})^{1+\delta}).$$

Integration over u in the frequency domain yields the asserted bound for $\|R_n\|_{H^{-1}}$. □

The expected H^{-1}-norm of the main statistical error term is bounded using $\mathrm{Var}_{\mathbb{C}}[\varphi_n^{(k)}(u)] \leqslant n^{-1}\mathbb{E}[X_\Delta^{2k}]$ for $k=0,1,2$:

$$\begin{aligned}\mathbb{E}[\|\mathscr{F}^{-1}[\mathscr{F}K_h((\varphi_n-\varphi)/\varphi)'']\|^2_{H^{-1}}] &= \frac{1}{2\pi}\mathbb{E}[\|(1+u^2)^{-1/2}\mathscr{F}K_h((\varphi_n-\varphi)/\varphi)''\|^2_{L^2}] \\ &\lesssim M_h^2\int_{-1/h}^{1/h}(1+u^2)^{-1}\sum_{k=0}^2 \mathrm{Var}_{\mathbb{C}}[\varphi_n^{(k)}(u)]du \\ &\lesssim n^{-1}M_h^2.\end{aligned}$$

Altogether we have proved the following result, where the condition on M_h ensures that the second order term is negligible.

Proposition 6.5. *Suppose that the kernel K satisfies $\int|\eta|^{1/2}K(\eta)d\eta<\infty$ and that $h\to 0$ as $n\to\infty$ such that $M_h = o(n^{1/2}(\log(h^{-1})^{-1-\delta})$ holds for some $\delta>0$. Then the estimator $\nu_{\sigma,n}$ of ν_σ satisfies*

$$\|\nu_{\sigma,n}-\nu_\sigma\|_{H^{-1}} = O_P\big(h^{1/2}+n^{-1/2}M_h\big).$$

In the worst case (i.e., if $\sigma > 0$ and the characteristic function has Gaussian tails), we have $M_h \sim \exp(ch^{-2})$ for some $c > 0$ and the choice $h = \tilde{c}/\sqrt{\log n}$ with $\tilde{c} > \sqrt{2c}$ yields the universal rate $O_P((\log n)^{-1/4})$. If a pure compound Poisson process is observed with a jump measure satisfying $\int x^2 \nu(dx) < \infty$, then M_h is uniformly bounded in h because ψ, ψ' and ψ'' are uniformly bounded. In that case we may choose $h = n^{-1/2}$ and $\nu_{\sigma,n}$ converges with parametric rate $O_P(n^{-1/2})$ to ν_σ. Note, however, that this convergence is measured in the weak H^{-1}-norm and much finer results can be obtained, e.g. uniform central limit theorems in a Donsker-type fashion as in Nickl and Reiß [49]. All other convergence rates between logarithmic and parametric occur indeed and the choice of the bandwidth h depends heavily on the unknown (to the statistician) size M_h of the statistical error. We refer to Comte and Genon-Catalot [20] for a data-driven choice based on a model selection approach (in the finite variation case $\int |x| \nu(dx) < \infty$ and for L^2-loss). Nonparametric testing based on this general approach via ν_σ is discussed in Reiß [50].

7 Estimating the Blumenthal-Getoor index

In this section we consider the problem of estimating the Blumenthal-Getoor index of a Lévy process observed at low frequency. The results are mainly based on Belomestny [12]. An extension to more general models can be found in Belomestny and Panov [13, 14]. In the case of high-frequency data the problem was studied in Aït-Sahalia and Jacod [2].

7.1 Setup

For a one-dimensional Lévy process $X = (X_t)_{t\geq 0}$ with a Lévy measure ν, the Blumenthal-Getoor (BG) index of X is defined as

$$\mathtt{BG}(X) = \inf\left\{ r > 0 : \int_{|x|\leq 1} |x|^r \nu(dx) < \infty \right\}.$$

The Blumenthal-Getoor index is a fundamental characteristic of the Lévy process X that determines the activity of jumps of X. If $\nu([-\varepsilon, \varepsilon]) < \infty$, then the process X has finite activity of jumps and $\mathtt{BG}(X) = 0$. If the Lévy measure $\nu((-\infty, -\varepsilon] \cup [\varepsilon, \infty))$ diverges near $\varepsilon = 0$ at a rate $\varepsilon^{-\alpha}$ for some $\alpha > 0$, then the BG index of X is equal to α. From a practical point of view, the importance of the Blumenthal-Getoor index lies in the fact that it determines the smoothness properties of the marginal density of X and has significant impact on the convergence of different approximation algorithms for X (see, e.g., Dereich, [24]). Recently, the problem of estimation of the BG index from discrete observations of a Lévy process X or some other processes based on X has drawn much attention in the literature. Aït-Sahalia and Jacod, [2] (see also [4]) studied the problem of estimating the so-called jump activity index, that is defined for any Itô semimartingale Y via

$$\mathtt{JAI}(Y) = \inf\left\{ r > 0 : \sum_{0 \le s \le T} |\Delta Y_s|^r < \infty \right\},$$

where $\Delta Y_s = Y_s - Y_{s-}$ is the size of the jump at time s and T is a fixed time horizon. Note that, in general, $\mathtt{JAI}(Y)$ is a random quantity, which is to be determined pathwise. In the case of a Lévy process Y, $\mathtt{JAI}(Y)$ is deterministic and coincides with the Blumenthal-Getoor index of Y. Obviously, one can compute $\mathtt{JAI}(Y)$ if the whole path of the process Y up to time T is observed. In a more realistic situation when the process Y is observed on a discrete grid $\{0, \Delta, \ldots, \Delta n\}$ with $\Delta n = T$ and $\Delta \to 0$ as $n \to \infty$ (*high-frequency* data), Aït-Sahalia and Jacod proposed a method which is able to consistently estimate $\mathtt{JAI}(Y)$ and is based on a statistics that counts the “big” increments of the process Y. Turning to the case of *low-frequency* data, i.e., the case of fixed $\Delta > 0$ and $T \to \infty$, one may wonder if any kind of statistical inference is possible in this situation at all. The first results showing that a consistent estimation of the BG index based on the low-frequency data is possible, were obtained in Belomestny, [12] for the case of a Lévy process X. The inference in [12] relied on the kind of Abelian theorem which characterises the decay of the characteristic function of X. Such Abelian theorems are well known in the literature: Bismut [15] showed that the tail integral $\nu\big((-\infty, -x) \cup (x, +\infty)\big)$ behaves asymptotically like $x^{-\alpha}$ as $x \to +0$ for some $\alpha \in [0, 2)$ if and only if the characteristic exponent of the corresponding Lévy process X with $\sigma = 0$, $\gamma = 0$ and Lévy measure ν is of order $-|u|^\alpha$ for large $|u|$. In [12] the following deeper result was proved.

Proposition 7.1. *Let the Lévy density $\nu(x)$ of a one-dimensional Lévy process X satisfy for $\eta > 0$, $\alpha \in (0, 2)$*

$$\int_{\mathbb{R}} (e^{\mathtt{i}xu} - 1 - \mathtt{i}xu\mathbf{1}(|x| \le 1))\, \nu(dx) = -\eta |u|^\alpha \tau(u), \quad u \in \mathbb{R}, \tag{7.1}$$

where the function τ fulfills

$$\tau(u) = 1 + D_\pm |u|^{-\kappa} + o(|u|^{-\kappa}), \quad u \to \pm\infty \tag{7.2}$$

with some constants $\kappa \in (0, 1)$, D_+ and D_-. Then

$$\int_{|x| < \varepsilon} x^2 \nu(x)\, dx = c\varepsilon^{2-\alpha} \vartheta(\varepsilon), \tag{7.3}$$

where $c > 0$ is a constant depending on η and α and the function $\vartheta(\varepsilon)$ satisfies

$$|\vartheta(\varepsilon)-1| \lesssim |\varepsilon|^{\kappa}, \quad \varepsilon \to 0.$$

It is clear that the parameter α in (7.3) coincides with the BG index of X. Thus the asymptotic behaviour of $\psi(u)$ for large u is connected to the BG index of X. This fact can be used to infer on α using the spectral approach of Section 4. Consider a Lévy process X with

$$\psi(u) = \mathtt{i}\gamma u + \vartheta(u), \quad \gamma \in \mathbb{R}, \tag{7.4}$$

where the function ϑ is of the form

$$\vartheta(u) = -\eta |u|^{\alpha} \tau(u) \tag{7.5}$$

with $\mathrm{Re}[\tau(u)] > 0$ for $u \in \mathbb{R} \setminus \{0\}$ and $\tau(u) \to 1$ as $|u| \to \infty$. The formula

$$\begin{aligned} \mathscr{Y}(u) &:= \log(-\log(|\varphi(u)|^2)) \\ &= \log(2\eta\Delta) + \alpha \log(u) + \log(\mathrm{Re}\,\tau(u)), \quad u > 0, \end{aligned} \tag{7.6}$$

with $\varphi(u) = \exp(\Delta\psi(u))$ suggests now how to estimate α from φ. Indeed, in terms of the new "data" $\mathscr{Y}$ we have a linear semiparametric problem with the non-parametric part $\log(\mathrm{Re}\,\tau(u))$ which can be viewed as a "nuisance" parameter. Since $\log(\mathrm{Re}\,\tau(u)) \to 0$ as $|u| \to \infty$, we can get rid of this component by using frequencies u with large $|u|$. On the other hand, if we plug-in an estimate φ_n instead of φ, the variance of $\mathscr{Y}_n(u) := \log(-\log(|\varphi_n(u)|^2))$ will increase exponentially with $|u|$ (because of the exponential decay of $\varphi(u)$) and we have to regularize the problem by damping (or cutting off) high frequencies. An appropriate weighting scheme would allow to take both effects into account. Let

$$\widetilde{w}^{U_n}(u) := (1/U_n)\, \widetilde{w}(u/U_n)$$

with a bounded non-constant function $\widetilde{w}(u)$ supported on the interval $[0,1]$, such that $\widetilde{w}(u) > 0$ on $(0,1)$ and $\int |\widetilde{w}(u)| \log^2(u)\, du < \infty$. Consider the optimisation problem

$$(\mu_n, \alpha_n) := \underset{(\mu,\alpha)}{\operatorname{argmin}} \int_0^{\infty} \widetilde{w}^{U_n}(u)(\mathscr{Y}_n(u) - \alpha\log(u) - \mu)^2\, du. \tag{7.7}$$

As can be easily seen, the solution α_n of (7.7) is equal to

$$\alpha_n = \int_0^{\infty} w_{\alpha}^{U_n}(u) \mathscr{Y}_n(u)\, du, \tag{7.8}$$

where

$$w_{\alpha}^{U_n}(u) := \widetilde{w}^{U_n}(u) \frac{\int_0^{\infty} \widetilde{w}^{U_n}(s)\log(s)\,ds - \left(\int_0^{\infty} \widetilde{w}^{U_n}(s)\,ds\right)\log(u)}{\left(\int_0^{\infty} \widetilde{w}^{U_n}(s)\log(s)\,ds\right)^2 - \int_0^{\infty} \widetilde{w}^{U_n}(s)\log^2(s)\,ds \bullet \int_0^{\infty} \widetilde{w}^{U_n}(s)\,ds}.$$

As a result $w_\alpha^{U_n}(u)$ satisfies

$$\int_0^{U_n} \log(u) w_\alpha^{U_n}(u)\,\mathrm{d}u = 1, \quad \int_0^{U_n} w_\alpha^{U_n}(u)\,\mathrm{d}u = 0. \tag{7.9}$$

In the next section we discuss the convergence of the estimate α_n.

7.2 Minimax upper bounds

To state minimax upper bounds we first need to specify a class of Lévy processes.

Definition 7.2. *Let* $\mathscr{A}(\bar{\alpha}, \eta_-, \eta_+, \varkappa, c_\tau)$ *denote the class of Lévy processes with characteristic exponents of the form*

$$\psi(u) = \mathrm{i}\gamma u + \vartheta(u), \quad \vartheta(u) = -\eta |u|^\alpha \tau(u), \quad u \in \mathbb{R}, \tag{7.10}$$

where $0 < \alpha \le \bar{\alpha} \le 2$*,*

$$0 < \eta_- \le \eta \le \eta_+ < \infty \tag{7.11}$$

and

$$|1 - \tau(u)| \le \frac{c_\tau}{|u|^\varkappa}, \quad |u| \to \infty \tag{7.12}$$

for some $0 < \varkappa \le \alpha$ *and* $c_\tau > 0$.

We will write

$$(\alpha, \eta, \tau) \in \mathscr{A}(\bar{\alpha}, \eta_-, \eta_+, \varkappa, c_\tau)$$

to indicate that the Lévy process with characteristics (α, η, τ) is in the class $\mathscr{A}$. The following theorem shows that the uniform convergence rates of α_n over the class $\mathscr{A}$ are of order $\log^{\varkappa/\bar{\alpha}}(n)$.

Theorem 7.3. *Choosing for* $\beta = 1 + \varkappa/\bar{\alpha}$

$$U_n = \left[\frac{1}{2\eta_+\Delta} \log\left(n \log^{-\beta}(n)\right)\right]^{1/\bar{\alpha}},$$

we obtain for the risk of α_n *the uniform convergence rate*

$$\alpha_n - \alpha = O_{\mathrm{P},\mathscr{A}}\left((\Delta/\log(n))^{\varkappa/\bar{\alpha}}\right). \tag{7.13}$$

Remark 7.4.

(a) The convergence rates depend on $\bar{\alpha}$, the prior upper bound for α. If there is no prior information on $\bar{\alpha}$ one may take $\bar{\alpha} = 2$.
(b) The case of Lévy processes X with $\sigma > 0$ can be handled in a similar way. Indeed, consider a Lévy process X with the characteristic exponent of the form

$$\psi(u) = \mathtt{i}\gamma u - \sigma^2 u^2/2 + \vartheta(u), \quad \mu \in \mathbb{R}, \sigma^2 > 0. \tag{7.14}$$

Fixe some $\xi > 2$ and introduce the function

$$\vartheta_\xi(u) := \xi^2 \operatorname{Re}(\psi(u)) - \operatorname{Re}(\psi(\xi u)),$$

then we have

$$\vartheta_\xi(u) = -c_\xi(\alpha)|u|^\alpha \tau_\xi(u),$$

where $c_\xi(\alpha) = \eta(\xi^2 - \xi^\alpha)$ and $\tau_\xi(u)$ fulfills

$$|1 - \tau_\xi(u)| \lesssim \frac{1}{|u|^\varkappa}, \quad |u| \to \infty. \tag{7.15}$$

Thus $\vartheta_\xi(u)$ has a structure similar to the structure of $\vartheta(u)$ in (7.10) and we can carry over the results of the previous section to a class of Lévy models with $\sigma^2 > 0$.

Let us describe the main steps in the proof of Theorem 7.3. We replace $\mathscr{Y}_n$ by $\mathscr{Y}$ in (7.8) and introduce

$$\bar{\alpha}_n := \int_0^\infty w_\alpha^{U_n}(u)\mathscr{Y}(u)\,du. \tag{7.16}$$

First one can get by (7.12) the following bound for the "model bias" $\bar{\alpha}_n - \alpha$:

$$|\bar{\alpha}_n - \alpha| = \left| \int_0^\infty w_\alpha^{U_n}(u) \log(\operatorname{Re} \tau(u))\,du \right| \le C U_n^{-\varkappa}$$

with some constant $C > 0$ depending on c_τ. Next using the Taylor expansion of the function $\log(-\log(x))$, we get for the statistical error $\alpha_n - \bar{\alpha}_n$:

$$\mathbb{P}\left\{ |\alpha_n - \bar{\alpha}_n| > C_1 \frac{\varepsilon_n}{c_\tau U_n^\alpha} \right\} \le C_2 n^{-1-\delta},$$

with some $\delta > 0$, provided

$$\varepsilon_n := \frac{\log n}{\sqrt{n}}\, e^{2\eta \Delta U_n^\alpha} \to 0, \quad U_n \to \infty, \quad n \to \infty.$$

Finally choosing U_n in a proper way, that balances the deterministic and stochastic errors (bias-variance trade-off), we arrive at (7.13).

Example 7.5. Let us consider the *generalized hyperbolic* (GH) Lévy model which was introduced in [27]. The characteristic function Φ_{GH} of increments in the GH Lévy model with parameters $(\kappa, \beta, \delta, \lambda)$ is given by

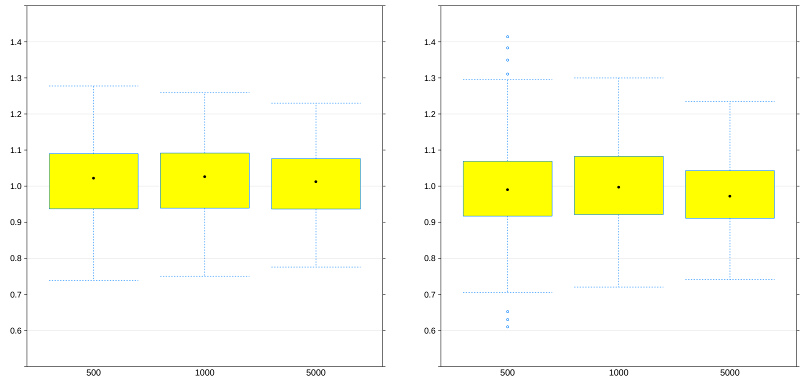

Fig. 4 Box plots of the estimate α_n in the GH Lévy model with (right) and without (left) the Brownian component in dependence on n.

$$\Phi_{GH}(u) = \frac{\left(\sqrt{\kappa^2+\beta^2}\right)^{\lambda}}{\left(\sqrt{\kappa^2-(\beta+\mathrm{i}u)^2}\right)^{\lambda}} \frac{K_\lambda\left(\delta\sqrt{\kappa^2-(\beta+\mathrm{i}u)^2}\right)}{K_\lambda\left(\delta\sqrt{\kappa^2+\beta^2}\right)},$$

where K_λ is the modified Bessel function of the second kind. Φ_{GH} has the Lévy-Khintchine representation

$$\Phi_{GH}(u) = \exp\left(-\sigma^2 u^2/2 + \int_{-\infty}^{\infty}(e^{\mathrm{i}ux}-1-\mathrm{i}ux)g(x)\,dx\right).$$

The function $g(x)$, the density of the corresponding Lévy measure ν, can be represented in an integral form. From this representation the following expansion for $\bar{g}(x) = x^2 g(x)$ can be obtained

$$\bar{g}(x) = \frac{\delta}{\pi} + \frac{\lambda+\frac{1}{2}}{2}|x| + \frac{\delta\beta}{\pi}x + o(|x|), \quad x \to 0.$$

A direct consequence of this expansion is that

$$\int_{|x|>\varepsilon} g(x)\,dx \asymp 1/\varepsilon, \quad \varepsilon \to 0$$

and hence the BG index of the GH Lévy model is equal to 1. In our simulation study we simulate the GH Lévy process X with parameters $\beta = 0$, $\lambda = 1$, $\kappa = 1$ and $\delta = 5$ at $n+1$ equidistant points $\{0,\ldots,n\}$. The results of the BG index estimation are presented in Figure 4, where 1000 samples of the length n are used to construct the box plots. The left side of Figure 4 corresponds to the GH Lévy model with no diffusion part and the right side deals with the case of a non-zero diffusion part ($\sigma = 0.5$).

7.3 Minimax lower bounds

In this section we address the question of optimality of the rates in (7.13).

Theorem 7.6. *We have*

$$\lim_{K\to\infty}\limsup_{n\to\infty}\inf_{\alpha_n}\sup_{(\alpha,\eta,\tau)\in\mathscr{A}}\mathbb{P}_{(\alpha,\eta,\tau)}\Big(|\alpha_n-\alpha|>K(\Delta/\log(n))^{\varkappa/\bar{\alpha}}\Big)>0. \quad (7.17)$$

Proof. The proof is based on Theorem 5.6. First we turn to the construction of models f_0 and f_1. Let us consider a symmetric stable Lévy process with

$$\psi(u)=\mathtt{i}\mu u+\vartheta(u),\quad \vartheta(u)=-\eta_+|u|^{\alpha},\quad 0<\alpha\le 1,\quad u\in\mathbb{R}$$

Now for any δ satisfying $0<\delta<\alpha$ and $M>0$ define

$$\psi_\delta(u):=\mathtt{i}\mu u+\vartheta_\delta(u),$$

where

$$\vartheta_\delta(u):=-\eta_+|u|^{\alpha}\Big(\mathbf{1}(|u|\le M)+\frac{M^{\delta}(1+c|u|^{-\varkappa})}{|u|^{\delta}(1+cM^{-\varkappa})}\mathbf{1}(|u|>M)\Big).$$

Then $\varphi_\delta(u):=\exp(\mathtt{i}\mu u+\vartheta_\delta(u))$ is the characteristic function of some Lévy process and

$$\varphi_\delta(u)=\varphi(u),\quad |u|\le M,$$

where $\varphi(u)=\exp(\mathtt{i}\mu u+\vartheta(u))$. Indeed, the function $\vartheta_\delta(u)$ is a continuous, non-positive, symmetric function which is convex on $\mathbb{R}_+$ for large enough M and small enough $c>0$. According to the well known Pólya criteria (see e.g. [61], Theorem 1.3.8), the function $\exp(\xi\vartheta_\delta(u))$ is a characteristic function of some absolutely continuous distribution for any $\xi>0$. In particular, for any natural q the function $\exp(\vartheta_\delta(u)/q)$ is a characteristic function of some absolutely continuous distribution. Hence, $\exp(\vartheta_\delta(u))$ is a characteristic function of some infinitely divisible distribution. Define

$$f_0=(\alpha,\eta_+,1),\quad f_1=(\alpha-\delta,\eta_+,\tau_{\delta,M}) \quad (7.18)$$

and $\varphi_{f_0}(u)=\varphi(u)$, $\varphi_{f_1}(u)=\varphi_\delta(u)$ with

$$\tau_{\delta,M}(u):=|u|^{\delta}\mathbf{1}(|u|\le M)+\frac{M^{\delta}}{(1+cM^{-\varkappa})}(1+c|u|^{-\varkappa})\mathbf{1}(|u|>M).$$

If $M^{\delta}=1+cM^{-\varkappa}$, i.e.

$$\delta=\log(1+cM^{-\kappa})/\log M\asymp cM^{-\kappa}/\log M,\quad M\to\infty, \quad (7.19)$$

then

$$|\tau_{\delta,M}(u)-1| \lesssim |u|^{-\varkappa}, \quad |u| \to \infty$$

and hence $f_1 \in \Theta = \mathscr{A}(\bar{\alpha}, \eta_-, \eta_+, \varkappa)$. Furthermore one can show that

$$\chi^2(p_{f_0}^{\otimes n}, p_{f_1}^{\otimes n}) = n\chi^2(p_{f_0}, p_{f_1}) \lesssim M^{7-\alpha+\delta} e^{-2\eta M^{\alpha-\delta}}$$

and the choice $M \asymp \left[\frac{1}{2\eta_+}\log\left(n\log^{-\beta}(n)\right)\right]^{1/(\alpha-\delta)}$ with $\beta \geq \frac{7-(\alpha-\delta)}{2(\alpha-\delta)}$ yields

$$\chi^2(p_{f_0}^{\otimes n}, p_{f_1}^{\otimes n}) < 1$$

for large enough n. □

8 Spectral estimation of time-changed Lévy processes

In this section we are going to study the problem of estimating the characteristics of a multidimensional Lévy process X from the low-frequency observations $Y_0, Y_\Delta, \ldots, Y_{n\Delta}$ of the time changed Lévy process $Y_t = X_{\mathscr{T}(t)}$. The presentation follows Belomestny [9].

8.1 Setting

The main difficulty in constructing nonparametric estimates for the Lévy density ν of X lies in the fact that the jumps are unobservable variables since in practice only discrete observations of the process Y are available. The more high-frequent the observations are, the more relevant information about the jumps of the underlying process are contained in the sample. Such a high-frequency based statistical approach has played a central role in the recent literature on nonparametric estimation for Lévy-type processes. For instance, under discrete observations of a pure Lévy process X_t at times $t_j = j\Delta$, $j = 0, \ldots, n$, [62] proposed the quantity

$$\widehat{\beta}(f) = \frac{1}{n\Delta}\sum_{k=1}^{n} f(X_{t_k} - X_{t_{k-1}})$$

as a consistent estimator for the functional

$$\beta(f) = \int f(x)\nu(x)\,dx,$$

where f is a given bounded “test function”. Turning back to the time-changed Lévy processes, it was shown in [30] that in the case where the rate process ρ in

$$\mathscr{T}(t) = \int_0^t \rho(s_-)\,ds,$$

is a positive ergodic diffusion independent of the Lévy process X, $\widehat{\beta}(f)$ is still a consistent estimator for $\beta(f)$ up to a constant, provided the time horizon $n\Delta$ and the sampling frequency Δ^{-1} converge to infinity at suitable rates. In the case of low-frequency data (Δ is fixed) we cannot be sure to what extent the increment $X_{t_k} - X_{t_{k-1}}$ is due to one or several jumps or just to the diffusion part of the Lévy process, so that at first sight it may appear surprising that some kind of inference in this situation is possible at all. Suppose that the sequence $\mathscr{T}(j\Delta) - \mathscr{T}((j-1)\Delta),\ j = 1,\ldots,n$, is stationary and ergodic with the invariant stationary distribution π, then for any bounded "test function" f

$$\frac{1}{n}\sum_{j=1}^{n} f\left(X_{\mathscr{T}(t_j)} - X_{\mathscr{T}(t_{j-1})}\right) \to \mathbb{E}_\pi[f(X_{\mathscr{T}(\Delta)})], \quad n \to \infty. \tag{8.1}$$

The limiting expectation in (8.1) is then given by

$$\mathbb{E}_\pi[f(X_{\mathscr{T}(\Delta)})] = \int_0^\infty \mathbb{E}[f(X_s)]\,\pi(ds).$$

Taking $f(z) = \exp(\mathtt{i}\langle u, z\rangle)$ for some $u \in \mathbb{R}^d$ we arrive at the following representation for the characteristic function of $X_{\mathscr{T}(s)}$:

$$\mathbb{E}\left[\exp\left(\mathtt{i}\langle u, X_{\mathscr{T}(\Delta)}\rangle\right)\right] = \int_0^\infty \exp(t\psi(u))\,d\pi(dt) = \mathscr{L}_\Delta(-\psi(u)), \tag{8.2}$$

where $\psi(u)$ is the characteristic exponent of the Lévy process X and $\mathscr{L}_\Delta$ is the Laplace transform of π. In fact, the most difficult part of the estimation procedure consists in reconstructing the characteristics of the underlying Lévy process X from an estimate for $\mathscr{L}_\Delta(-\psi(u))$. Taking into account (2.3), we can reformulate our problem as a problem of semi-parametric estimation of the characteristic exponent ψ under the structural assumption (2.3) from an empirical estimate of (8.2) based on the observations of Y. The formula (8.2) shows that the characteristic function

$$\varphi_Y(u|\Delta) = \mathbb{E}\left[\exp\left(\mathtt{i}\langle u, X_{\mathscr{T}(\Delta)}\rangle\right)\right]$$

can be viewed as a composite function and our statistical problem is hence closely related to the problem of statistical inference on the components of a composite function. The latter type of problems in a regression setup has attracted much attention recently (see, e.g., [36] and [39]). Our problem has, however, some features not reflected in that literature. First, the unknown link function $\mathscr{L}_\Delta$ is a completely monotone function, since it is the Laplace transform of the time change $\mathscr{T}(\Delta)$. Second, the complex-valued function ψ is of the form (2.3) implying, for example, a certain asymptotic behaviour of

$\psi(u)$ as $u \to \infty$. Finally, we are not in a regression setup and $\varphi_Y(u|\Delta)$ needs to be estimated from its empirical counterpart

$$\widehat{\varphi}(u) = \frac{1}{n}\sum_{j=1}^{n} e^{\mathtt{i}\langle u, Y_{\Delta j} - Y_{\Delta(j-1)}\rangle}.$$

8.2 Specification analysis

It is clear that without further restrictions on the class of time-changed Lévy processes our problem of estimating ν is not well defined because even in the case of a perfectly known distribution of the process Y the parameters of the Lévy process X are in general not identifiable. Moreover, the corresponding statistical procedure will suffer from the "curse of dimensionality" as the dimension d increases. In order to avoid these undesirable features, we have to impose additional restrictions on the structure of the time-changed process Y. In the statistical literature one can find basically two types of restricted composite models: additive models and single-index models. While the latter class of models is too restrictive in our situation, the former one naturally appears if one assumes the independence of the components of X_t. Here, we study the class of time-changed Lévy processes satisfying the following two assumptions.

Assumption 8.1

(AXI) The Lévy process X_t has independent components such that at least two of them are non-zero, i.e.,

$$\varphi_Y(u|\Delta) = \mathscr{L}_t(-\psi_1(u_1) - \ldots - \psi_d(u_d)), \tag{8.3}$$

where ψ_k, $k = 1, \ldots, d$, are the characteristic exponents of the components of X_t of the form

$$\psi_k(u) = \mathtt{i}\mu_k u - \sigma_k^2 u^2/2 + \int_{\mathbb{R}} \left(e^{\mathtt{i}ux} - 1 - \mathtt{i}ux\bullet\mathbf{1}_{\{|x|\le 1\}}\right) \nu_k(dx), \tag{8.4}$$

for $k = 1, \ldots, d$, and

$$|\mu_l| + \sigma_l^2 + \int_{\mathbb{R}} x^2 \nu_l(dx) \neq 0 \tag{8.5}$$

for at least two different indices l.

(ATI) The time change process $\mathscr{T}$ is normalised to $\mathbb{E}[\mathscr{T}(t)] = t$.

The advantage of the modeling framework (8.3) is twofold. On the one hand, models of this type are rather flexible: the distribution of Y_t for a fixed t is in general determined by $d+1$ non-parametric components and $2 \times d$ parametric ones. On the other hand, these models remain parsimonious and, as we will see later, admit statistical inference not suffering from the "curse of dimensionality" as d becomes large. The latter feature of our model is in accordance with the well documented behaviour of additive models in the regression setting and may become particularly important if one is going to use it, for instance, to model large portfolios of assets. The non-degeneracy assumption (8.5) basically excludes one-dimensional models and is not restrictive since it can be always checked prior to estimation by testing that

$$-\partial_{u_l u_l}\varphi_n(u)|_{u=0} = \frac{1}{n}\sum_{j=1}^{n}\left(Y_{\Delta j,l} - Y_{\Delta(j-1),l}\right)^2 > 0$$

for at least two different indices l.

Remark 8.2. Note that the condition (ATI) ensures the identifiability in our model and is often used in financial literature to guarantee the unbiased reflection of calendar time.

Let us make a few remarks on the one-dimensional case, where

$$\varphi_Y(u|t) = \mathscr{L}_t(-\psi_1(u)), \quad t \geq 0. \tag{8.6}$$

If $\mathscr{L}_\Delta$ is known, i.e., the distribution of the random variable $\mathscr{T}(\Delta)$ is known, we can consistently estimate the Lévy measure ν_1 by inverting $\mathscr{L}_\Delta$. In the case when the function $\mathscr{L}_\Delta$ is unknown, one needs rather restrictive assumptions to ensure identifiability. Indeed, consider the class of one-dimensional Lévy processes of the so-called compound exponential type with the characteristic exponent of the form

$$\psi(u) = \log\left[\frac{1}{1-\widetilde{\psi}(u)}\right],$$

where $\widetilde{\psi}(u)$ is the characteristic exponent of another one-dimensional Lévy process $\widetilde{L}_t$. It is well known (see, e.g., Section 3 in Chapter 4 of [56]) that $\exp(\psi(u))$ is the characteristic function of some infinitely divisible distribution if $\exp(\widetilde{\psi}(u))$ is. Introduce

$$\widetilde{\mathscr{L}_\Delta}(z) = \mathscr{L}_\Delta(\log(1+z)).$$

As can be easily seen, the function $\widetilde{\mathscr{L}_\Delta}$ is completely monotone with $\widetilde{\mathscr{L}_\Delta}(0) = 1$ and $\widetilde{\mathscr{L}'_\Delta}(0) = \mathscr{L}'_\Delta(0)$. Moreover, it fulfills $\widetilde{\mathscr{L}_\Delta}(-\widetilde{\psi}(u)) = \mathscr{L}_\Delta(-\psi(u))$ for all $u \in \mathbb{R}$. The existence of the time change (increasing) process $\mathscr{T}$ with the given marginal $\mathscr{T}(\Delta)$ can be derived from the general theory of stochastic partial ordering (see, [40]). The above construction shows that even under the assumption $\mathbb{E}[\mathscr{T}(t)] = t, t \geq 0$, one cannot, in general, consistently estimate the parameters of the one-dimensional time-changed Lévy process Y_t from the low-frequency observations.

8.3 Main ideas

Assume that the Lévy measures of the component processes $X_t^1,\dots,X_t^d$ are absolutely continuous with densities $\nu_1(x),\dots,\nu_d(x)$ that satisfy

$$\int_{\mathbb{R}} x^2 \nu_k(x)\,dx < \infty, \quad k=1,\dots,d.$$

Consider the functions (see also Section 6.2)

$$\bar{\nu}_k(x) := x^2 \nu_k(x), \quad k=1,\dots,d.$$

By differentiating ψ_k two times, we get

$$\psi_k''(u) = -\sigma_k^2 - \int_{\mathbb{R}} e^{\mathtt{i}ux} \bar{\nu}_k(x)\,dx.$$

For the sake of simplicity, we will assume in the sequel that the (σ_k) are known. Otherwise one should consider the derivatives of higher order, e.g.

$$\psi_k^{(4)}(u) = \int_{\mathbb{R}} e^{\mathtt{i}ux} x^4 \nu_k(x)\,dx,$$

provided $\int_{\mathbb{R}} x^4 \nu_k(x)\,dx < \infty$. Alternatively one can first estimate σ_k via the weighted least-squares approach using the asymptotic identity

$$\psi_k''(u) \to -\sigma_k^2, \quad u \to \infty.$$

Introduce the functions $\bar{\psi}_k(u) = \psi_k(u) + \sigma_k^2 u^2/2$ to get

$$\mathscr{F}[\bar{\nu}_k](u) = -\bar{\psi}_k''(u) = -\sigma_k^2 - \psi_k''(u). \tag{8.7}$$

Denote $Z = Y_\Delta$, $\varphi_k(u) = \partial_{u_k}\varphi_Z(u)$, $\varphi_{kl}(u) = \partial_{u_k u_l}\varphi_Z(u)$ and $\varphi_{jkl}(u) = \partial_{u_j u_k u_l}\varphi_Z(u)$ for $j,k,l \in \{1,\dots,d\}$ with

$$\varphi_Z(u) = \mathbb{E}\left[\exp(\mathtt{i}\langle u, Z\rangle)\right] = \mathscr{L}_\Delta(-\psi_1(u_1) - \ldots - \psi_d(u_d)). \tag{8.8}$$

Fix some $k \in \{1,\dots,d\}$ and for any real number u introduce a vector $u^{(k)} = (0,\dots,0,u,0,\dots,0) \in \mathbb{R}^d$, with u being placed at the kth coordinate of the vector $u^{(k)}$. Choose some $l \neq k$, such that the component X_t^l is not degenerated. Then we get from (8.8)

$$\frac{\varphi_k(u^{(k)})}{\varphi_l(u^{(k)})} = \frac{\psi_k'(u)}{\psi_l'(0)} \tag{8.9}$$

if $\mu_l \neq 0$ and

$$\frac{\varphi_k(u^{(k)})}{\varphi_{ll}(u^{(k)})} = \frac{\psi'_k(u)}{\psi''_l(0)} \tag{8.10}$$

in the case $\mu_l = 0$. The identities $\varphi_l(\mathbf{0}) = -\psi'_l(0)\mathscr{L}'_\Delta(0)$ and $\varphi_{ll}(\mathbf{0}) = [\psi'_l(0)]^2\mathscr{L}''_\Delta(0) - \psi''_l(0)\mathscr{L}'_\Delta(0)$ imply $\psi'_l(0) = -[\mathscr{L}'_\Delta(0)]^{-1}\varphi_l(\mathbf{0}) = \Delta^{-1}\varphi_l(\mathbf{0})$ and $\psi''_l(0) = -[\mathscr{L}'_\Delta(0)]^{-1}\varphi_{ll}(\mathbf{0}) = \Delta^{-1}\varphi_{ll}(\mathbf{0})$ if $\psi'_l(0) = 0$ since $\mathscr{L}'_\Delta(0) = -\mathbb{E}[\mathscr{T}(\Delta)] = -\Delta$. Combining this with (8.9) and (8.10), we derive

$$\psi''_k(u) = \Delta^{-1}\varphi_l(\mathbf{0})\frac{\varphi_{kk}(u^{(k)})\varphi_l(u^{(k)}) - \varphi_k(u^{(k)})\varphi_{lk}(u^{(k)})}{\varphi_l^2(u^{(k)})}, \quad \mu_l \neq 0 \tag{8.11}$$

$$\psi''_k(u) = \Delta^{-1}\varphi_{ll}(\mathbf{0})\frac{\varphi_{kk}(u^{(k)})\varphi_{ll}(u^{(k)}) - \varphi_k(u^{(k)})\varphi_{llk}(u^{(k)})}{\varphi_{ll}^2(u^{(k)})}, \quad \mu_l = 0. \tag{8.12}$$

Note that in the above derivations we have repeatedly used the assumption (ATI), that turns out to be crucial for the identifiability. The basic idea of the algorithm, we shall develop in Section 8.4, is to estimate $\bar{\nu}_k$ by applying the regularised Fourier inversion formula to an estimate of $\bar{\psi}''_k(u)$. As indicated by formulas (8.11) and (8.12), one could, for example, estimate $\bar{\psi}''_k(u)$, if some estimates for the functions $\varphi_k(u)$, $\varphi_{lk}(u)$ and $\varphi_{llk}(u)$ are available.

8.4 Algorithm

The estimation procedure consists of three steps.

Step 1 First, we are interested in estimating partial derivatives of the function $\varphi_Z(u)$ up to the third order. To this end define

$$\widehat{\varphi}_k(u) := \frac{\mathtt{i}}{n}\sum_{j=1}^{n}\Delta_j Y^k \exp(\mathtt{i}\langle u, \Delta_j Y\rangle), \tag{8.13}$$

$$\widehat{\varphi}_{lk}(u) := -\frac{1}{n}\sum_{j=1}^{n}\Delta_j Y^l \Delta_j Y^k \exp(\mathtt{i}\langle u, \Delta_j Y\rangle), \tag{8.14}$$

$$\widehat{\varphi}_{llk}(u) := -\frac{\mathtt{i}}{n}\sum_{j=1}^{n}(\Delta_j Y^l)^2 \Delta_j Y^k \exp(\mathtt{i}\langle u, \Delta_j Y\rangle). \tag{8.15}$$

with $\Delta_j Y := Y_{\Delta j} - Y_{\Delta(j-1)}$, $j = 1, \ldots, n$.

Step 2 In a second step we estimate the second derivative of the characteristic exponent $\psi_k(u)$. Put

$$\widehat{\psi}_{k,2}(u) := \Delta^{-1}\widehat{\varphi}_l(\mathbf{0})\frac{\widehat{\varphi}_{kk}(u^{(k)})\widehat{\varphi}_l(u^{(k)}) - \widehat{\varphi}_k(u^{(k)})\widehat{\varphi}_{lk}(u^{(k)})}{[\widehat{\varphi}_l(u^{(k)})]^2}, \quad |\widehat{\varphi}_l(\mathbf{0})| > \kappa/\sqrt{n},$$

$$\widehat{\psi}_{k,2}(u) := \Delta^{-1}\widehat{\varphi}_{ll}(\mathbf{0})\frac{\widehat{\varphi}_{kk}(u^{(k)})\widehat{\varphi}_{ll}(u^{(k)}) - \widehat{\varphi}_k(u^{(k)})\widehat{\varphi}_{llk}(u^{(k)})}{[\widehat{\varphi}_{ll}(u^{(k)})]^2}, \quad |\widehat{\varphi}_l(\mathbf{0})| \leq \kappa/\sqrt{n},$$

where κ is a positive number.

Step 3 Finally, we construct an estimator for $\bar{\nu}_k(x)$ by applying the Fourier inversion formula combined with a regularization to $\widehat{\psi}_{k,2}(u)$:

$$\widehat{\nu}_k(x) := -\frac{1}{2\pi}\int_{\mathbb{R}} e^{-\mathrm{i}ux}[\widehat{\psi}_{k,2}(u)+\sigma_k^2]w_\nu(uh_n)\,du, \tag{8.16}$$

where w_ν is a regularizing kernel supported on $[-1,1]$ and h_n is a sequence of bandwidths which tends to 0 as $n\to\infty$. The choice of the sequence h_n will be discussed below.

Remark 8.3. The parameter κ determines the testing error for the hypothesis $H:\mu_l>0$. Indeed, if $\mu_l=0$, then $\varphi_l(\mathbf{0})=0$ and by the central limit theorem

$$\mathbb{P}\left(|\widehat{\varphi}_l(\mathbf{0})|>\kappa/\sqrt{n}\right)=\mathbb{P}\left(\sqrt{n}|\widehat{\varphi}_l(\mathbf{0})-\varphi_l(\mathbf{0})|>\kappa\right)\to\mathbb{P}\left(|\xi|>\kappa/\sqrt{\mathrm{Var}[Y_\Delta^{(l)}]}\right),$$

for $n\to\infty$ where $\xi\sim\mathscr{N}(0,1)$.

8.5 Asymptotic analysis

In this section we are going to study the asymptotic properties of the estimates $\widehat{\nu}_k(x)$, $k=1,\ldots,d$. In particular, we consider almost sure uniform as well as pointwise convergence rates for $\widehat{\nu}_k(x)$.

8.5.1 Global vs. local smoothness of Lévy densities

Let L_t be a one-dimensional Lévy process with a Lévy density ν. Denote $\bar{\nu}(x):=x^2\nu(x)$ and assume that $\int\bar{\nu}(x)\,dx<\infty$. For $\beta\geqslant 0$ and $\gamma\in[0,2]$ consider the two following classes of Lévy densities ν:

$$\mathfrak{S}_\beta := \left\{\nu : \int_{\mathbb{R}}(1+|u|^\beta)|\mathscr{F}[\bar{\nu}](u)|\,du<\infty\right\} \tag{8.17}$$

and

$$\mathfrak{B}_\gamma := \left\{\nu : \int_{|y|>\varepsilon}\nu(y)\,dy \asymp \frac{\Pi(\varepsilon)}{\varepsilon^\gamma},\ \varepsilon\to+0\right\}, \tag{8.18}$$

where Π is some positive function on $\mathbb{R}_+$ satisfying $0<\Pi(0+)<\infty$. The parameter β measures the smoothness of $\bar{\nu}$ while γ is the Blumenthal-Geetor index of L_t.

Let us now investigate the connection between classes $\mathfrak{S}_\beta$ and $\mathfrak{B}_\gamma$. First, consider an example. Let L_t be a tempered stable Lévy process with a Lévy density

$$\nu(x)=\frac{2^{\gamma}\bullet\gamma}{\Gamma(1-\gamma)}x^{-(\gamma+1)}\exp\left(-\frac{x}{2}\right)\mathbf{1}_{(0,\infty)}(x),$$

where $\gamma\in(0,1)$. It is clear that $\nu\in\mathfrak{B}_\gamma$, but what is about $\mathfrak{S}_\beta$? Since

$$\bar{\nu}(x)=\frac{2^{\gamma}\bullet\gamma}{\Gamma(1-\gamma)}x^{1-\gamma}\exp\left(-\frac{x}{2}\right)\mathbf{1}_{(0,\infty)}(x),$$

we derive

$$\mathscr{F}[\bar{\nu}](u)=\int_0^\infty e^{\mathrm{i}ux}\bar{\nu}(x)\,dx\asymp 2^{\gamma}\gamma(1-\gamma)e^{\mathrm{i}\pi(1-\gamma/2)}u^{-2+\gamma},\quad u\to+\infty$$

by the Erdélyi lemma (see [28]). Hence, ν cannot belong to $\mathfrak{S}_\beta$ as long as $\beta>1-\gamma$. The message of this example is that given the activity index γ, the parameter β determining the smoothness of $\bar{\nu}$, cannot be taken arbitrary large. The above example can be straight-forwardly generalized to a class of Lévy densities supported on $\mathbb{R}_+$. It turns out that if the Lévy density ν is supported on $[0,\infty)$, is infinitely smooth in $(0,\infty)$ and $\nu\in\mathfrak{B}_\gamma$ for some $\gamma\in(0,1)$, then $\nu\in\mathfrak{S}_\beta$ for all β satisfying $0\le\beta<1-\gamma$ and $\nu\notin\mathfrak{S}_\beta$ for $\beta>1-\gamma$. As a matter of fact, in the case $\gamma=0$ (finite activity case) the situation is different and β can be arbitrary large.

The above discussion indicates that in the case $\nu\in\mathfrak{B}_\gamma$ with some $\gamma>0$ it is reasonable to look at the local smoothness of the transformed Lévy density $\bar{\nu}_k$ instead of the global one. To this end fix a point $x_0\in\mathbb{R}$ and a positive integer number $s\ge 1$. Consider a class $\mathfrak{H}_s(x_0,\delta)$ of Lévy densities ν such that $\bar{\nu}(x)\in C^s(]x_0-\delta,x_0+\delta[)$ and

$$\sup_{x\in]x_0-\delta,x_0+\delta[}|\bar{\nu}^{(l)}(x)|\le L \tag{8.19}$$

for $1\le l\le s$ and some constant $L>0$.

8.5.2 Assumptions

In order to prove the convergence of $\widehat{\nu}_k(x)$ we need the assumptions listed below.

Assumption 8.4

(AL1) The Lévy densities $\nu_1,\ldots,\nu_d$ are in the class $\mathfrak{B}_\gamma$ for some $\gamma>0$.

(AL2) For some $p>2$, the Lévy densities $\nu_k, k=1,\ldots,d$, have finite absolute moments of the order p:

$$\int_{\mathbb{R}}|x|^p\nu_k(x)\,dx<\infty,\quad k=1,\ldots,d.$$

(AT1) The sequence $T_k = \mathscr{T}(\Delta k) - \mathscr{T}(\Delta(k-1))$, $k \in \mathbb{N}$, *is strictly stationary,* α*-mixing with the mixing coefficients* $(\alpha_T(j))_{j\in\mathbb{N}}$ *satisfying*

$$\alpha_T(j) \le \bar{\alpha}_0 \exp(-\bar{\alpha}_1 j), \quad j \in \mathbb{N},$$

for some positive constants $\bar{\alpha}_0$ *and* $\bar{\alpha}_1$. *Moreover, assume that*

$$\mathbb{E}\left[\mathscr{T}^{-2/\gamma}(\Delta)\right] < \infty, \quad \mathbb{E}\left[\mathscr{T}^{2p}(\Delta)\right] < \infty$$

with γ *and* p *being from the assumptions (AL1) and (AL2), respectively.*

(AT2) The Laplace transform $\mathscr{L}_t(z)$ *of* $\mathscr{T}(t)$ *fulfills*

$$\mathscr{L}_t'(z) = o(1), \quad \mathscr{L}_t''(z)/\mathscr{L}_t'(z) = O(1), \quad |z| \to \infty, \quad \operatorname{Re} z > 0.$$

(AK) The regularizing kernel w_ν *is uniformly bounded, is supported on* $[-1,1]$ *and satisfies*

$$w_\nu(u) = 1, \quad u \in [-a_K, a_K]$$

with some $0 < a_K < 1$.

(AH) The sequence of bandwidths h_n *is assumed to satisfy*

$$h_n^{-1} = O(n^{1-\delta}), \quad M_n \sqrt{\frac{\log n}{n}} \sqrt{\frac{1}{h_n} \log \frac{1}{h_n}} = o(1), \quad n \to \infty$$

for some positive number δ *fulfilling* $2/p < \delta \le 1$, *where*

$$M_n = \max_{l \neq k} \sup_{\{|u| \le 1/h_n\}} |\varphi_l^{-1}(u^{(k)})|.$$

Remark 8.5. By requiring $\nu_k \in \mathfrak{B}_\gamma$, $k = 1, \ldots, d$, with some $\gamma > 0$, we exclude from our analysis pure compound Poisson processes and some infinite activity Lévy processes with $\gamma = 0$. This is mainly done for the sake of brevity: we would like to avoid additional technical calculations related to the fact that the distribution of Y_t is not in general absolutely continuous in the case of a compound Poisson process L_t.

Remark 8.6. Assumption (AT1) is satisfied if, for example, the process $\mathscr{T}(t)$ is of the form (2.11), where the rate process $\rho(u)$ is strictly stationary, geometrically α-mixing and fulfills

$$\mathbb{E}[\rho^{2p}(u)] < \infty, \quad u \in [0,\Delta], \quad \mathbb{E}\left(\int_0^\Delta \rho(u)\,du\right)^{-2/\gamma} < \infty. \tag{8.20}$$

In the case of the square-root process (Cox-Ingersoll-Ross process) ρ, the assumptions (8.20) are satisfied for any $p > 0$ and any $\gamma > 0$.

8.5.3 Uniform rates of convergence

Fix some k from the set $\{1,2,\ldots,d\}$. Define a weighting function $w(x) := \log^{-1/2}(e+|x|)$ and denote

$$\|\bar{\nu}_k - \widehat{\nu}_k\|_{L^\infty(\mathbb{R},w)} := \sup_{x\in\mathbb{R}}[w(|x|)|\bar{\nu}_k(x) - \widehat{\nu}_k(x)|].$$

Let ξ_n be a sequence of positive r.v. and q_n be a sequence of positive real numbers. We shall write $\xi_n = O_{a.s.}(q_n)$ if there is a constant $D>0$ such that $\mathbb{P}(\limsup_{n\to\infty} q_n^{-1}\xi_n \le D) = 1$. In the case $\mathbb{P}(\limsup_{n\to\infty} q_n^{-1}\xi_n = 0) = 1$ we shall write $\xi_n = o_{a.s.}(q_n)$.

Theorem 8.7. *Suppose that the assumptions (AL1), (AL2), (AT1), (AT2), (AK) and (AH) are fulfilled. Let $\widehat{\nu}_k(x)$ be the estimate for $\bar{\nu}_k(x)$ defined in Section 8.4. If $\nu_k \in \mathfrak{S}_\beta$ for some $\beta > 0$, then*

$$\|\bar{\nu}_k - \widehat{\nu}_k\|_{L^\infty(\mathbb{R},w)} = O_{a.s.}\left(\sqrt{\frac{\log^{3+\varepsilon} n}{n}\int_{-1/h_n}^{1/h_n} \mathfrak{R}_k^2(u)\,du} + h_n^\beta\right),$$

for arbitrary small $\varepsilon > 0$, where

$$\mathfrak{R}_k(u) = \frac{(1+|\psi_k'(u)|)^2}{|\mathscr{L}_\Delta'(-\psi_k(u))|}.$$

Corollary 8.8. *Suppose that $\sigma_k = 0$, $\gamma \in (0,1]$ in the assumption (AL1) and*

$$|\mathscr{L}_\Delta'(z)| \gtrsim \exp(-a|z|^\eta), \quad |z|\to\infty, \quad \operatorname{Re} z \ge 0$$

for some $a>0$ and $\eta>0$. If $\mu_k > 0$, then

$$\|\bar{\nu}_k - \widehat{\nu}_k\|_{L^\infty(\mathbb{R},w)} = O_{a.s.}\left(\sqrt{\frac{\log^{3+\varepsilon} n}{n}}\exp\left(ach_n^{-\eta}\right) + h_n^\beta\right) \tag{8.21}$$

with some constant $c>0$. In the case $\mu_k = 0$ we have

$$\|\bar{\nu}_k - \widehat{\nu}_k\|_{L^\infty(\mathbb{R},w)} = O_{a.s.}\left(\sqrt{\frac{\log^{3+\varepsilon} n}{n}}\exp\left(ach_n^{-\gamma\eta}\right) + h_n^\beta\right). \tag{8.22}$$

Choosing h_n in such a way that the right-hand sides of (8.21)-(8.22) are minimized, we obtain the rates shown on the right side of Table 1. Table 2 (right) shows the rates for the case $\sigma_k > 0$.

Corollary 8.9. *If $\gamma \in (0,1]$ in the assumption (AL1) and*

$$|\mathscr{L}_\Delta'(z)| \gtrsim |z|^{-\alpha}, \quad |z|\to\infty, \quad \operatorname{Re} z \ge 0$$

$\lvert\mathscr{L}'_\Delta(z)\rvert \gtrsim \lvert z\rvert^{-\alpha}$		$\lvert\mathscr{L}'_\Delta(z)\rvert \gtrsim \exp(-a\lvert z\rvert^{\eta})$	
$\mu_k > 0$	$\mu_k = 0$	$\mu_k > 0$	$\mu_k = 0$
$n^{-\frac{\beta}{(2\alpha+2\beta+1)}} \log^{\frac{(3+\varepsilon)\beta}{(2\alpha+2\beta+1)}}(n)$	$n^{-\frac{\beta}{(2\alpha\gamma+2\beta+1)}} \log^{\frac{(3+\varepsilon)\beta}{(2\alpha\gamma+2\beta+1)}}(n)$	$\log^{-\beta/\eta} n$	$\log^{-\beta/\gamma\eta} n$

Table 1 Theorem 8.7: uniform convergence rates for the estimates $\widehat{\nu}_k$, $k = 1,\ldots,d$, in the case $\sigma_k = 0$.

$\lvert\mathscr{L}'_\Delta(z)\rvert \gtrsim \lvert z\rvert^{-\alpha}$	$\lvert\mathscr{L}'_\Delta(z)\rvert \gtrsim \exp(-a\lvert z\rvert^{\eta})$
$n^{-\frac{\beta}{(4\alpha+2\beta+1)}} \log^{\frac{(3+\varepsilon)\beta}{(4\alpha+2\beta+1)}}(n)$	$\log^{-\beta/2\eta} n$

Table 2 Theorem 8.7: uniform convergence rates for the estimates $\widehat{\nu}_k$, $k = 1,\ldots,d$, in the case $\sigma_k > 0$.

for some $\alpha > 0$, then

$$\|\bar{\nu}_k - \widehat{\nu}_k\|_{L^\infty(\mathbb{R},w)} = O_{a.s.}\left(\sqrt{\frac{\log^{3+\varepsilon} n}{n}} h_n^{-1/2-\alpha} + h_n^{\beta}\right),$$

provided $\mu_k > 0$. In the case $\mu_k = 0$ one has

$$\|\bar{\nu}_k - \widehat{\nu}_k\|_{L^\infty(\mathbb{R},w)} = O_{a.s.}\left(\sqrt{\frac{\log^{3+\varepsilon} n}{n}} h_n^{-1/2-\alpha\gamma} + h_n^{\beta}\right).$$

The choices $h_n = n^{-1/(2(\alpha+\beta)+1)} \log^{(3+\varepsilon)/(2(\alpha+\beta)+1)}(n)$ and

$$h_n = n^{-1/(2(\alpha\gamma+\beta)+1)} \log^{(3+\varepsilon)/(2(\alpha\gamma+\beta)+1)}(n)$$

for the cases $\mu_k > 0$ and $\mu_k = 0$, respectively, lead to the bounds shown in Table 1 on the left side. In the case $\sigma_k > 0$ the rates of convergence are given in Table 2 on the left side.

Remark 8.10. As one can see, the assumption (AH) is always fulfilled for the optimal choices of h_n given in Corollary 8.9, provided $\alpha\gamma + \beta > 0$ and $p > 2 + 1/(\alpha\gamma + \beta)$.

The proof of Theorem 8.7 can be found in [9] and is based on the following representation:

$$\psi_k''(u) - \widehat{\psi}_{k,2}(u) = \frac{\psi_k''(u)}{\psi_l'(0)}\left(\varphi_l(\mathbf{0}) - \widehat{\varphi}_l(\mathbf{0})\right) + \mathscr{R}_0(u) + \mathscr{R}_1(u) + \mathscr{R}_2(u) \quad (8.23)$$

with

$$\begin{aligned}
\mathscr{R}_0(u) = {} & \left[V_1(u)\psi_k''(u) - V_2(u)\psi_k'(u)\right]\left(\varphi_l(u^{(k)}) - \widehat{\varphi}_l(u^{(k)})\right) \\
& + V_2(u)\left(\varphi_k(u^{(k)}) - \widehat{\varphi}_k(u^{(k)})\right) \\
& - V_1(u)\left(\varphi_{kk}(u^{(k)}) - \widehat{\varphi}_{kk}(u^{(k)})\right) \\
& + V_1(u)\psi_k'(u)\left(\varphi_{lk}(u^{(k)}) - \widehat{\varphi}_{lk}(u^{(k)})\right),
\end{aligned}$$

$$\begin{aligned}
\mathscr{R}_1(u) = {} & \left[\widetilde{V}_1(u)\psi_k''(u) - \widetilde{V}_2(u)\psi_k'(u)\right]\left(\varphi_l(u^{(k)}) - \widehat{\varphi}_l(u^{(k)})\right) \\
& + \widetilde{V}_2(u)\left(\varphi_k(u^{(k)}) - \widehat{\varphi}_k(u^{(k)})\right) \\
& - \widetilde{V}_1(u)\left(\varphi_{kk}(u^{(k)}) - \widehat{\varphi}_{kk}(u^{(k)})\right) \\
& + \widetilde{V}_1(u)\psi_k'(u)\left(\varphi_{lk}(u^{(k)}) - \widehat{\varphi}_{lk}(u^{(k)})\right),
\end{aligned}$$

$$\begin{aligned}
\mathscr{R}_2(u) = {} & \Gamma^2(u)\frac{\varphi_l(\mathbf{0})\left(\varphi_{lk}(u^{(k)}) - \widehat{\varphi}_{lk}(u^{(k)})\right)}{\left[\varphi_l(u^{(k)})\right]^2}\left[\left(\varphi_l(u^{(k)}) - \widehat{\varphi}_l(u^{(k)})\right)\psi_k'(u) - \right. \\
& \left. - \left(\varphi_k(u^{(k)}) - \widehat{\varphi}_k(u^{(k)})\right)\right] + \frac{\left(\widehat{\varphi}_l(\mathbf{0}) - \varphi_l(\mathbf{0})\right)}{\varphi_l(u^{(k)})}\left[\frac{\mathscr{R}_0 + \mathscr{R}_1}{\varphi_l(\mathbf{0})}\right]
\end{aligned}$$

with

$$\begin{aligned}
V_1(u) &= \frac{\varphi_l(\mathbf{0})}{\Delta\varphi_l(u^{(k)})} = -\frac{1}{\mathscr{L}_\Delta'(-\psi_k(u))}, \\
V_2(u) &= \frac{\varphi_l(\mathbf{0})\varphi_{lk}(u^{(k)})}{\Delta\left[\varphi_l(u^{(k)})\right]^2} = -V_1(u)\psi_k'(u)\frac{\mathscr{L}_\Delta''(-\psi_k(u))}{\mathscr{L}_\Delta'(-\psi_k(u))},
\end{aligned}$$

$$\widetilde{V}_1(u) = (\Gamma(u) - 1)V_1(u), \quad \widetilde{V}_2(u) = (\Gamma^2(u) - 1)V_2(u)$$

and

$$\Gamma(u) = \left[1 - \frac{1}{\varphi_l(u^{(k)})}\left(\varphi_l(u^{(k)}) - \widehat{\varphi}_l(u^{(k)})\right)\right]^{-1}.$$

The representation (8.23) and the Fourier inversion formula imply the representation for the deviation $\bar{\nu}_k - \widehat{\nu}_k$:

$$
\begin{aligned}
\widehat{\nu}_k(x) - \bar{\nu}_k(x) = {} & \frac{1}{2\pi}\frac{(\varphi_l(\mathbf{0}) - \widehat{\varphi}_l(\mathbf{0}))}{\psi_l'(0)} \int_{\mathbb{R}} e^{-\mathrm{i}ux}\psi_k''(u) w_\nu(uh_n)\,du \\
& + \frac{1}{2\pi}\int_{\mathbb{R}} e^{-\mathrm{i}ux}\mathscr{R}_0(u) w_\nu(uh_n)\,du \\
& + \frac{1}{2\pi}\int_{\mathbb{R}} e^{-\mathrm{i}ux}\mathscr{R}_1(u) w_\nu(uh_n)\,du \\
& + \frac{1}{2\pi}\int_{\mathbb{R}} e^{-\mathrm{i}ux}\mathscr{R}_2(u) w_\nu(uh_n)\,du \\
& + \frac{1}{2\pi}\int_{\mathbb{R}} e^{-\mathrm{i}ux}(1 - w_\nu(uh_n))(\psi_k''(u) + \sigma_k^2)\,du,
\end{aligned}
$$

where each summand can be bounded above in terms of a uniform bound for the deviation between the derivatives of the empirical characteristic function $\widehat{\varphi}$ and the ones of the original characteristic function φ. Such a bound can be obtained in a way similar to Theorem 3.4 (see [9] for details).

8.5.4 Pointwise rates of convergence

Since the transformed Lévy density $\bar{\nu}_k$ is usually not smooth at 0 (see Section 8.5.1), pointwise rates of convergence might be more informative than the uniform ones if $\nu_k \in \mathfrak{B}_\gamma$ for some $\gamma > 0$. It is remarkable that the same estimate $\widehat{\nu}_k$ as before will achieve the optimal pointwise convergence rates in the class $\mathfrak{H}_s(x_0, \delta)$, provided the kernel w_ν satisfies (AK) and is sufficiently smooth.

Theorem 8.11. *Suppose that the assumptions (AL1), (AL2), (AT1), (AT2), (AK) and (AH) are fulfilled. If $\nu_k \in \mathfrak{H}_s(x_0, \delta)$ for some $s \geq 1, \delta > 0$ and $w_\nu \in C^m(\mathbb{R})$ for some $m \geq s$, then*

$$
|\widehat{\nu}_k(x_0) - \bar{\nu}_k(x_0)| = O_{a.s.}\left(\sqrt{\frac{\log^{3+\varepsilon} n}{n}\int_{-1/h_n}^{1/h_n} \mathfrak{R}_k^2(u)\,du} + h_n^s\right) \tag{8.24}
$$

with $\mathfrak{R}_k(u)$ as in Theorem 8.7. As a result, the pointwise rates of convergence for different asymptotic behaviours of the Laplace transform $\mathscr{L}_t$ coincide with those given in Table 1 and Table 2, replacing β by s.

The proof of Theorem 8.11 can be found in [9]

Remark 8.12. If the kernel w_ν is infinitely smooth, then it is suitable for any pointwise smoothness of $\bar{\nu}_k$, i.e., (8.24) will hold for arbitrarily large $s \geq 1$, provided $\nu_k \in \mathfrak{H}_s(x_0, \delta)$. An example of infinitely smooth kernels satisfying (AK) is given by the so called flat-top kernels.

8.6 Simulation study

We consider a model based on time-changed normal inverse Gaussian (NIG) Lévy processes. NIG Lévy processes form a relatively new class of processes introduced in [5] as a model for log returns of stock prices. They are characterised by the property that their increments have an NIG distribution. Barndorff-Nielsen [5] considered classes of normal variance-mean mixtures and defined the NIG distribution as the case when the mixing distribution is inverse Gaussian. Shortly after its introduction it was shown that the NIG distribution fits very well the log returns on German stock market data, making the NIG Lévy processes of great interest for practioneers. A NIG distribution has in general four parameters: $\alpha \in \mathbb{R}_+$, $\beta \in \mathbb{R}$, $\delta \in \mathbb{R}_+$ and $\mu \in \mathbb{R}$ with $|\beta| < \alpha$. The NIG distribution is infinitely divisible with characteristic function

$$\varphi(u) = \exp\left\{\delta\left(\sqrt{\alpha^2-\beta^2} - \sqrt{\alpha^2-(\beta+\mathtt{i}u)^2} + \mathtt{i}\mu u\right)\right\}.$$

Each parameter in the $\mathrm{NIG}(\alpha,\beta,\delta,\mu)$ distribution can be interpreted as having a different effect on the shape of the distribution: α is responsible for the tail heaviness or steepness, β has to do with symmetry, δ scales the distribution and μ determines its mean value. One can define the NIG Lévy process $(L_t)_{t\geq 0}$ which starts at zero and has independent and stationary increments such that each increment $L_{t+\Delta} - L_t$ has a $\mathrm{NIG}(\alpha,\beta,\Delta\delta,\Delta\mu)$ distribution. The NIG process has no diffusion component making it a pure jump process with the Lévy density

$$\nu(x) = \frac{2\alpha\delta}{\pi}\frac{\exp(\beta x)K_1(\alpha|x|)}{|x|} \tag{8.25}$$

where $K_\lambda(z)$ is the modified Bessel function of the third kind. Taking into account the asymptotic relations

$$K_1(z) \asymp 2/z, \quad z \to +0 \text{ and } K_1(z) \asymp \sqrt{\frac{\pi}{2z}}e^{-z}, \quad z \to +\infty,$$

we conclude that $\nu \in \mathfrak{B}_1$ and $\nu \in \mathfrak{H}_s(x_0,\delta)$ for arbitrary large $s > 0$ if $x_0 \neq 0$. Moreover, Assumption (AL2) is fulfilled for any $p > 0$ and the identity

$$\frac{d^2}{du^2}\log\varphi(u) = -\alpha^2/\left(\alpha^2-(\beta+\mathtt{i}u)^2\right)^{3/2}$$

implies $\nu \in \mathfrak{S}_{2-\delta}$ for arbitrary small $\delta > 0$.

One way to construct a time-changed Lévy process from the NIG Lévy process L_t is to use a time change of the form (2.11) with some rate process $\rho(t)$. A possible candidate for the rate of the time change is given by the

Cox-Ingersoll-Ross process (CIR process). The CIR process is defined as a solution of the following SDE:

$$dZ_t = \kappa(\eta - Z_t)\,dt + \zeta\sqrt{Z_t}\,dW_t, \quad Z_0 = 1$$

where W_t is a Wiener process. This process is mean reverting with $\kappa > 0$ being the speed of mean reversion, $\eta > 0$ being the long-run mean rate and $\zeta > 0$ controlling the volatility of Z_t. Additionally, if $2\kappa\eta > \zeta^2$ then Z_t is stationary and exponentially α-mixing. The time change $\mathscr{T}$ is then defined as

$$\mathscr{T}(t) = \int_0^t Z_t\,dt.$$

Simple calculations show that the Laplace transform of $\mathscr{T}(t)$ is given by

$$\mathscr{L}_t(z) = \frac{\exp(\kappa^2\eta t/\zeta^2)\exp(-2z/(\kappa+\gamma(z)\coth(\gamma(z)t/2)))}{(\cosh(\gamma(z)t/2)+\kappa\sinh(\gamma(z)t/2)/\gamma(z))^{2\kappa\eta/\zeta^2}}$$

with $\gamma(z) = \sqrt{\kappa^2 + 2\zeta^2 z}$. It is easy to see that $\mathscr{L}_t(z) \asymp \exp\left(-\frac{\sqrt{2z}}{\zeta}[1+t\kappa\eta]\right)$ as $|z| \to \infty$ with $\mathrm{Re}\,z \geq 0$. Moreover $\mathbb{E}|\mathscr{T}(t)|^p < \infty$ for any $p > 0$ and any fixed $t > 0$ since $\mathscr{L}_t(z)$ is finite for real z satisfying $z > -\kappa^2/2\zeta^2$.

Let L_t be a three-dimensional NIG Lévy process with independent components distributed as $\mathrm{NIG}(1,-0.05,1,-0.5)$, $\mathrm{NIG}(3,-0.05,1,-1)$ and $\mathrm{NIG}(1,-0.03,1,2)$, respectively. Construct the time-changed process $Y_t = L_{\mathscr{T}(t)}$. Note that the process Y_t is not any longer a Lévy process and has in general dependent increments. Let us estimate $\bar{\nu}_1$, the transformed Lévy density of the first component of L_t. First note that according to Theorem 8.7 the estimate $\widehat{\nu}_1$ from Section 8.4 has the following logarithmic convergence rates

$$\|\bar{\nu}_1 - \widehat{\nu}_1\|_{L^\infty(\mathbb{R},w)} = O_{a.s.}\left(\log^{-2(2-\delta)}(n)\right), \quad n \to \infty$$

for arbitrary small $\delta > 0$, provided the bandwidth sequence is chosen in the optimal way. We construct an estimate $\widehat{\nu}_1$ as described before. In particular, we first estimate the derivatives φ_1, φ_2, φ_{11} and φ_{12} by means of (8.13) and (8.14). Then we estimate $\psi_1''(u)$. Finally, we get $\widehat{\nu}_1$ from (8.16) where the kernel w_ν is chosen to be the kernel of the form

$$w_\nu(x) = \begin{cases} 1, & |x| \leq 0.05, \\ \exp\left(-\frac{e^{-1/(|x|-0.05)}}{1-|x|}\right), & 0.05 < |x| < 1, \\ 0, & |x| \geq 1. \end{cases}$$

The kernel w_ν obviously satisfies the assumption (AK). Let us turn to the finite sample performance of the estimate $\widehat{\nu}_1$. It turns out that the choice of the sequence h_n is crucial for a good performance of ν_1. For this choice

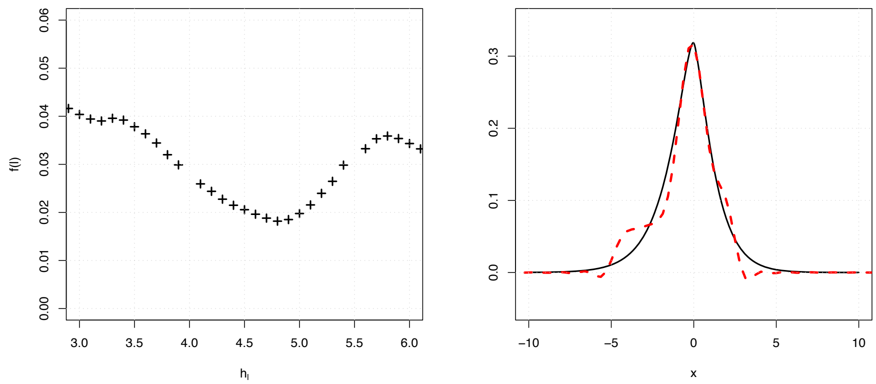

Fig. 5 Left hand-side: objective function $f(l)$ for the “quasi-optimality” approach versus the corresponding bandwidths h_l. Right hand-side: adaptive estimate $\widetilde{\nu}_1$ (dashed line) together with the true function $\bar{\nu}_1$ (solid line).

we adopt the so called “quasi-optimality” approach proposed in [7]. This approach is aimed to perform a model selection in inverse problems without taking into account the noise level. Although one can prove the optimality of this criterion on average only, it leads in many situations to quite reasonable results. In order to implement the “quasi-optimality” algorithm in our situation, we first fix a sequence of bandwidths $h_1,\ldots,h_L$ and construct the estimates $\nu_1^{(1)},\ldots,\nu_1^{(L)}$ using the formula (8.16) with bandwidths $h_1,\ldots,h_L$, respectively. Then one finds $l^\star = \operatorname{argmin}_l f(l)$ with

$$f(l) = \|\widehat{\nu}_1^{(l+1)} - \widehat{\nu}_1^{(l)}\|_{L_1(\mathbb{R})}, \quad l = 1,\ldots,L.$$

Denote by $\widetilde{\nu}_1 = \widehat{\nu}_1^{l^*}$ a new adaptive estimate for $\bar{\nu}_1$. In our implementation of the “quasi-optimality” approach we take $h_l = 0.5 + 0.1 \times l$, $l = 1,\ldots,40$. The parameters of the used CIR process are $\kappa = 1$, $\eta = 1$ and $\zeta = 0.1$. Finite sample performance of $\widehat{\nu}_1$ is illustrated in Figure 5, where the sequence of estimates $\widehat{\nu}_1^{(1)},\ldots,\widehat{\nu}_1^{(L)}$ was constructed from the time series $Y_\Delta,\ldots,Y_{n\Delta}$ with $n = 5000$ and $\Delta = 0.1$. We can compute some functionals of $\widetilde{\nu}_1$. We have, for example, the following estimates for the integral and for the mean of $\bar{\nu}_1$: $\int \widetilde{\nu}_1(x)\,dx = 1.081376$ ($\int \bar{\nu}_1(x)\,dx = 1.015189$) and $\int \tilde{x}\nu_1(x)\,dx = -0.4772505$ ($\int x\bar{\nu}_1(x)\,dx = -0.3057733$).

Figure 6 (left) shows the boxplots of the resulting error $\|\bar{\nu}_1 - \widetilde{\nu}_1\|_{L^\infty(\mathbb{R},w)}$ computed using 100 trajectories each of the length $n = 5000$, where the time span between observations is $\Delta = 0.1$.

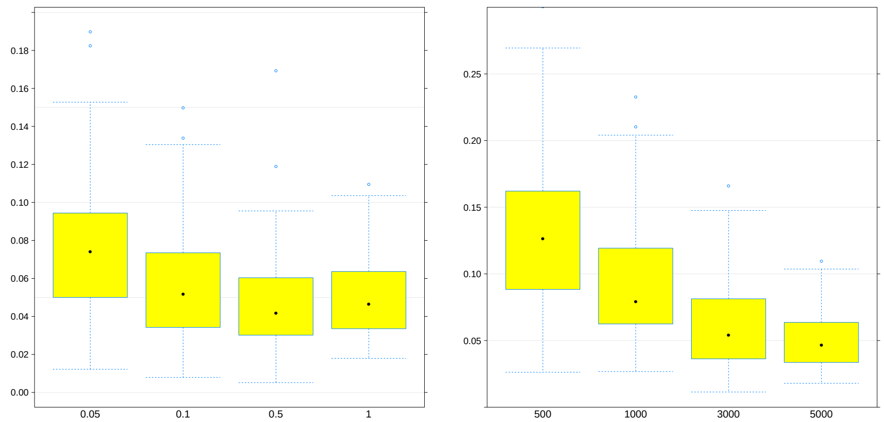

Fig. 6 Boxplots of the error $\|\bar{\nu}_1 - \widetilde{\nu}_1\|_{L^\infty(\mathbb{R},w)}$ for different values of the mean reversion speed parameter κ (left) and different numbers of observations n (right).

9 Spectral calibration from option data

Statistical estimators based on historical data yield parameters of the model under the so-called physical or real-world probability measure $\mathbb{P}$. By contrast, option pricing and calibration refers to expectations relative to some risk-neutral measure $\mathbb{Q}$. Here, we present a calibration method for exponential Lévy models from European option prices. This type of indirect observation is closely related to the direct low-frequency observation setting. The presentation is mainly based on Belomestny and Reiß [10], where also all the proofs can be found.

9.1 The exponential Lévy model and option prices

A European call option with maturity T and strike K for an underlying asset grants the holder the right to buy the asset at the future time T for the price K. A risk neutral price at time $t = 0$ for this option is given by

$$C(K,T) = e^{-rT}\mathbb{E}_{\mathbb{Q}}[(S_T - K)^+], \tag{9.1}$$

where $(A)^+ := \max(A,0)$, S_T is the (random) asset price at time T and $\mathbb{Q}$ is a martingale measure equivalent to the real world probability $\mathbb{P}$. By considering option prices we immediately draw inference on this pricing measure $\mathbb{Q}$. The measure $\mathbb{Q}$ is assumed to be settled by the market and to be identical for all options traded.

From now on we suppose that S follows an exponential Lévy model

$$S_t = Se^{rt+X_t} \text{ with a Lévy process } X_t \text{ for } t \geqslant 0, \tag{9.2}$$

under $\mathbb{Q}$ where $S>0$ is the present value of the asset and $r \geqslant 0$ is the riskless interest rate, which is assumed to be known and constant. Risk neutral pricing requires that the discounted price process $e^{-rt}S_t$ is a martingale on the filtered probability space $(\Omega, \mathfrak{F}, \mathbb{Q}, (\mathfrak{F}_t))$, fixed throughout.

An excellent reference for this and similar Lévy-based models in finance is the monograph by Cont and Tankov [21]. Here we shall mainly consider Lévy processes X with a jump component of finite variation and absolutely continuous jump distribution. Its characteristic function is given by the Lévy-Khintchine representation

$$\varphi_T(u) := \mathbb{E}[\exp(iuX_T)] = \exp\Big(T\Big(-\frac{\sigma^2}{2}u^2 + i\gamma u + \int_{-\infty}^{\infty}(e^{iux}-1)\nu(x)\,\mathrm{d}x\Big)\Big). \tag{9.3}$$

The corresponding characteristic triplet is denoted by $\mathscr{T} := (\sigma^2, \gamma, \nu)$.

By the independence of increments in X the martingale condition may be explicitly stated as

$$\forall t \geqslant 0 : \mathbb{E}[e^{X_t}] = 1 \iff \frac{\sigma^2}{2} + \gamma + \int_{-\infty}^{\infty}(e^x - 1)\nu(x)\,\mathrm{d}x = 0, \tag{9.4}$$

where here and below expectations are always taken with respect to $\mathbb{Q}$. Observe that we have imposed implicitly the exponential moment condition $\int_0^\infty (e^x-1)\nu(x)\,\mathrm{d}x < \infty$ to ensure the existence of $\mathbb{E}[S_t]$. Another consequence is that the characteristic function φ_T is defined on the whole strip $\{z \in \mathbb{C} \mid \mathrm{Im}(z) \in [-1,0]\}$ in the complex plane, which will be important later. We reduce the number of parameters by introducing the negative log-forward moneyness

$$x := \log(K/S) - rT,$$

such that the call price in terms of x is given by

$$\mathscr{C}(x,T) = S\mathbb{E}[(e^{X_T} - e^x)^+].$$

The analogous formula for the price of a put option, which gives the owner the right to sell an asset at time T for the price K, is $\mathscr{P}(x,T) = S\mathbb{E}[(e^x - e^{X_T})^+]$. Then the well-known put-call parity is easily established:

$$\mathscr{C}(x,T) - \mathscr{P}(x,T) = S\mathbb{E}[e^{X_T} - e^x] = S(1-e^x). \tag{9.5}$$

9.2 The observations

We focus on the calibration from options with a fixed maturity $T>0$. We observe the prices of N call options (or by the put-call parity alternatively put options) at different strikes K_j, $j = 1, \dots, N$, corrupted by noise

$$Y_j = C(K_j, T) + \sigma_j \varepsilon_j, \quad j = 1, \dots, N. \tag{9.6}$$

We assume the observational noise (ε_j) to consist of independent centred random variables with $\mathbb{E}[\varepsilon_j^2] = 1$ and $\sup_j \mathbb{E}[\varepsilon_j^4] < \infty$. The noise levels (σ_j) are assumed to be positive and known. This random observation model reflects the bid-ask spread and other frictions at the market.

As we want to employ Fourier techniques, we introduce the function

$$\mathscr{O}(x) := \begin{cases} S^{-1}\mathscr{C}(x,T), & x \geqslant 0, \\ S^{-1}\mathscr{P}(x,T), & x < 0 \end{cases} \tag{9.7}$$

in the spirit of [19]. $\mathscr{O}$ records normalised call prices for $x \geqslant 0$ and normalised put prices for $x \leqslant 0$. The following properties of $\mathscr{O}$ are important.

Proposition 9.1.

(a) We have $\mathscr{O}(x) = S^{-1}\mathscr{C}(x,T) - (1-e^x)^+$ *for all* $x \in \mathbb{R}$.
(b) $\mathscr{O}(x) \in [0, 1 \wedge e^x]$ *holds for all* $x \in \mathbb{R}$.
(c) If $C_\alpha := \mathbb{E}[e^{\alpha X_T}]$ *is finite for some* $\alpha \geqslant 1$, *then* $\mathscr{O}(x) \leqslant C_\alpha e^{(1-\alpha)x}$ *holds for all* $x \geqslant 0$.
(d) At any $x \in \mathbb{R} \setminus \{0\}$, *respectively* $x \in \mathbb{R} \setminus \{0, \gamma T\}$ *in the case* $\sigma = 0$ *and* $\lambda < \infty$, *the function* $\mathscr{O}$ *is twice differentiable with*

$$\int_{\mathbb{R}\setminus\{0,\gamma T\}} |\mathscr{O}''(x)|\, dx \leqslant 3.$$

The first derivative $\mathscr{O}'$ *has a jump of height* -1 *at zero and, in the case* $\sigma = 0$ *and* $\lambda < \infty$, *a jump of height* $+e^{T(\gamma-\lambda)}$ *occurs in* $\mathscr{O}'$ *at* γT.
(e) The Fourier transform of $\mathscr{O}$ *satisfies*

$$\mathscr{F}\mathscr{O}(v) = \frac{1 - \varphi_T(v-i)}{v(v-i)}, \quad v \in \mathbb{R}. \tag{9.8}$$

This identity extends to all complex values v *with* $\operatorname{Im}(v) \in [0,1]$. *Note the properties* $\varphi_T(0) = 1$ *and* $\varphi_T(-i) = 1$ *derived from the general property of characteristic functions and the martingale condition* (9.4), *respectively.*

We transform our observations (Y_j) and predictors (K_j) to

$$O_j := Y_j/S - (1 - K_j e^{-rT}/S)^+ = \mathscr{O}(x_j) + \delta_j \varepsilon_j, \tag{9.9}$$

$$x_j := \log(K_j/S) - rT, \tag{9.10}$$

where $\delta_j = S^{-1}\sigma_j$. In practice, the design (x_j) will be rather dense around $x = 0$ (at the money) and sparse for options further out of the money or in the money.

In order to facilitate the subsequent analysis we make a mild moment assumption on the price process, which guarantees by Proposition 9.1(b,c) the exponential decay of $\mathscr{O}$.

Assumption 9.2 *We assume that* $C_2 := \mathbb{E}[e^{2X_T}]$ *is finite. This is equivalent to postulating for the asset price a finite second moment:* $\mathbb{E}[S_T^2] < \infty$.

9.3 The estimation method

Let us assume here that the Lévy process has finite intensity λ. Later we shall impose also a certain regularity on the jump density ν. We make use of the exact inversion formula, that is the mapping from the option prices to the parameters derived in equation (9.11) below. This has the advantage that no numerical minimisation technique needs to be employed and the propagation of errors is more transparent. Moreover, the method and the proofs are closely related to the basic spectral estimation procedure from Section 4.

Since our asset follows an exponential Lévy model, the jumps in the Lévy process appear exponentially transformed in the asset prices and it is intuitive that inference on the exponentially weighted jump measure

$$\mu(x) := e^x \nu(x), \quad x \in \mathbb{R},$$

will lead to spatially more homogeneous properties of the estimator than for ν itself. Our calibration procedure relies essentially upon the formula

$$\begin{aligned} \psi(v) &:= \frac{1}{T}\log\Big(1 + iv(1+iv)\mathscr{F}\mathscr{O}(v)\Big) = \frac{1}{T}\log(\varphi_T(v-i)) \\ &= -\frac{\sigma^2 v^2}{2} + i(\sigma^2+\gamma)v + (\sigma^2/2 + \gamma - \lambda) + \mathscr{F}\mu(v), \end{aligned} \tag{9.11}$$

which is a simple consequence of the formulae (9.3) and (9.8). Note that the function ψ is up to a shift in the argument the cumulant-generating function of the Lévy process and a continuous version of the logarithm must be taken such that $\psi(0) = 0$, which is implied by the martingale condition. Formula (9.11) shows that the Lévy triplet is uniquely identifiable given the observation of the whole option price function $\mathscr{O}$ without noise: $\mathscr{F}\mu(v)$ tends to zero as $|v| \to \infty$ due to the Riemann-Lebesgue Lemma and σ^2, γ, λ are identifiable as coefficients in the polynomial, which in turn yields the function $\mathscr{F}\mu(v)$. A properly refined application of this approach will equip us with estimators for the whole triplet $\mathscr{T} = (\sigma^2, \gamma, \mu)$ (we parametrize Lévy triplets equivalently with μ or ν).

Let us formulate the basic algorithm to be used when a certain smoothness property is imposed on μ, that is under the prior knowledge $\mu \in \mathscr{G}$, where $\mathscr{G}$ is a smoothness class. The procedure consists of four steps: (a) we build an approximation $\tilde{\mathscr{O}}$ of $\mathscr{O}$ from the data; (b) we obtain an approximation $\tilde{\psi}$ of ψ by formula (9.11); (c) we estimate the coefficients of the quadratic polynomial on the right-hand side in (9.11) from $\tilde{\psi}$ under the presence of a noise component and the nonparametric nuisance part $\mathscr{F}\mu$; (d) we obtain an estimator for $\mathscr{F}\mu$ by considering the remainder.

The model (9.11) has a similar structure as partial linear models, well known in statistics, but in fact there is one substantial difference: the function $\mathscr{F}\mu$ is not supposed to be smooth, but instead it is decaying for high frequencies because we work in the spectral domain. This is also why we shall regularise the problem by cutting off frequencies $|v|$ higher than a certain threshold level U, which depends on the noise level and the smoothness assumptions in $\mathscr{G}$.

We now give a detailed description of the different steps in the procedure.

(a) We approximate the function $\mathscr{O}$ by building $\tilde{\mathscr{O}}$ from the observations (O_j) in the form

$$\tilde{\mathscr{O}}(x) = \beta_0(x) + \sum_{j=1}^{N} O_j b_j(x), \quad x \in \mathbb{R},$$

and consequently $\mathscr{F}\mathscr{O}$ by

$$\mathscr{F}\tilde{\mathscr{O}}(u) = \mathscr{F}\beta_0(u) + \sum_{j=1}^{N} O_j \mathscr{F} b_j(u), \quad u \in \mathbb{R},$$

where (b_j) are some basis functions to be chosen and the function β_0 is added to take care of the jump in the derivative of $\mathscr{O}$ at zero: $\beta_0'(0+) - \beta_0'(0-) = -1$. Taking into account the decay properties of $\mathscr{O}$, we interpolate the data by specifying

$$\forall x \in \mathbb{R}: \ b_k(x) \in [0,1], \quad \forall j,k = 1,\ldots,N: \ b_k(x_j) = \delta_{jk}, \ \lim_{|u|\to\infty} b_k(u) = 0.$$

We stress here that step (a) should not be understood as a smoothing step, but rather as a means to find a reasonable approximation of $\mathscr{F}\mathscr{O}$ based on discrete data. As can be seen in the theoretical analysis and the numerical simulations below, it suffices to use simple linear B-splines as basis functions. A B-spline consists of polynomial pieces, connected in a special way. For example, a linear B-spline consists of 2 polynomial peaces that joint at one inner knot in such a way that at the joining point the function is continuous. Moreover, any linear B-spline is positive on a domain spanned by 3 knots; everywhere else it is zero.

(b) For $\kappa(v) \in (0,1)$, specified later, we calculate

$$\tilde{\psi}(v) := \frac{1}{T} \log_{\geqslant \kappa(v)} \Big(1 + iv(1+iv)\mathscr{F}\tilde{\mathscr{O}}(v) \Big), \quad v \in \mathbb{R}, \tag{9.12}$$

where the function $\log_{\geqslant\kappa} : \mathbb{C} \setminus \{0\} \to \mathbb{C}$ is given by

$$\log_{\geqslant\kappa}(z) := \begin{cases} \log(z), & |z| \geq \kappa \\ \log(\kappa z/|z|), & |z| < \kappa \end{cases} \tag{9.13}$$

and $\log(\bullet)$ is taken in such a way that $\tilde{\psi}(v)$ is continuous with $\tilde{\psi}(0) = 0$ (almost surely the argument of the logarithm in (9.12) does not vanish because zero is a polar set of the process, see [53] for the fine result). If we observe option prices for different maturities T_k, we perform the steps (a) and (b) for each T_k separately and aggregate at this point the different estimators for ψ to obtain one estimator with less variance.

(c) With an estimate $\tilde{\psi}$ of ψ at hand, we obtain estimators for the parametric part $(\sigma^2, \gamma, \lambda)$ by an averaging procedure taking into account the polynomial structure in (9.11). Upon fixing the spectral cut-off value $U = U(\mathscr{G}, (\delta_j), (x_j))$, we set

$$\hat{\sigma}^2 := \int_{-U}^{U} \mathrm{Re}(\tilde{\psi}(u)) w_\sigma^U(u)\,\mathrm{d}u, \tag{9.14}$$

$$\hat{\gamma} := -\hat{\sigma}^2 + \int_{-U}^{U} \mathrm{Im}(\tilde{\psi}(u)) w_\gamma^U(u)\,\mathrm{d}u, \tag{9.15}$$

$$\hat{\lambda} := \frac{\hat{\sigma}^2}{2} + \hat{\gamma} - \int_{-U}^{U} \mathrm{Re}(\tilde{\psi}(u)) w_\lambda^U(u)\,\mathrm{d}u, \tag{9.16}$$

where the weight functions w_σ^U, w_γ^U and w_λ^U satisfy

$$\int_{-U}^{U} w_\sigma^U(u)\,\mathrm{d}u = 0,\ \int_{-U}^{U} u^2 w_\sigma^U(u)\,\mathrm{d}u = -2; \quad \int_{-U}^{U} u w_\gamma^U(u)\,\mathrm{d}u = 1; \tag{9.17}$$

$$\int_{-U}^{U} u^2 w_\lambda^U(u)\,\mathrm{d}u = 0,\ \int_{-U}^{U} w_\lambda^U(u)\,\mathrm{d}u = 1. \tag{9.18}$$

For the construction of weighting functions satisfying the above conditions see Section 4. The estimate of the coefficients can be understood as an orthogonal projection estimate with respect to an L^2-scalar product weighted according to the supposed decay property of $\mathscr{F}\mu$.

(d) Finally, we define the estimate for μ as the inverse Fourier transform of the remainder:

$$\hat{\mu}(u) := \mathscr{F}^{-1}\left[\left(\tilde{\psi}(\bullet) + \tfrac{\hat{\sigma}^2}{2}(\bullet - i)^2 - i\hat{\gamma}(\bullet - i) + \hat{\lambda}\right)\mathbf{1}_{[-U,U]}(\bullet)\right](u), \quad u \in \mathbb{R}. \tag{9.19}$$

Note that the computational complexity of this basic estimation procedure is very low. The only time consuming steps are the three integrations in step (c) and the inverse Fourier transform (inverse FFT) in step (d). In step (a) we just take a data-dependent linear combination of the functions $\mathscr{F}b_k$ and the function $\mathscr{F}\beta_0$, which with our choice as linear B-splines can be computed explicitly:

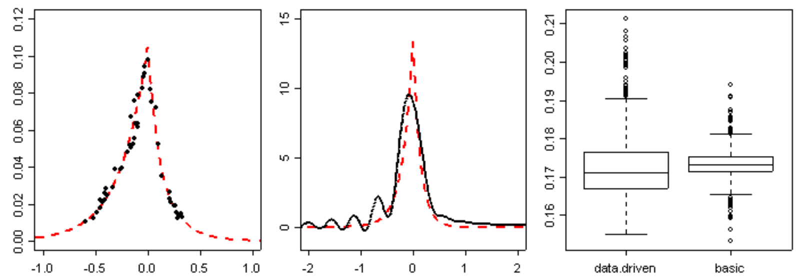

Fig. 7 Kou model. Left: Sample (O_j) and true function $\mathscr{O}$ (dashed line). Center: True μ (dashed) and estimated $\hat{\mu}$ (black) modified Lévy densities. Right: Box plot for the L^2-loss of the data-driven and the basic procedure based on 1000 Monte-Carlo simulations.

$$\mathscr{F}b_k(u) := u^{-2}\Big(\frac{e^{iux_k}-e^{iux_{k-1}}}{x_k-x_{k-1}} - \frac{e^{iux_{k+1}}-e^{iux_k}}{x_{k+1}-x_k}\Big),$$

$$\mathscr{F}\beta_0(u) \;=\; u^{-2}\Big(1+\frac{e^{iux_{j_0}}x_{j_0-1}-e^{iux_{j_0-1}}x_{j_0}}{x_{j_0}-x_{j_0-1}}\Big)$$

with $k = 1,\dots,N$, some extrapolated design points x_0 and x_{N+1}, where we set $\tilde{O}(x_0) = \tilde{O}(x_{N+1}) = 0$, and with the index $j_0 \in \{1,\dots,N\}$ defined by $x_{j_0-1} < 0 \leqslant x_{j_0}$.

9.4 A numerical example

Two empirical findings (*stylized facts*) in financial data have attracted attention recently: the leptokurtic return distribution of assets with a higher peak and two (asymmetric) heavier tails than those of the normal distribution, and the implied volatility smile. To incorporate these features, the double exponential jump diffusion model was proposed by Kou [45]. In his model the Lévy triplet is specified by the jump density

$$\nu(x) = \lambda\left(p\lambda_+ e^{-\lambda_+ x}\mathbf{1}_{[0,\infty)}(x) + (1-p)\lambda_- e^{\lambda_- x}\mathbf{1}_{(-\infty,0)}(x)\right), \quad x \in \mathbb{R},$$

and the parameters $\sigma, \lambda, \lambda_+, \lambda_- \geqslant 0$ and $p \in [0,1]$, while γ is uniquely determined by the martingale condition. We simulate the Kou model with parameters $\sigma = 0.1, \lambda = 5, \lambda_- = 4, \lambda_+ = 8, p = 1/3$ and apply the nonparametric estimation procedure given the observation of noisy European option data with $T = 0.25$, $N = 50$, $r = 0.06$ and $\delta_j = \mathscr{O}(x_j)/10$.

In Figure 7 (left) the simulated observations (O_j) and the true curve $\mathscr{O}$ are depicted as functions of the log-forward moneyness. The estimated transformed Lévy density μ in the center is obtained using the basic procedure, as specified in the mathematical analysis, with a human-driven choice of the cut-off parameter U. The parameters were estimated as $\hat{\sigma} = 0.035$, $\hat{\lambda} = 7.56$, $\hat{\gamma} = 0.556\,(\gamma = 0.423)$. We observe that the estimated transformed Lévy density recovers the main features of the Kou model like the mode at zero and the skewness. From the functional form of the estimator we can easily derive estimates for other important quantities, e.g. for the proportion of negative jumps by calculating $\hat{\lambda}^{-1}\int_{-\infty}^{0}\hat{\nu}(x)\,\mathrm{d}x = \hat{\lambda}^{-1}\int_{-\infty}^{0}e^{-x}\hat{\mu}(x)\,\mathrm{d}x$, which in the simulation example evaluates to 0.72 (true value: $1-p = 2/3$).

In the right part of Figure 7 we compare the performance of the completely data-driven estimator, as described in Bauer and Reiß [7], with the oracle estimator (i.e, choosing the best possible U) obtained from the basic procedure in terms of the empirical L^2-loss. A box plot is shown for 1000 Monte-Carlo replications.

9.5 Real data: DAX options

This part is mainly based on the work by Söhl and Trabs [55]. The calibration methods are applied to a data set from the Deutsche Börse database Eurex[1]. It consists of settlement prices of European put and call options on the DAX index from May 2008. Therefore, the prices are observed before the latest financial crises and thus the market activity is relatively stable. The interest rate r is chosen for each maturity separately according to the put–call parity at the respective strike prices. The expiry months of the options are between July and December, 2008, and thus the time to maturity T, measured in years, ranges from two to seven months. The number of observations N is between 50 to 100 different strikes for each maturity and trading day.

In addition to applying the calibration method for Lévy processes with finite jump activity, as described above, we shall also report the estimation results for a pure-jump exponential Lévy model of self-decomposable type where the Lévy measure has a density

$$\nu(dx) = \frac{k(x)}{|x|}\,dx \text{ with } k:\mathbb{R}\to\mathbb{R}^+ \text{ increasing on } (-\infty,0)\text{, decreasing on } (0,\infty).$$

The class of self-decomposable distributions has nice probabilistic characterisations, e.g. as invariant measures of all Lévy-Ornstein-Uhlenbeck process, see Sato [52], and they include infinite activity jump process of small intensity like the important class of Gamma processes. The main parameter that measures the (usually infinite) small jump intensity is

$$\alpha := k(0-) + k(0+) \in [0,\infty).$$

[1] provided through the SFB 649 “Economic Risk”, Humboldt-Universität zu Berlin

		T=0.136, N=61	T=0.564, N=106
FA	$\hat{\sigma}$	0.110 (0.0021)	0.124 (0.0013)
	$\hat{\lambda}$	3.392 (0.2015)	0.637 (0.0181)
SD	$\hat{\alpha}$	8.662 (0.1534)	5.181 (1.0030)

Table 3 Estimated parameters ϑ and estimated standard deviation $\hat{s}_\vartheta$ (in brackets) for $\vartheta \in \{\sigma, \lambda, \alpha\}$ using option prices from May 29, 2008, with N observed strikes for each maturity T.

For statistical estimation of the function k or the parameter α a spectral calibration method works, if adapted in a clever way, similar to the one presented here, see Trabs [59] for the details. In view of the discussions about the right model world for financial data (ranging from continuous semi-martingales to pure-jump processes) it is very reasonable to check the model validity by estimators for these two structurally different Lévy classes, but see also the discussion in Section 6 for the behaviour of the finite activity estimator under infinite jump activity models.

In addition to including a second estimator, estimated standard deviations are presented, which are, of course, of major interest as a quantification of the statistical uncertainty. Here, we follow the construction by Söhl [54], which provides asymptotically honest joint confidence intervals for the real triplet parameters and the Lévy density evaluated at prescribed points and reveals very interesting interdependencies. The numerical results are described in full detail in Söhl and Trabs [55].

9.5.1 Finite activity versus self-decomposable Lévy models

Let us first focus on option prices of May 29, 2008, an arbitrarily chosen day, where options are calibrated to both, the finite activity (FA) and the self-decomposable (SD) exponential Lévy model. The results are summarized in Table 3 and Figure 8. Using the complete estimation of the models, we generate the corresponding option functions $\hat{\mathscr{O}}$. They are graphically compared to the given data points. Both methods yield good fits to the data. For the longer maturity, however, some problems occur in the SD calibration. Although the sample size is larger, the estimated standard deviation is larger for longer maturities in the SD scenario, too. The calibration at other trading days confirms this weakness of the SD method for larger T. This coincides with the asymptotic analysis of Trabs [59] where longer durations lead to slower convergence rates of the risk.

Moreover, Figure 8 shows that the estimated option function $\hat{\mathscr{O}}$ which results from the SD calibration does not exactly recover the tails of $\mathscr{O}$. At all maturities and in both models the Lévy density has more weight on the negative half line and thus there are more negative jumps than positive ones

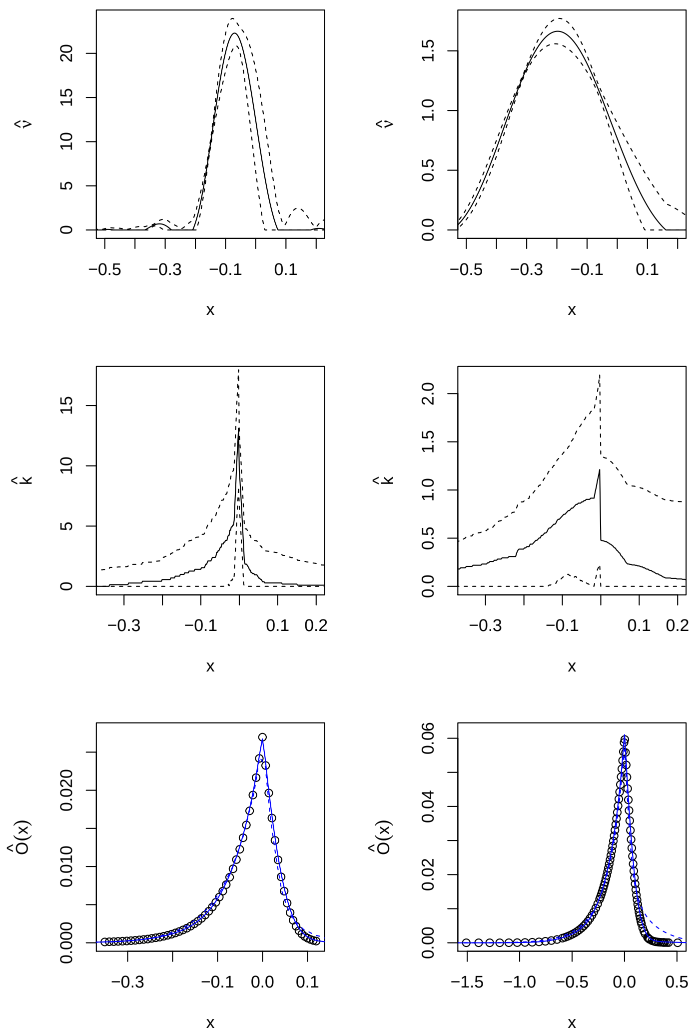

Fig. 8 Estimated jump densities *(top)*, k–functions *(center)* with pointwise 95% confidence intervals as well as calibrated option functions in the FA *(bottom, solid)* and SD *(bottom, dashed)* setting and given DAX option data from May 29, 2008 *(bottom, points)*. The times to maturity are $T = 0.136$ years *(left)* and $T = 0.564$ years *(right)*.

priced into the options. This coincides with the empirical findings in the literature, see e.g. Cont and Tankov [22].

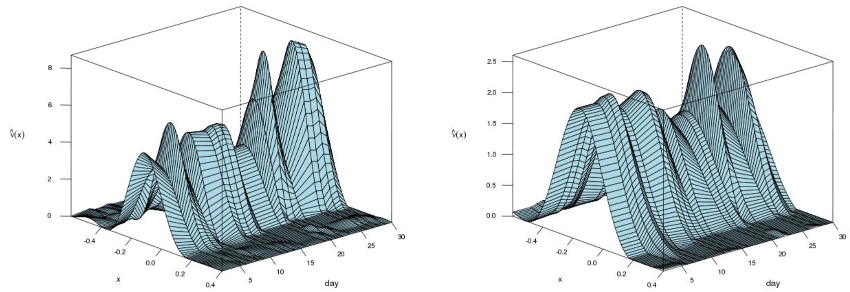

Fig. 9 Estimation of ν for maturity in September *(left)* and December (*right)*.

9.5.2 Estimation across trading days

By considering more than one day the stability of the finite activity estimation procedure is studied. Moreover, calibrating the model across the trading days in May, 2008, shows the development of the model along the time line and with small changes in the maturities. To profit from the higher observation number, the calibration procedure is applied to options with maturity in September and December.

The estimated volatility $\hat{\sigma}$ fluctuates around 0.1 and 0.12. Figure 9 displays the estimated jump densities. All jump measures have a similar shape. In contrast to Cont and Tankov [22] the densities are unimodal or have only minor additional modes in the tails, which may be artifacts of the spectral calibration method. The tails of $\hat{\nu}$ do not differ significantly, while the different heights reflect the development of the jump activities $\hat{\lambda}$. There is an obvious trend to small negative jumps in all data sets, which is in line with the stylized facts of option pricing models. The calibration is stable for consecutive market days.

9.6 Risk bounds

9.6.1 The main results

In order to assess the quality of the estimators introduced in Section 9.3, we quantify their risks under a smoothness condition of order s on the transformed jump density μ.

Definition 9.3. *For $s \in \mathbb{N}$ and $R, \sigma_{max} > 0$ let $\mathscr{G}_s(R, \sigma_{max})$ denote the set of all Lévy triplets $\mathscr{T} = (\sigma^2, \gamma, \mu)$, satisfying the martingale condition and Assumption 9.2 with $C_2 \leqslant R$, such that μ is s-times (weakly) differentiable and*

	σ^2	γ	λ	μ
$\sigma_{max} > 0$	$\lvert\log(\varepsilon)\rvert^{-(s+3)/2}$	$\lvert\log(\varepsilon)\rvert^{-(s+2)/2}$	$\lvert\log(\varepsilon)\rvert^{-(s+1)/2}$	$\lvert\log(\varepsilon)\rvert^{-s/2}$
$\sigma_{max} = 0$	0	$\varepsilon^{(2s+4)/(2s+5)}$	$\varepsilon^{(2s+2)/(2s+5)}$	$\varepsilon^{2s/(2s+5)}$

Table 4 The minimax rates $v_{q,\sigma_{max}}$ for the different parameters $q \in \{\sigma^2, \gamma, \lambda, \mu\}$.

$$\sigma \in [0, \sigma_{max}], \quad |\gamma|, \lambda \in [0, R], \quad \max_{0 \leqslant k \leqslant s} \|\mu^{(k)}\|_{L^2(\mathbb{R})} \leqslant R, \quad \|\mu^{(s)}\|_{L^\infty(\mathbb{R})} \leqslant R.$$

Since the underlying Lévy triplet is only identifiable if $\mathscr{O}(x)$ is known for all $x \in \mathbb{R}$, we consider the asymptotics of a growing number of observations with

$$\Delta := \max_{j=2,\dots,N} (x_j - x_{j-1}) \to 0 \qquad \text{and} \qquad A := \min(x_N, -x_1) \to \infty. \tag{9.20}$$

In contrast to standard regression estimates we shall always track explicitly the dependence on the level (δ_k) of the noise in the observations, which is usually rather small for observed option prices. The subsequent analysis can certainly be improved for a concrete design (x_j) and concrete noise levels (δ_j), but for revealing the main features it is more transparent and concise to state the results in terms of the abstract noise level

$$\varepsilon := \Delta^{3/2} + \Delta^{1/2} \|\delta\|_{l^\infty}, \tag{9.21}$$

comprising the level of the numerical interpolation error and of the stochastic error simultaneously. Here we use the norms $\|\delta\|_{l^\infty} := \sup_k \delta_k$ and $\|\delta\|_{l^2}^2 := \sum_k \delta_k^2$.

We now state the main results about the risk upper bounds of the estimators obtained by the basic procedure.

Theorem 9.4. *Assume $e^{-A} \lesssim \Delta^2$ and $\Delta \|\delta\|_{l^2}^2 \lesssim \|\delta\|_{l^\infty}^2$. For any $\bar{\sigma} > \sigma_{max}$ we choose*

$$U_{\bar{\sigma}} := \bar{\sigma}^{-1} \big(2 \log(\varepsilon^{-1})/T\big)^{1/2}, \quad U_0 := \varepsilon^{-2/(2s+5)}, \tag{9.22}$$

in the cases $\sigma_{max} > 0$ and $\sigma_{max} = 0$, respectively. Then every estimator $\hat{q} \in \{\hat{\sigma}^2, \hat{\gamma}, \hat{\lambda}, \hat{\mu}\}$ for the corresponding parameter q satisfies the following asymptotic risk bound:

$$\sup_{\mathscr{T} \in \mathscr{G}_s(R, \sigma_{max})} \mathbb{E}_{\mathscr{T}}[\|\hat{q} - q\|^2]^{1/2} \lesssim v_{q,\sigma_{max}},$$

where $\|\bullet\|$ denotes the absolute value for $q \in \{\sigma^2, \gamma, \lambda\}$ and the $L^2(\mathbb{R})$-norm for $q = \mu$ and the rate $v_{q,\sigma_{max}}$ is given in Table 4.

The two assumptions in the theorem are not very severe: because of the exponential decay of $\mathscr{O}$ the width A of the design only needs to grow logarithmically and the error levels (δ_k) need only be square summable after renormalisation. The latter condition can certainly be further relaxed since this term is caused by a rough bound on the quadratic remainder term.

For the lower bounds we refer to the equivalence between the regression and the Gaussian white noise model, as established by [16], and consider merely the idealized observation model

$$\mathrm{d}Z(x) = \mathscr{O}(x)\,\mathrm{d}x + \varepsilon\,\mathrm{d}W(x), \quad x \in \mathbb{R}, \tag{9.23}$$

with the noise level asymptotics $\varepsilon \to 0$, a two-sided Brownian motion W and with $\mathscr{O} = \mathscr{O}_{\mathscr{T}}$ denoting the option price function from (9.7) for the given triplet $\mathscr{T}$. Here, the noise level ε corresponds exactly to the regression error $\Delta^{1/2}\|\delta\|_{l^\infty}$. Due to Assumption 1 the option price functions $\mathscr{O}$ decrease exponentially and the results by Brown and Low [16] remain valid for unbounded intervals. This simplification avoids tedious numerical approximations in the proofs that can be found in [10].

Theorem 9.5. *Let $s \in \mathbb{N}$, $R > 0$ and $\sigma_{max} \geqslant 0$ be given. For the observation model (9.23) and any quantity $q \in \{\sigma^2, \gamma, \lambda, \mu\}$ the following asymptotic risk lower bounds hold:*

$$\inf_{\hat{q}} \sup_{\mathscr{T} \in \mathscr{G}_s(R,\sigma_{max})} \mathbb{E}_{\mathscr{T}}[\|\hat{q} - q\|^2]^{1/2} \gtrsim v_{q,\sigma_{max}},$$

where $\|\bullet\|$ denotes the absolute value for $q \in \{\sigma^2, \gamma, \lambda\}$ and the $L^2(\mathbb{R})$-norm for $q = \mu$, the infimum is always taken over all estimators and the rate $v_{q,\sigma_{max}}$ is as in Table 4.

Compared to Theorem 5.7 on lower bounds for i.i.d. observations of a Lévy process, here the choice of alternatives is more restricted because the martingale condition needs to remain fulfilled, see Belomestny and Reiß [10], but the proof itself becomes easier since in this regression-type setting it suffices to bound directly the L^2-distance of the densities p_0, p_1, avoiding the problem of a density in the denominator.

9.6.2 Discussion of the results

As we want to identify the Lévy triplet exactly in the limit, we have to assume the asymptotics $\Delta \to 0$ and $A \to \infty$ in the upper bound result. The numerical interpolation error term $\Delta^{3/2}$ contained in ε can be made smaller by using higher-order schemes. On the other hand, the statistical error term $\Delta^{1/2}\|\delta\|_{l^\infty}$ cannot be avoided as proved by the lower bound. Another way to study the calibration problem is to keep the number N of observations fixed and just to consider the asymptotics $\|\delta\|_{l^\infty} \to 0$. In this case the original Lévy triplet is not identifiable and the triplet of interest has to be properly defined in the set of triplets giving rise to the uncorrupted option prices, cf. Cont and Tankov [22] for a minimum relative entropy approach.

Recall that the severe ill-posedness in the case $\sigma > 0$ is due to an underlying deconvolution problem with the Gaussian kernel of variance σ^2: the law of the diffusion part of X_T is convolved with that of the compound Poisson part to give the density of X_T. For small values of σ and finite samples the performance is not so bad, compare the simulations in Section 9.4; it just needs a lot more observations to improve on that.

At first sight the rates for the parametric estimation part are astonishing. They are worse than in usual semi-parametric problems which also indicates that misspecified parametric models will give unreliable estimates for the volatility and jump intensity. In the case $\sigma = 0$, however, these rates are easily understood when employing the language of distributions. With δ_0 denoting the Dirac distribution in zero and δ_0' its derivative we have

$$\log(\varphi_T(u)) = T\mathscr{F}\big(\gamma\delta_0' + \nu - \lambda\,\delta_0\big)(u).$$

Estimating the density of X_T and similarly its characteristic function from the noisy observations of $\mathscr{O}$ amounts roughly to differentiate the observed function twice, cf. Ait-Sahalia and Duarte [1]. This gives the minimax rate for ν and μ as that of estimating the second derivative of a regression function of regularity $s+2$. For the parameter λ it suffices to estimate the jump in the antiderivative of $\mathscr{F}^{-1}(\log(\varphi_T))$, which corresponds to a pointwise estimation problem in the first derivative of a regression function, while for γ the analogy is the estimation of the regression function itself at zero. This explains also why in the class $\mathscr{G}_s$ we have measured the regularity not only in L^2, but also uniformly. In fact, if we only assume an L^2-Sobolev condition, then the same lower bound techniques will yield slower rates for the parameters, as is typical for pointwise estimation problems.

Observe that the estimation of the jump density at zero is only possible by imposing a certain regularity there, otherwise it is clearly not possible to detect jumps of height zero.

10 Open ends

Finally, let us point out two important, but yet unresolved topics where we see a high potential for future research.

10.1 Multi- and high-dimensional spectral inference

So far, the main research focus was on observations of one-dimensional processes with the notable exception of the time-changed Lévy case. The extension of the spectral estimation method to the multidimensional case is in principle mostly straight-forward. In the finite intensity Lévy case the characteristic exponent is the sum of a polynomial of degree 2 in the frequency variables plus the Fourier transform of the jump measure ν on $\mathbb{R}^d$. The same weighted least squares approach as in Section 4 can then be used to estimate $\sigma \in \mathbb{R}^{d\times d}$, $b \in \mathbb{R}^d$ and $\lambda \in \mathbb{R}^+$. In the general case, however, an analogue of the measure ν_σ from Section 6 that naturally (from a topological point of view) incorporates both, the diffusion and the jump part of the Lévy process, is to be found. One candidate is derived from the Hessian of the characteristic exponent

$$-\nabla^2 \psi(u) = \mathscr{F}\bar{\nu}_\sigma \text{ with } \bar{\nu}_\sigma(dx) = \sigma^2\delta_0(dx) + (x \otimes x)\nu(dx).$$

Note that $\bar{\nu}_\sigma$ is an $\mathbb{R}^{d\times d}$-valued finite measure on $\mathbb{R}^d$, assuming a second finite moment for the Lévy process. A probabilistic question is then to derive a continuity result like Proposition 6.2 also for $\bar{\nu}_\sigma$. For the statistical analysis then a much finer result is needed to derive optimal results of convergence, which basically depends on a quantification of the modulus of continuity.

From a statistical perspective a multivariate problem generates completely new questions, particularly on the dependence between the marginal processes. Main features of multivariate jump processes are not covered by the linear correlation structure. To this end, the copula concept has been transferred to Lévy processes and one key inference question is to test whether a certain copula structure should be rejected or not based on empirical data, which has been addressed by Bücher and Vetter [17] in the high-frequency case. For low-frequency observations this is still a completely open question. In particular in view of financial applications, high interest lies in particular in the quantification of the tail dependence, which can describe how different assets react together on larger shocks and thus how well diversified the risk of a portfolio is in times of crises.

In mathematical statistics the problem of high-dimensional inference has been attracting major interest recently, assuming that the dimension d tends to infinity as the sample size n tends to infinity. These asymptotics cover features of real data better where the dimension is not small relative to the sample size. Moreover, interesting probabilistic and statistical questions turn up. In particular, it is shown that under sparsity or low rank assumptions a high-dimensional covariance matrix can be estimated much more accurately than for general models of dimension d, see e.g. Cai *et al.* [18] and the references therein. It seems that similar results can be obtained for the diffusion matrix $\sigma \in \mathbb{R}^{d\times d}$ of a Lévy process from low or high frequency observations. In the latter case, however, the nuisance of the jump part may interfere and worse rates might result as in dimension one, see Jacod and Reiß [37]. The corresponding estimation problem for the jump measure ν on $\mathbb{R}^d$ is equally interesting. In particular, important subclasses of multidimensional Lévy processes have been introduced, e.g. extensions of stable or self-decomposable processes (cf. Sato [52]), and the construction of asymptotically optimal estimators for these subclasses remains a challenging problem, both theoretically and for applications.

10.2 Spectral estimation of affine processes

Let X be a regular affine process. The formulas (2.6) and (2.7) imply

$$e^{-\mathtt{i}\langle u,x\rangle} \frac{\partial \varphi(u|t,x)}{\partial t}\Big|_{t=0} = F_0(u) + \langle x, F_1(u)\rangle. \tag{10.1}$$

So the right hand side of (10.1) is a linear function of x with the functions $F_0(u)$ and $F_1(u)$ of Lévy-Khintchine form (see Theorem 2.6). Hence, the spectral estimation principle of Section 4 can be applied to estimate the parameters of X provided a consistent estimate for the derivative $e^{-\mathtt{i}\langle u,x\rangle} \frac{\partial \varphi(u|t,x)}{\partial t}\Big|_{t=0}$ is available for all $u \in \mathbb{R}^d$. Assume that the process X is observed on a time grid $0, \Delta, 2\Delta, \ldots, n\Delta$ with $\Delta \to 0$ and $T := n\Delta \to \infty$. Now one can estimate the vector (F_0, F_1) by solving the least-squares problem

$$(\widehat{F}_0, \widehat{F}_1) = \operatorname*{arginf}_{(a,b)\in\mathbb{R}\times\mathbb{R}^d} \sum_{k=1}^{n} \left[\frac{e^{\mathtt{i}\langle u, X_{k\Delta} - X_{(k-1)\Delta}\rangle} - 1}{\Delta} - a - \langle X_{(k-1)\Delta}, b\rangle \right]^2. \tag{10.2}$$

Based on the estimates $(\widehat{F}_0(u), \widehat{F}_1(u))$, we can estimate the parameters of X. There are several open questions, in particular:

- Does the estimate $(\widehat{F}_0, \widehat{F}_1)$ converge to (F_0, F_1) and at which rate?
- How can the parameters of the affine process X be estimated based on $(\widehat{F}_0, \widehat{F}_1)$ and how large are the errors?
- Are the convergence rates optimal?

For a different approach towards the estimation of affine models see Belomestny [8]. This approach is based on blockwise local polynomial smoothing in time and space.

Acknowledgements We are indebted to Jakob Söhl and Mathias Trabs for careful proofreading, all remaining mistakes are our own.

Partial financial support by the Deutsche Forschungsgemeinschaft through SFB 649 "Economic Risk", Research Unit 1749 "Structural Inference in Statistics" and through SFB 823 "Statistical modelling of nonlinear dynamic processes", by Laboratory for Structural Methods of Data Analysis in Predictive Modeling, MIPT, RF government grant, ag. 11.G34.31.0073 and by the International Laboratory of Qualitative Finance, NRU HSE, RF government grant, ag. 14.A12.31.0007 is gratefully acknowledged.

References

1. Ait-Sahalia, Y. and Duarte, J. (2003). Nonparametric option pricing under shape restrictions. *Journal of Econometrics* **116** (1), 9–47.
2. Aït-Sahalia, Y. and Jacod, J. (2009). Estimating the degree of activity of jumps in high frequency data. *The Annals of Statistics* **37** (5A), 2202–2244.
3. Aït-Sahalia, Y. and Jacod, J. (2012). Analyzing the spectrum of asset returns: Jump and volatility components in high frequency data. *The Journal of Economic Literature* **50** (4), 1007–50.
4. Aït-Sahalia, Y. and Jacod, J. (2012). Identifying the successive blumenthal-getoor indices of a discretely observed process. *The Annals of Statistics* **40** (3), 1430–1464.
5. Barndorff-Nielsen, O. E. (1997). Processes of normal inverse gaussian type. *Finance and Stochastics* **2** (1), 41–68.
6. Basawa, I. V. and Brockwell, P. J. (1982). Non-parametric estimation for non-decreasing lévy processes. *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 262–269.
7. Bauer, F. and Reiss, M. (2008). Regularization independent of the noise level: an analysis of quasi-optimality. *Inverse Problems* **24** (5), 1–16.
8. Belomestny, D. (2011). Spectral estimation of the lévy density in partially observed affine models. *Stochastic Processes and their Applications* **121** (6), 1217–1244.
9. Belomestny, D. (2011). Statistical inference for time-changed lévy processes via composite characteristic function estimation. *The Annals of Statistics* **39** (4), 2205–2242.
10. Belomestny, D. and Reiss, M. (2006). Spectral calibration of exponential lévy models. *Finance and Stochastics* **10** (4), 449–474.
11. Belomestny, D. and Schoenmakers, J. (2011). A jump-diffusion libor model and its robust calibration. *Quantitative Finance* **11** (4), 529–546.
12. Belomestny, D. (2010). Spectral estimation of the fractional order of a Lévy process. *Ann.Stat.* **38** (1), 317–351.
13. Belomestny, D., Panov V. (2013). Abelian theorems for stochastic volatility models with application to the estimation of jump activity. *Stochastic Processes and their Applications* **123** (1), 15–44.
14. Belomestny, D., Panov V. (2013). Estimation of the activity of jumps in time-changed Lévy models. *Electronic Journal of Statistics* **7**, 2970–3003.
15. Bismut, J.-M. (1983). Calcul des variations stochastique et processus de sauts. *Zeitschrift für Wahrscheinlichkeitstheorie und Verw. Gebiete* **63**, 147–235.
16. Brown, L. D. and Low, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *The Annals of Statistics* **24** (6), 2384–2398.
17. Bücher, A. and Vetter, M. (2012). Nonparametric inference on lévy measures and copulas. *arXiv preprint arXiv:1205.0417, to appear in The Annals of Statistics* .
18. Cai, T. T., Liu, W., and Zhou, H. H. (2012). Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation. *arXiv preprint arXiv:1212.2882, to appear in The Annals of Statistics* .

19. Carr, P. and Madan, D. (1999). Option valuation using the fast fourier transform. *Journal of Computational Finance* **2** (4), 61–73.
20. Comte, F. and Genon-Catalot, V. (2011). Estimation for lévy processes from high frequency data within a long time interval. *The Annals of Statistics* **39** (2), 803–837.
21. Cont, R. and Tankov, P. (2004). *Financial modelling with jump processes*, vol. 133. Chapman & Hall/CRC Boca Raton.
22. Cont, R. and Tankov, P. (2004). Non-parametric calibration of jump-diffusion option pricing models. *Journal of Computational Finance* **7** (3), 1–50.
23. Csorgo, S. (1981). Limit behaviour of the empirical characteristic function. *The Annals of Probability* **9** (1), 130–144.
24. Dereich, S. (2011). Multilevel monte carlo algorithms for lévy-driven sdes with gaussian correction. *The Annals of Applied Probability* **21** (1), 283–311.
25. Duffie, D., Filipovic, D., and Schachermayer, W. (2003). Affine processes and applications in finance. *Ann. Appl. Probab* **13** (3), 984–1053.
26. Duffie, Filipović and Schachermayer (2003). Affine processes and applications in finance. *Ann. Appl. Prob.* **13** (3), 984–1053.
27. Eberlein, E. and Keller, U. (1995). Hyperbolic distributions in finance. *Bernoulli* **1** (3), 281–299.
28. Erdélyi, A. (1955). Asymptotic expansions. Tech. rep., DTIC Document.
29. Feuerverger, A. and Mureika, R. A. (1977). The empirical characteristic function and its applications. *The annals of Statistics* pp. 88–97.
30. Figueroa-López, J. E. (2009). Nonparametric estimation of time-changed lévy models under high-frequency data. *Advances in Applied Probability* **41** (4), 1161–1188.
31. Figueroa-López, J. E. (2011). Sieve-based confidence intervals and bands for lévy densities. *Bernoulli* **17** (2), 643–670.
32. Figueroa-López, J. E. and Houdré, C. (2006). Risk bounds for the non-parametric estimation of lévy processes. *Lecture Notes-Monograph Series* pp. 96–116.
33. Gnedenko, B. and Kolmogorov, A. (1954). *Limit Distributions for Sums of Independent Random Variables*. Cambridge, Mass.
34. Gugushvili, S. (2009). Nonparametric estimation of the characteristic triplet of a discretely observed Lévy process. *J. Nonparametr. Stat.* **21** (3), 321–343.
35. Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* **58** (301), 13–30.
36. Horowitz, J. and Mammen, E. (2007). Rate-optimal estimation for a general class of nonparametric regression models with unknown link functions. *Annals of Statistics* **35** (6), 2589–2619.
37. Jacod, J. and Reiss, M. (2012). A remark on the rates of convergence for integrated volatility estimation in the presence of jumps. *arXiv preprint arXiv:1209.4173263, to appear in The Annals of Statistics* .
38. Jacod, J. and Shiryaev, A. (1987). *Limit Theorems for Stochastic Processes*. Springer.
39. Juditsky, A. B., Lepski, O. V., and Tsybakov, A. B. (2009). Nonparametric estimation of composite functions. *The Annals of Statistics* **37** (3), 1360–1404.
40. Kamae, T. and Krengel, U. (1978). Stochastic partial ordering. *The Annals of Probability* **6** (6), 1044–1049.
41. Kappus, J. (2014). Adaptive nonparametric estimation for lévy processes observed at low frequency. *Stochastic Processes and their Applications* **124** (1), 730–758.
42. Kappus, J. and Reiss, M. (2010). Estimation of the characteristics of a lévy process observed at arbitrary frequency. *Statistica Neerlandica* **64** (3), 314–328.
43. Kawata, T. (1972). *Fourier analysis in probability theory*. Academic Press, New York. Probability and Mathematical Statistics, No. 15.

44. Keller-Ressel, M., Schachermayer, W., and Teichmann, J. (2013). Regularity of affine processes on general state spaces. *Electron. J. Probab* **18** (43), 1–17.
45. Kou, S. G. (2002). A jump-diffusion model for option pricing. *Management science* **48** (8), 1086–1101.
46. Liese, F. (1987). Estimates of Hellinger integrals of infinitely divisible distributions. *Kybernetika (Prague)* **23** (3), 227–238.
47. Massart, P. and Picard, J. (2007). *Concentration inequalities and model selection*, vol. 1896. Springer.
48. Neumann, M. H. and Reiss, M. (2009). Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli* **15** (1), 223–248.
49. Nickl, R. and Reiss, M. (2012). A donsker theorem for lévy measures. *Journal of Functional Analysis* **263**, 3306–3332.
50. Reiss, M. (2013). Testing the characteristics of a lévy process. *Stochastic Processes and their Applications* **123** (7), 2808–2828.
51. Rubin, H. and Tucker, H. G. (1959). Estimating the parameters of a differential process. *The Annals of Mathematical Statistics* pp. 641–658.
52. Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge.
53. Söhl, J. (2010). Polar sets for anisotropic gaussian random fields. *Statistics & Probability Letters* **80** (9), 840–847.
54. Söhl, J. (2012). Confidence sets in nonparametric calibration of exponential lévy models. *arXiv preprint arXiv:1202.6611, to appear in Finance and Stochastics* .
55. Söhl, J. and Trabs, M. (2012). Option calibration of exponential lévy models: Confidence intervals and empirical results. *arXiv preprint arXiv:1202.5983, to appear in J. Comp. Finance* .
56. Steutel, F. W. and Van Harn, K. (2003). *Infinite divisibility of probability distributions on the real line*. CRC Press.
57. Todorov, V. and Tauchen, G. (2012). Realized laplace transforms for pure-jump semimartingales. *The Annals of Statistics* **40** (2), 1233–1262.
58. Trabs, M. (2013). Information bounds for inverse problems with application to deconvolution and lévy models. *arXiv preprint arXiv:1307.6610* .
59. Trabs, M. (2014). Calibration of self-decomposable lévy models. *Bernoulli* **20** (1), 109–140.
60. Tsybakov, A. B. (2009). *Introduction to nonparametric estimation*. Springer Series in Statistics, Springer, New York. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.
61. Ushakov, N.G. (1999). *Selected topics in characteristic functions*. VSP, Utrecht.
62. Woerner, J. H. (2003). Variational sums and power variation: a unifying approach to model selection and estimation in semimartingale models. *Statistics & Decisions* **21** (1), 47–68.